A.A.S. **4**, II, 199-239 (2009)

MODELLING INCOME PROTECTION CLAIM TERMINATION RATES BY CAUSE OF SICKNESS I: RECOVERIES

By S. Y. Ling, H. R. Waters and A. D. Wilkie

ABSTRACT

In this paper we present methods and results for the estimation and modelling of the recovery intensity for Income Protection (IP) insurance claims, allowing for different causes of claim. We use UK data supplied by the Continuous Mortality Investigation relating to claims paid in the years 1975 to 2002, inclusive. Each claim is classified by one of 70 possible causes according to ICD8.

We group causes where appropriate, and then use the Cox model and generalised linear models to model the recovery intensity.

In two subsequent papers we complete our modelling of IP claim termination rates by discussing the modelling of the mortality of IP claimants.

There are two main reasons why it is useful to incorporate cause of sickness in the modelling of IP claim terminations:

(i) The cause of sickness will be known to the insurer for a claim in the course of payment. A reserve can be set more accurately for such a claim if a model of the termination rates appropriate for this cause is available.

(ii) Different causes of claim will become more or less significant over time. For example, tuberculosis may have been an important cause of sickness in the past, but is likely to be far less significant now; the swine flu pandemic starting in 2009 is likely to have a significant effect on observed aggregate claim termination rates, skewing them towards higher rates at shorter durations. Information about trends in morbidity, together with a model of termination rates by cause of claim, allows future aggregate claim termination rates to be predicted more accurately, reserves to be set at more appropriate levels and policies to be priced more accurately.

One of the covariates included in our models for recovery intensities is Calendar Year. Aggregate recovery intensities have been decreasing over the period considered, 1975 to 2002, and this is generally reflected in the models for recovery intensities by cause of sickness. However, when these intensities are projected for years beyond 2002, the results are not always plausible.

KEYWORDS

Income Protection Insurance; Recoveries; Cause of Claim; Proportional Hazards; Cox Model; Generalised Linear Models

CONTACT ADDRESS

Howard R. Waters, Department of Actuarial Mathematics and Statistics, and the Maxwell Institute for the Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. Tel: 00 44 131 4513211; E-mail: H.R.Waters@hw.ac.uk

1. Introduction

Income Protection (IP) is a form of individual long term insurance which pays a regular benefit during periods when the life assured is sick, provided the current sickness has lasted longer than the *Deferred Period* specified in the policy. The most common deferred periods in the UK are 1, 4, 13, 26 and 52 weeks, and we will refer to these as DP1, DP4, DP13, DP26 and DP52, respectively. The policyholder pays premiums during the term of the policy at times when sickness benefit is not being paid. For more information, see Kluwer (2001) and Sanders & Silby (1988).

The Continuous Mortality Investigation (CMI) is an organisation established by the Faculty of Actuaries and the Institute of Actuaries in the UK with responsibility for collecting, analysing and modelling mortality and morbidity data from participating insurance companies. In particular, the CMI collects data on IP policies on a yearly basis — policies in force at the start/end of the year and claims paid during the year.

In CMIR12 (1991), the CMI published details of the stochastic model it proposed, and now uses, to analyse and model IP data. This model is represented in Figure 1.

Key features of this model are as follows:

(i) This is an attempt to model an individual's life history in terms of his/her state of health. In other words, it models the underlying *sickness* process rather than the *claims* process.

(ii) The model is specified in terms of the transition intensities for moves between the three states.

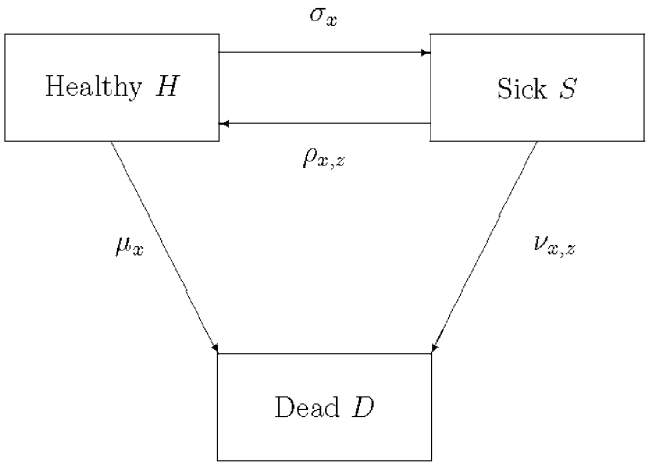


Figure 1. The model for IP insurance in CMIR12 (1991)

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

(iii) The transition intensities are functions of current age, x, and also, for sickness terminations, the duration of the current sickness, z. Individuals of a given age x who are not sick, are assumed to be 'equally healthy'. Also, the sickness termination intensities do not depend on the cause of the current sickness. Note that in this paper we will model the recovery intensity as a function of cause of sickness, sickness duration and *exact age at falling sick*, as well as other covariates. See Section 4.3. In Papers II and III we will model the force of mortality as a function of cause of sickness, sickness duration and *exact attained age*, as well as other covariates.

(iv) CMIR12 (1991) gives details of the modelling of all four intensities using data from 1975 to 1978. The modelling of the sickness termination intensities based on data from 1991 to 1998 was reported in CMIWP5 (2004).

The purpose of the research reported in this paper, and in its two accompanying papers, Ling *et al.* (2009b & 2009c), is to go some way to extending the model in Figure 1 to a model for IP insurance which allows for different causes of sickness. Such a model is represented in Figure 2.

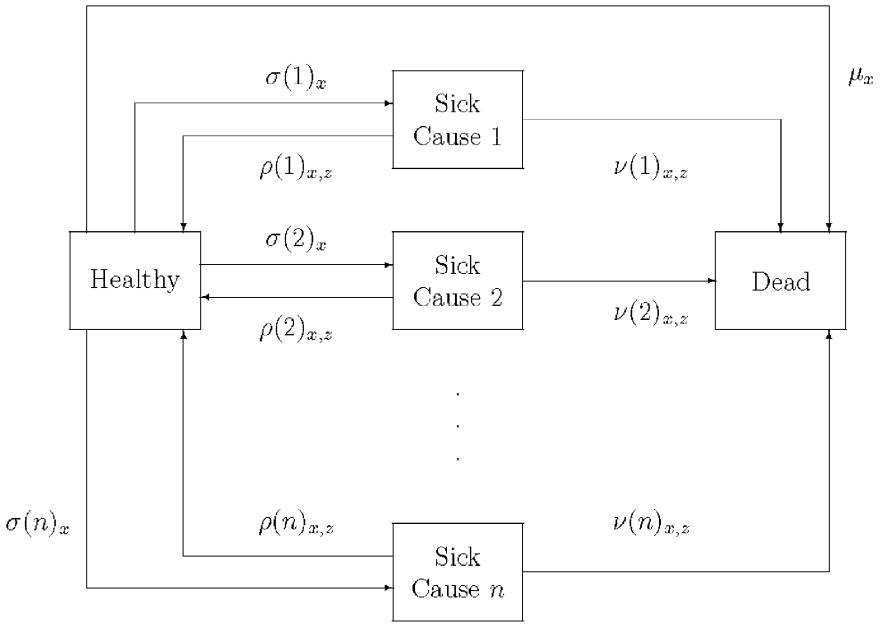


Figure 2. A model for IP insurance by cause of claim

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Key features of the model in Figure 2 are as follows:

(i) The 'Sick' state in Figure 1 has been divided into n separate states, one for each possible cause of sickness.
(ii) While suffering from a particular sickness, it is not possible for the individual to fall sick from another cause unless and until (s)he recovers from the current sickness.
(iii) Each of the n causes of sickness has its own inception intensity, $\sigma(i)_x$, and its own termination intensities, $\rho(i)_{x,z}$ (recovery) and $\nu(i)_{x,z}$ (mortality).
(iv) We could divide 'Dead' into n separate states, but these would be divided by cause of sickness rather than cause of death. This will be a point to note in Papers II and III.

In this paper we discuss the modelling of the recovery intensities, $\rho(i)_{x,z}$. In Papers II and III, Ling *et al.* (2009b & 2009c), we discuss the modelling of the mortality intensities, $\nu(i)_{x,z}$. We do not discuss in any of these papers the modelling of the sickness inception intensities, $\sigma(i)_x$. Some applications are presented in Paper III.

In CMIR12 (1991), the recovery intensity, $\rho_{x,z}$, was modelled using data from 1975 to 1978 relating to male assured lives whose policies were not rated. A single function adequately fitted the data for all deferred periods, although for deferred periods greater than 1 week there was a 4-week 'run in' period following the end of the deferred period where the recovery intensity increased to the aggregate level. The model for $\rho_{x,z}$ in CMIWP5 (2004) also used data from males only, in this case for the period 1991 to 1998 and for Occupation Class 1 (occupation class was not recorded for the 1975-78 data). The 'run in' period was still a feature for this model. One difference between the models in these two reports is that in CMIWP5 (2004) deferred period was included in the model as a multiplicative factor, so that, after any run-in period, the recovery intensity for any deferred period was a constant multiple of the recovery intensity for, say, DP1.

In Section 2 we describe the data made available to us by the CMI. Previous research into claim termination rates by cause of sickness includes Cordeiro (1998 & 2002) and CMIWP23 (2006), both of whom used CMI data. CMIWP23 (2006) provides a very full discussion of the CMI's IP data, the need for an analysis by cause of claim, previous research and the reasons, and criteria, for a preliminary grouping of causes of sickness. We comment on these papers in Section 3 and discuss the preliminary aggregation of our data into 12 sickness categories. In Section 4 we describe the methodology we used to analyse claim termination rates, illustrating this methodology by applying it to one of our categories: diseases of the circulatory system. Results for the other 11 categories are summarised in the Appendix, together with a list of individual causes comprising the category and some notes on the modelling. We discuss some of the results in Section 5.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Full details of the research reported in this paper can be found in Ling (2009), particularly Chapter 3, and Ling (2008).

The Acknowledgements and References for this paper and its accompanying papers, Ling *et al.* (2009b & 2009c), are given at the end of this paper.

2. Data

We were provided by the CMI with a set of IP data, comprising claim records relating to payments made during any of the investigation years from 1975 to 2002 inclusive. Information in each claim record included:

(a) Sex: male or female
(b) Deferred period: 1, 4, 13, 26 or 52 weeks
(c) Occupational rating
(d) Age last birthday at sickness inception
(e) Calendar year of the claim payment
(f) Occupation class
(g) Cause of claim (see Section 3)
(h) Month and year of birth
(i) Date of sickness inception
(j) Date of payment commencement
(k) Mode of payment commencement: continuation of claim payment from preceding year, new claim, new claim after interruption, revival and benefit change
(l) Date of payment cessation
(m) Mode of payment cessation: recovery, death, expiry and continuation of claim payment in succeeding year.

Details of the coding of records is given in CMIR2 (1976). The CMI did not collect 'Occupational class' code information until 1990, but did assign an occupational rating to records from 1975. Details are in CMIR18 (2000). For consistency, we classified each pre-1990 claim record as 'non-rated' if its 'Occupational rating' was 0, and each post-1990 claim record as 'non-rated' if its 'Occupation class' was 1 or was not given and its 'Occupational rating' was 0. All other claim records were classified as 'rated'.

Using the information described above, we are able to calculate the following quantities for each claim record:

(i) Sickness duration at the start of the claim payment. This is given by the difference between the date of payment commencement and the date of sickness inception.
(ii) Sickness duration at the end of the claim payment. This is given by the difference between the date of payment cessation and the date of sickness inception.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

(iii) Exact age at the start of claim payment. With only month and year of birth given, we assume that the day of birth is the 15th. The exact age of the claimant at the start of claim payment is given by the difference between the date of payment commencement and the date of birth.

3. Preliminary Classification of Cause of Claim

There are 70 possible causes of sickness for each of the claim records supplied by the CMI and they are coded according to the Abbreviated List C in ICD8 (1967). We have not included cause 0, 'Cause unknown', for obvious reasons, or causes 76 (ME) and 77 (HIV/AIDS) since, due to policy exclusions, these are not well represented in our data.

Our objective in this paper is to produce for each of the 70 causes a model for the recovery intensity. Since there is very little data for some of these causes, this suggests the use of generalised linear models (GLMs) with cause of sickness being treated as a covariate. However, such a purely statistical approach could, through lack of data, lead to causes relating to very different medical conditions having similar, or even identical, models for their recovery intensities. This would be undesirable since it would be expected that medical advances would affect recovery patterns for medically similar conditions in similar ways and those for medically different conditions in different ways. For this reason our approach was to group the 70 possible causes of sickness into *sickness categories* based on medical similarity and then to use statistical methods to model the recovery intensities for the different causes within each category using the ICD8 codes as covariates.

This procedure — preliminary grouping on medical grounds followed by statistical modelling within groups — was applied by Cordeiro (1988 & 2002) to CMI data from the period 1979-82. Cordeiro started with 18 groups and, as a result of her statistical analysis, combined these so that she concluded that only five different models were needed to describe the claim termination rates for IP policies. However, the data supplied to Cordeiro were very much less detailed than the data supplied to the current authors so that our methods for statistical analysis are very different from hers.

CMIR8 (1986) and CMIWP23 (2006), using CMI data from 1975-78 and 1991-2002, grouped the 70 causes into 14 and 11 categories, respectively, on the grounds of medical similarities before carrying out some statistical analyses. No modelling of the data was attempted in either paper. CMIWP23 (2006) examined the variation in sickness experience by quadrennium, deferred period and occupational class for each sickness category, separately for each sex. The marginal analyses presented in CMIWP23 (2006) do not

Table 1. The grouping of the 70 causes of sickness into 12 sickness categories

Sickness category	ICD8 code	Number of inceptions	Exposed to risk (days)	Number of	
				Recoveries	Deaths
G1 Infections and acute respiratory	01-19	6,355	537,461	4,883	40
G2 Neoplasms	20, 21	4,499	2,212,088	2,165	1,766
G3 Endocrine and metabolic	22-26	868	651,123	596	48
G4 Mental illness	27	8,512	8,616,813	5,280	219
G5 Nervous system and sensory organs	28-31	4,473	3,960,268	2,865	220
G6 Circulatory	32-38	9,360	7,976,298	6,173	552
G7 Respiratory	39-45	14,041	974,101	10,440	102
G8 Digestive (non-infectious)	47-51	6,115	1,038,801	5,753	114
G9 Genito-urinary	52-55	2,986	526,179	2,606	58
G10 Musculoskeletal	61, 62	17,200	9,746,827	13,131	144
G11 Injuries	66-70	15,636	4,376,070	13,758	88
G12 All other known causes	46, 56-60, 63-65	6,542	2,925,889	5,091	147
All sickness categories		96,587	43,532,918	72,741	3,498

shed light on how the covariates jointly relate to the recovery and mortality intensities from sick.

We grouped the 70 causes of sickness into 12 sickness categories, labelled G1 to G12, as set out in Table 1. Note that the grouping we adopted is very similar to that used in CMIWP23 (2006). We include in Table 1, for each sickness category, the ICD8 codes of the constituent causes, the total number of claim inceptions during the investigation years, the total number of days exposed to risk of claim termination (recovery or death) and the total numbers of recoveries and deaths.

The description of each cause within a category is given in the Appendix. We will refer to an individual cause by the label *cs* followed by its ICD8 code.

4. Modelling the Recovery Intensity

4.1 *Overview*

Our objective is to produce a model for the recovery intensity for each of the 70 causes of sickness of the form:

$$\rho(z, \mathbf{x}) = \rho_0(z) \exp(\mathbf{x}^T \beta(z)) \tag{1}$$

where z denotes duration of sickness, $\mathbf{x}$ is a vector of covariates and $\beta(z)$ is a vector of regression coefficients whose elements may be functions of z.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

If the elements of the regression vector were constants, rather than functions of z, this would be the classical Cox proportional hazards (PH) model — a baseline hazard, $\rho_0(z)$, which is a function of z, and then every covariate having a proportional effect on the baseline hazard. See Cox (1972).

The classical Cox PH model is a powerful tool since it allows the regression vector β to be estimated without the need to specify the baseline hazard. However, this power comes at a price — the basic assumption, that the effect of each covariate is multiplicative, is very strong. It is interesting to note that in relation to the modelling of claim termination rates for Australian IP data, Pitt (2007) reported that there were *clear violations of the assumption of proportional hazards for some of the key rating factors*. However, Pitt did not allow for the possibility of duration-varying effects, i.e. that the elements of the regression vector could be functions of z rather than constants, as we do.

Fitting even the classical Cox model is computationally intensive, particularly with a large data set such as ours. Fitting our more general model for $\rho(z, \mathbf{x})$ for a large number of causes of sickness is even more computationally intensive. For this reason, we adopt a pragmatic approach to fitting our models by proceeding step by step and, when computational efficiency requires it, by not allowing the structure of the model to change and holding some parameters fixed while others are being estimated. In the remainder of this section we illustrate this process by describing the fitting of models of the form (2) to the data for the causes of sickness in category G6, diseases of the circulatory system. We also describe the tests used to assess the goodness of fit of the models.

4.2 *Preliminary Data Analysis*

The recovery intensities for causes of sickness belonging to different categories are never modelled together due to the medical differences between such causes. Within a category, the first step in the modelling of the recovery intensities is to identify those causes with so little data that their data needs to be amalgamated with the data for another cause in the same category. The exposure and number of recoveries for each of the six individual causes in G6 are shown in Table 2.

Two causes, *cs*32 and *cs*33, have relatively few recoveries but both have reasonable amounts of exposure. It was decided that it was not necessary to amalgamate data for any of the causes in G6.

The next step is to identify those causes within the category whose data can reasonably be pooled for the purpose of estimating the recovery intensities, with the individual causes being included in the model as covariates. Pooling data in this way is good modelling practice — inference for causes with relatively little data can be strengthened by using information from causes with larger amounts of data — *provided the recovery intensities*

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 2. The causes of sickness in sickness category G6, circulatory

ICD8	Cause of sickness	Exposed to risk (days)	Recoveries
32	Active rheumatic fever	7,172	11
33	Chronic rheumatic heart disease	49,898	23
34	Hypertensive disease	657,927	566
35	Ischaemic heart disease	4,701,560	3,426
36	Cerebrovascular disease	1,503,893	389
37	Venous thrombosis and embolism	221,987	288
38	Other diseases of the circulatory system	824,861	1,470

for the different causes are sufficiently similar, where similar means that their baseline intensities are similar in shape. Our criterion for deciding whether the recovery intensities for any group of causes within a category are sufficiently similar is based on the following procedure. We model the recovery intensities for all the causes in the category relative to a reference cause, *cs*38 in this case, assuming a Cox PH model with cause as the only covariate. Hence, the model being used is:

$$\rho(z, \mathbf{x}) = \rho_0(z) \exp(\mathbf{x}^T \beta)$$

where $\mathbf{x}$ is a vector of indicator functions $(I_{cs32}, I_{cs33}, I_{cs34}, I_{cs35}, I_{cs36}, I_{cs37})$ each taking the value 1 if that particular cause is being considered. There is no need to specify the baseline hazard, $\rho_0(z)$, which is the recovery intensity for the reference cause, *cs*38, assumed to be a function of sickness duration alone. The model assumes that the recovery intensity for any of the other causes is a multiple of the recovery intensity for *cs*38, with the multiple being given by the exponential of the corresponding β coefficient.

The proportional hazards assumption can be tested graphically for each of the five causes separately, relative to the reference cause, *cs*38, by calculating the Schoenfeld residuals. Consider, for example, *cs*32. Let $\{z_i\}$ denote the set of event (recovery) times (durations of sickness) in our data for sickness category G6. The Schoenfeld residual is calculated at each time z_i as the difference between the value of I_{cs32} for the particular individual who recovers at duration z_i — this will be either 1 or 0, depending on whether this individual's cause of sickness is *cs*32 or not — minus the weighted average of the values of I_{cs32} for all individuals still at risk at duration z_i, with weights given by their likelihood of recovery at duration z_i. Thernau & Grambsch (1994) refined this statistic by proposing a scaled Schoenfeld residual. This statistic has mean zero and a standard deviation which can be estimated. The Schoenfeld residual is defined only at event times and so it is customary to smooth these plots, for example using a spline function.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

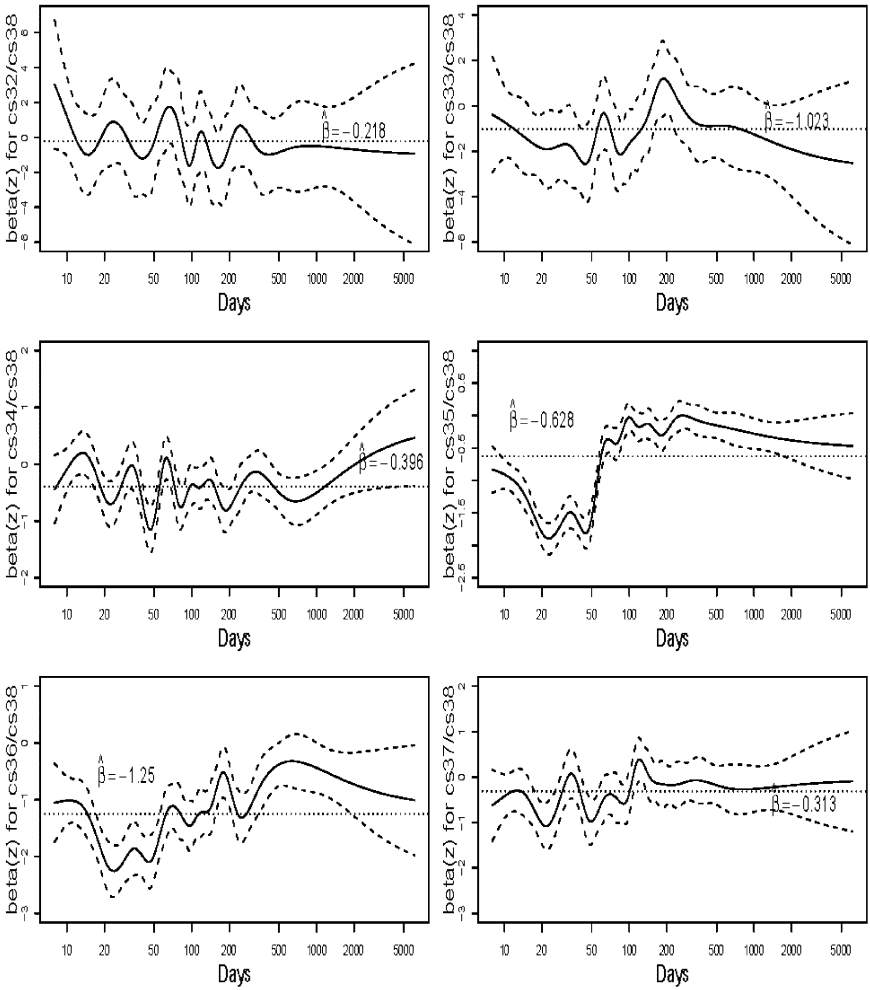


Figure 3. The smoothed plot of the Schoenfeld residuals for the causes of sickness in G6

Figure 3 shows for each of the six causes in category G6, relative to *cs*38, the plots of the smoothed scaled Schoenfeld residuals plus the assumed constant PH parameter estimated by the Cox model, $\hat{\beta}$, added for convenience, together with plus and minus two standard deviations. If the PH assumption is appropriate for a given cause, then the value of $\hat{\beta}$ should lie within the plus and minus two standard deviations envelope for significant

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

parts of the range of sickness durations. See Ling (2009, Section 3.2) and the references therein for further details of this test procedure. It can be seen that *cs*35 (Ischaemic heart disease) and *cs*36 (Cerebrovascular disease) fail the 'proportionality test', in the first case badly and in the second case not so badly, and so they will be modelled separately, whereas the other four causes pass the test, so that the data for causes *cs*32, *cs*33, *cs*34 and *cs*37 can be pooled with the data for *cs*38 for the modelling of the recovery intensities (with cause as a covariate). Note that this selection procedure takes account only of duration of sickness and cause. It takes no account of any other possible covariates, for example age and deferred period.

4.3 *Selection of Covariates*

The next step is to determine the basic structure of the model for the recovery intensities for a group of causes being modelled together. This means determining which of the possible covariates, and which interactions between covariates, should be included in the final model. At this stage we are not concerned with the baseline hazard, $\rho_0(z)$. The possible covariates, in addition to cause of sickness if more than one cause is included in the modelling, are shown in Table 3.

Both Age and Year are treated as continuous covariates and modelled using Chebycheff polynomials, $\{C_n(y)\}_{n=0}^{\infty}$. These are defined as follows:

$$C_0(y) = 1, \quad C_1(y) = y, \quad C_{n+1}(y) = 2yC_n(y) - C_{n-1}(y) \qquad \text{for} \quad n \geq 1$$

so that $C_n(y)$ is a polynomial in y of order n. There are some advantages in terms of modelling in using these polynomials rather than just powers of the variables. See Forfar *et al.* (1988). For computational stability, Age and Year are scaled and are represented by x_{age} and x_{year} respectively, where:

$$x_{age} = (\text{Age} - 43)/26, \quad x_{year} = (\text{Year} - 1988)/13$$

so that x_{age} ranges from -1 to 1 when age ranges from 17 to 69 and x_{year} ranges from -1 to (approximately) 1 when year ranges from 1975 to 2002.

Table 3. The possible covariates for the recovery intensity model

Covariate	Description
Age	Age (exact) at sickness inception (17-69)
Year	Attained calendar year (1975-2002)
Rating indicator	Indicator variable for rated; rated = 1, non-rated = 0
Sex	Indicator variable for female; female = 1, male = 0
Deferred period	Possible values for deferred period are 1 week (DP1), 4 weeks (DP4), 13 weeks (DP13), 26 weeks (DP26) and 52 weeks (DP52)

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

To determine which covariates to include in the model, we fit the classical Cox PH model:

$$\rho(z, \mathbf{x}) = \rho_0(z) \exp(\mathbf{x}^T \beta)$$

for all possible combinations of the covariates in Table 3 and cause of sickness, if appropriate, and interactions between them. The Chebycheff polynomials of Age and Year up to and including degree four are included in these models. The 'best' model is the one with the lowest value for its Akaike Information Criterion (AIC) (Akaike, 1974), which is defined as follows:

$$\text{AIC} = -2 \log(L_p(\hat{\beta})) + 2n \tag{2}$$

where $L_p(\hat{\beta})$ is the partial likelihood evaluated at the estimated coefficient vector $\hat{\beta}$ and n is the number of parameters in the model.

Table 4. Covariates and interactions retained in the three models for category G6

Cause(s)		
*cs*32, *cs*33, *cs*34, *cs*37, *cs*38	*cs*35	*cs*36
α_{cs32}		
α_{cs33}		
α_{cs34}		
α_{cs37}		
α_{age1}	α_{age1}	α_{age1}
α_{age2}	α_{age2}	α_{age2}
	α_{age3}	
α_{year}	α_{year}	α_{year}
	α_{year2}	
α_{rated}	α_{rated}	α_{rated}
		α_{sex}
α_{DP1}	α_{DP1}	α_{DP1}
α_{DP4}	α_{DP4}	α_{DP4}
$\alpha_{DP4:cs34}$		
α_{DP13}	α_{DP13}	α_{DP13}
$\alpha_{DP13:cs37}$		
α_{DP26}	α_{DP26}	
α_{DP52}	α_{DP52}	α_{DP52}
$\alpha_{age1:DP26}$		
	$\alpha_{age3:DP26}$	
	$\alpha_{age2:rated}$	
		$\alpha_{rated:DP13}$

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

For category G6, we carry out this procedure three times, once for *cs*32, *cs*33, *cs*34, *cs*37 and *cs*38, which are being modelled together, and once each for *cs*35 and *cs*36, which are being modelled separately. The covariates and interactions retained for each of the three models are shown in Table 4. The subscript for an entry in a column indicates a covariate to be included in the model; a double subscript indicates an interaction. Note that the integer following 'age' or 'year' in a subscript denotes the order of the Chebycheff polynomial to be included. In all three cases, Age, Year, Rated and Deferred Period were found to be significant covariates. In addition, Sex was significant for *cs*36, but not for the other two cases. Cause of sickness was retained as a covariate for the group of five sicknesses; this was not applicable for *cs*35 and *cs*36 on their own. See Ling (2009, Section 3.6.3).

We will continue to describe in detail the modelling of the recovery intensities for the group of causes *cs*32, *cs*33, *cs*34, *cs*37 and *cs*38 in Sections 4.4 to 4.8; the final results for causes *cs*35 and *cs*36 will be presented in Section 4.9.

4.4 *Testing the PH Assumption*

The PH assumption on which the modelling described in Section 4.3 is based is tested for each covariate in turn by plotting the smoothed scaled Schoenfeld residuals as a function of sickness duration, z, and checking whether the estimated parameter value lies consistently within the 95% confidence limits. For the group of causes *cs*32, *cs*33, *cs*34, *cs*37 and *cs*38, the only covariates which fail this test are those for Year, DP4 and DP13; the relevant plots are shown in Figure 4. For these covariates, the coefficients in the regression vector, β, need to be functions of z.

4.5 *Duration-dependent Coefficients*

Duration-dependent coefficients can take many forms. To describe the possible forms, take $\beta_{DP4}(z)$ as an example. If there were no duration dependence, we would have $\beta_{DP4}(z) = \alpha_{DP4}$. If there is duration dependence, we have:

$$\beta_{DP4}(z) = \alpha_{DP4} + f_{DP4}(z)$$

for some function $f_{DP4}(z)$. The duration dependence may persist for all durations or only for a certain period of sickness duration. In the latter case, the point at which the duration-varying effect started or ended is known as a 'break point'. The effect may be present only for shorter sickness durations or only for longer durations. To model the effect for longer durations, it is convenient to transform the variable z to $t_k(z)$, defined as follows in terms of a parameter $k(> 0)$:

$$t_k(z) = \frac{z}{1 + kz}.$$

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

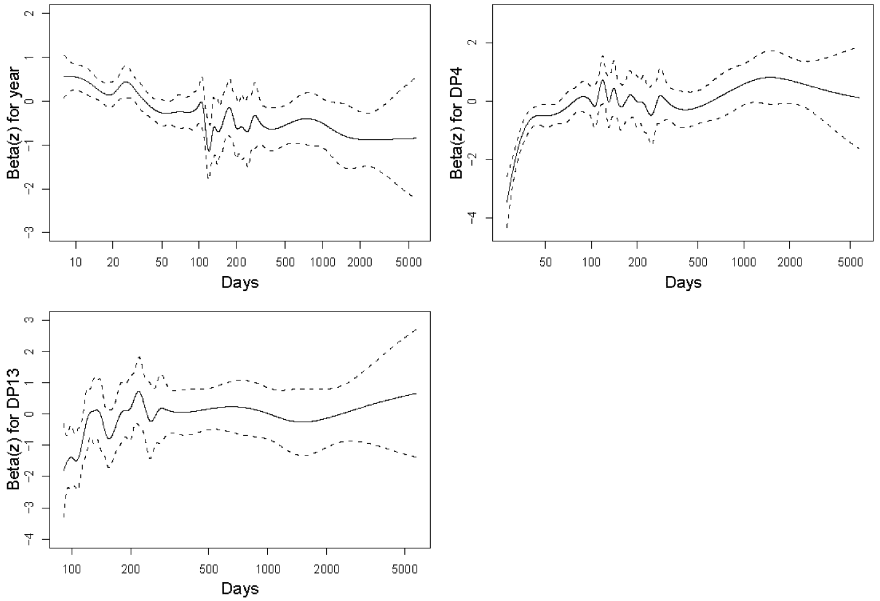


Figure 4. The smoothed plots of the scaled Schoenfeld residuals for x_{year}, x_{DP4} and x_{DP13} in the recovery intensity model for $cs32$, $cs33$, $cs34$, $cs37$ and $cs38$

The transformed duration variable $t_k(z)$ is used because it converges to an upper limit of $1/k$ as z increases without limit so that when the recovery intensity for very long sickness durations is calculated by extrapolating the graduation formula beyond the data range, the results obtained are more likely to be sensible. This transformed duration variable was also used, for the same reason, in CMIWP5 (2004).

The function $f_{DP4}(z)$ is assumed to consist of one or more of the following components that represent different types of duration dependency:

(i) $\gamma_{DP4}(\tau_{DP4} - z)_+$

(ii) $\theta_{DP4_i} C_i(t_k(z)), \quad i = 1, 2, \ldots$

(iii) $\zeta_{DP4_i}(C_1(t_k(\tau_{DP4})) - C_1(t_k(z)))_+^i, \quad i = 1, 2, \ldots$

(iv) $\phi_{DP4_i}(C_1(t_k(z)) - C_1(t_k(\tau_{DP4})))_+^i, \quad i = 1, 2, \ldots$

where $\gamma_{DP4}, \theta_{DP4_i}, \zeta_{DP4_i}$ and ϕ_{DP4_i} are regression parameters, τ_{DP4} is a break point and $y_+ = y$ if $y > 0$ and 0 otherwise. Note that $\theta_{DP4_1} \equiv \theta_{DP4}$, $\zeta_{DP4_1} \equiv \zeta_{DP4}$ and $\phi_{DP4_1} \equiv \phi_{DP4}$. In the case of two breakpoints, we will refer to the first one as τ_{DP4_1} and the second one as τ_{DP4_2}.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Inspection of Figure 4 indicates the following features:

Year: There is a long term near (log-)linear decline in the Year effect as a function of sickness duration. Note that the z-axis in Figure 4 is on a log scale.

DP4: The effect is piece-wise linear with two break points around 37.5 and 70.5 days; thereafter it is constant.

DP13: The effect is piece-wise linear with a single break point around 125.5 days; thereafter it is constant.

These considerations lead to the inclusion in the model of the terms:

$$\begin{aligned}\beta_{year}(z) &= \alpha_{year} + \theta_{year} C_1(t_k(z)) \\ \beta_{DP4}(z) &= \alpha_{DP4} + \gamma_{DP4_1}(\tau_{DP4_1} - z)_+ + \gamma_{DP4_2}(\tau_{DP4_2} - z)_+ \\ \beta_{DP13}(z) &= \alpha_{DP13} + \gamma_{DP13}(\tau_{DP13} - z)_+\end{aligned}$$

where $\tau_{DP4_1} = 37.5$, $\tau_{DP4_2} = 70.5$ and $\tau_{DP13} = 125.5$.

The τ values were chosen by visual inspection of Figure 4. More sophisticated methods were tried and found to be computationally intensive, time-consuming and unlikely to provide a better fit at the end of the modelling exercise. These values are not re-evaluated at any later stage of the modelling exercise.

At this stage a value for k is chosen somewhat crudely. Typically IP data for claims is heavily weighted to short durations; the value of k indicates how much the duration of sickness axis needs to be 'stretched' to distribute the data more evenly. Causes of sickness where duration is likely to be particularly short, for example some causes within category G1 (Infections and acute respiratory), are likely to require a small value of k, perhaps less than one, and sicknesses likely to last longer, for example those in category G6 (Circulatory), are likely to require a larger value of k, perhaps of the order of 7.

Having fixed the τ values and chosen, temporarily, a value for k, the coefficients in the regression vector $\beta(z)$ can be estimated by maximising the partial likelihood. Note that:

(a) The structure of the model is unchanged — this was fixed at an earlier stage. See Section 4.3.

(b) All the coefficients in $\beta(z)$ are re-estimated or, in the case of θ_{year}, γ_{DP4_1}, γ_{DP4_2} and γ_{DP13}, estimated for the first time.

(c) No attempt has yet been made to estimate the baseline intensity, $\rho_0(z)$.

See Ling (2009) for details.

We denote by $\hat{\beta}_g$ the vector of coefficients estimated at this stage.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

4.6 *Parameterising the Baseline Hazard*

The baseline hazard is a function only of sickness duration, z, and is assumed to have the form:

$$\rho_0(z) = \exp\left(\sum_{i=0}^{s} b_i C_i(t_k(z))\right)$$

where the parameters s, k and b_i need to be determined.

We first determine the values for s and k, so that the structure of the baseline hazard is fixed. We do this by fixing the regression vector β at the value $\hat{\beta}_g$ found in Section 4.5, maximising the full likelihood (as a function of the parameters $b_0, \ldots, b_s$ and k) for various combinations of s and k and then using the AIC to select the values for s and k which give the best fit.

The final stage in this process is to fix the τ parameters at their values found in Section 4.5, fix s at the value just determined, fix k, for both the baseline hazard *and* the duration-dependent parameters in $\beta(z)$, at the value just determined and to re-estimate all other parameters by maximising the full likelihood.

4.7 *The Final Model*

The general expression for a fully parameterised recovery intensity model is as follows:

$$\begin{aligned}\rho(z, \mathbf{x}) = \exp\Bigg(&\sum_{i=0}^{n} b_i C_i(t_k(z))\Bigg) \exp(x_{sex}\beta_{sex}(z) + x_{rated}\beta_{rated}(z) \\ &+ x_{DP4}\beta_{DP4}(z) + x_{DP13}\beta_{DP13}(z) + x_{DP26}\beta_{DP26}(z) + x_{DP52}\beta_{DP52}(z) \\ &+ \sum_{i=1}^{n} x_{yeari}\beta_{yeari}(z) + \sum_{i=1}^{n} x_{agei}\beta_{agei}(z) + \mathbf{\Phi}\Bigg)\end{aligned} \tag{3}$$

where each subscripted $\beta(z)$ denotes the duration-varying coefficient of its corresponding subscripted covariate x and $\mathbf{\Phi}$ denotes any interaction terms.

The final parameter values for the recovery intensity models for the group of causes *cs*32, *cs*33, *cs*34, *cs*37 and *cs*38 are shown in Table 5.

4.8 *Testing Goodness of Fit*

The adequacy of the final models of the recovery intensities for any cause or group of causes within a sickness category can be checked in a number of ways. The most straightforward way involves calculating a χ^2 statistic on the basis of actual and expected recoveries.

The methodology used to calculate a χ^2 statistic is identical to that set out in CMIR15 (1996). For each combination of covariate levels for Sex, Rating,

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 5. Parameters for the recovery intensity models for *cs*32, *cs*33, *cs*34, *cs*37 and *cs*38

Parameter	Value (standard error)	Parameter	Value (standard error)
k	6.7	α_{DP4}	0.1347
b_0	−49621.41		(0.0896)
	(6521.934)	$\alpha_{DP4:cs34}$	−0.2384
b_1	29804.68		(0.1113)
	(4072.891)	τ_{DP4_1}	37.5
b_2	−65398.31	γ_{DP4_1}	−0.2136
	(8597.399)		(0.0393)
b_3	9869.151	$\gamma_{DP4_1:cs37}$	0.1332
	(1348.994)		(0.0644)
b_4	−15776.75	τ_{DP4_2}	70.5
	(2075.904)	γ_{DP4_2}	−0.0197
α_{cs32}	0.5272		(0.0039)
	(0.3184)	α_{DP13}	−0.2037
α_{cs33}	−0.7275		(0.1101)
	(0.2114)	$\alpha_{DP13:cs37}$	0.5062
α_{cs34}	−0.2093		(0.1744)
	(0.0581)	τ_{DP13}	125.5
α_{cs37}	−0.2787	γ_{DP13}	−0.0394
	(0.0725)		(0.0095)
α_{age}	−0.6092	α_{DP26}	−0.1649
	(0.0719)		(0.1756)
α_{age2}	−0.2881	α_{DP52}	−0.7099
	(0.0707)		(0.2860)
α_{rated}	−0.1257	$\alpha_{age1:DP26}$	−1.0881
	(0.0605)		(0.3691)
$\alpha_{rated:cs32}$	−2.3618		
	(1.0506)		
α_{year}	0.5774		
	(0.1004)		
θ_{year}	−9.4564		
	(1.2166)		

Deferred Period and Year band, referred to as a tableau, data are laid out in a two dimensional array with (suitably chosen) age bands as rows and sickness duration bands as columns. Two such arrays are constructed for each tableau, one containing the actual numbers of recoveries (A) and the other the expected numbers of recoveries (E), calculated using the exposure and the modelled recovery intensities. In each cell (the intersection of each distinct row and column), we calculate $z = D/\sqrt{E}$, where D is the difference between A and E, allowing for a continuity correction since the actual number of recoveries is necessarily an integer.

For each z to approximate a normal variate, its expected number of recoveries E has to be greater than a certain number, taken as 8 in CMIR15 (1996). Thus, cells with fewer than 8 expected recoveries are merged with

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

adjacent cells until a total of at least 8 expected recoveries is obtained, and tableaux with a total of less than 8 expected recoveries have to be merged with other tableaux. Details about the merger of tableaux and the subsequent grouping of cells within each tableau can be found in Ling (2009, Appendix E).

The sum of the squares of the zs, after necessary grouping is carried out, has approximately a χ^2 distribution with the number of degrees of freedom equal to the number of cells (after grouping) minus the number of parameters fitted in the model.

For the group of causes *cs*32, *cs*33, *cs*34, *cs*37 and *cs*38 in sickness category G6, the value of the χ^2 statistic is 177.6. With 195 cells and 26 parameters fitted in the model, the probability value is 0.31, which does not suggest a poor fit to the data.

The χ^2 statistic tests the goodness of fit at an overall level; it does not necessarily detect a poor fit at the level of individual covariates. To take an extreme example, if A were greater than E for every cell involving males and less than E for every cell involving females, we would almost certainly not regard the model as satisfactory. However, if the differences between A and E were always small, the model would not fail the χ^2 test. One way to check for this feature for each level of each covariate is to plot the *partial residual effects* for each cause of sickness. To explain these plots in simple terms, consider, for example, the covariate Sex. For each (discretised interval of) sickness duration, we can calculate from the data an estimate of the recovery intensity:

(i) using just the data for males, weighted by exposure for males over all the covariates other than Sex; and
(ii) using just the data for females, weighted by exposure for females over all the covariates other than Sex.

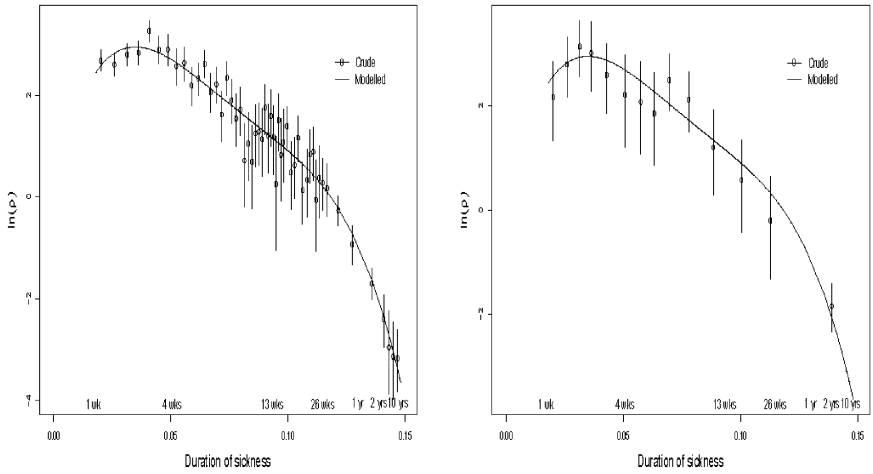
We can also calculate approximate 95% confidence intervals for these point estimates. From our model and the data we can calculate the recovery intensity as a function of:

(a) sickness duration and Sex (male), by weighting by exposure for males over all the covariates other than Sex; and
(b) sickness duration and Sex (female), by weighting by exposure for females over all the covariates other than Sex.

The plots of (i) and (a) and of (ii) and (b) against sickness duration are the partial residual effects plots for Sex for the given cause and these plots can be used to check the adequacy of the model for each sex separately. See Ling(2009) for more details.

Figure 5 shows the partial residual effects plots for Sex for *cs*38. Note that the duration of sickness axis in these plots has been rescaled using the function $z/(1+6.7z)$. It can be seen that the model is adequate in terms of Sex: it passes through most of the 95% confidence intervals and there is a reasonable spread of + and − values in terms of the point estimates.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press



(a) Male (b) Female

Figure 5. The partial residual effects plots for Sex for *cs*38

Further partial residual effects plots for each cause within category G6 (Circulatory), and for other causes, can be found in Ling (2008 & 2009).

One further test of the adequacy of the model, not illustrated here, is based on the *deviance residual*. This is a function of the actual and the expected recoveries in each cell and is similar to the z value used to calculate the χ^2 statistic. See McCullagh & Nelder (1989) and Ling (2009) for details, and Ling (2008 & 2009) for plots for each cause.

4.9 *Results for Causes cs35 and cs36*

The final parameter values for the recovery intensity models for causes *cs*35 and *cs*36, together with standard errors in parentheses, are shown in Table 6. Note that these recovery intensities were modelled separately from each other and from the other causes in category G6. The χ^2 statistics for each of these models are 270.3 with 271 degrees of freedom (probability value 0.50) and 19.7 with 17 degrees of freedom (probability value 0.29), respectively.

5. Some Results

In this section we present graphs illustrating some of the recovery intensities as functions of sickness duration for different causes and combinations of covariates.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 6. Parameters for the recovery intensity models for *cs*35 and *cs*36

Parameter	*cs*35 Value (standard error)	*cs*36 Value (standard error)
k	6.7	2.3
b_0	1128.687	32.1920
	(65.498)	(14.8636)
b_1	−8475.895	−107.9258
	(388.616)	(38.2616)
b_2	1124.992	29.5602
	(65.292)	(14.6404)
b_3	−2777.223	−28.7525
	(126.278)	(11.0018)
α_{sex}		−0.6349
		(0.2787)
α_{age}	0.4602	−0.8054
	(0.3787)	(0.1465)
α_{age2}	−0.6989	
	(0.2063)	
α_{age3}	0.3400	
	(0.1249)	
α_{rated}	−0.5489	−0.3976
	(0.1109)	(0.1744)
α_{year}	−0.3115	−0.4584
	(0.0391)	(0.1119)
τ_{year}	25.5	
γ_{year}	0.0648	
	(0.0103)	
α_{year2}	0.0923	
	(0.0387)	
α_{DP4}		0.4191
		(0.1487)
τ_{DP4}	72.5	81.5
γ_{DP4}	−0.0295	−0.0338
	(0.0031)	(0.0074)
α_{DP13}		0.3185
		(0.1743)
τ_{DP13}	185.5	180.5
γ_{DP13}	−0.0083	−0.0141
	(0.0011)	(0.0046)
α_{DP26}	0.1931	
	(0.1427)	
τ_{DP26}	265.5	
γ_{DP26}	−0.0137	
	(0.0035)	
α_{DP52}	−0.6343	−1.1757
	(0.1806)	(0.5913)
$\alpha_{age3:DP26}$	0.6477	
	(0.1707)	
$\alpha_{age2:rated}$	−0.4784	
	(0.1436)	
$\alpha_{rated:DP13}$		0.6679
		(0.2753)

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

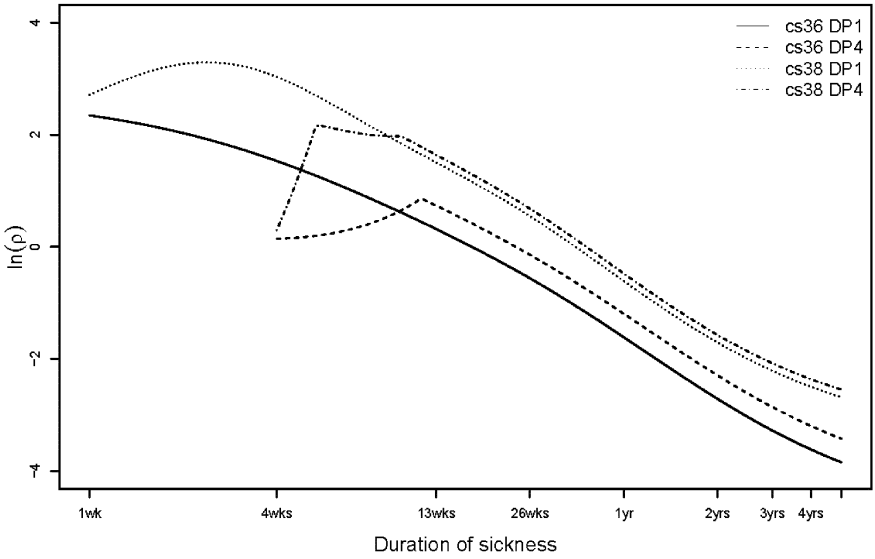


Figure 6. The recovery intensities for *cs*36 and *cs*38, DP1 and DP4, age 40, non-rated, 1988

5.1 *Figure 6*

Figure 6 shows the recovery intensities, on a log scale, for *cs*36 (males), cerebrovascular disease, and *cs*38, other diseases of the circulatory system, for both DP1 and DP4, for a life, non-rated, who is aged 40 at the start of sickness and for the calendar year 1988. The duration of sickness axis has been transformed using $t_k(z) = z/(1 + kz)$, where $k = 6.7$. Points to note about these intensities are as follows, with parameter values taken from Tables 5 (*cs*38) and 6 (*cs*36).

(i) The intensities for *cs*38 apply to both males and females; Sex is not a significant covariate for this cause. For *cs*36, the intensities for females are a constant multiple, $\exp(-0.6349) = 0.5300$, of those for males.

(ii) The graphs show the fundamentally different shapes for these causes for short durations of sickness. This endorses our decision, based on Figure 3, to model these causes separately.

(iii) For DP4 the graphs show clearly the 'run in' period for both causes of sickness, where the recovery intensity increases for a few weeks following the end of the deferred period.

For *cs*38 the run in period has two separate sections, with break points at 37.5 (τ_{DP4_1}) and 70.5 (τ_{DP4_2}) days. After 70.5 days, the intensity for DP4 is a constant multiple, $\exp(0.1347) = 1.1442$, of the intensity for DP1.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

For *cs*36 the run in period has a single section with a break point at 81.5 days. After 81.5 days, the intensity for DP4 is a constant multiple, $\exp(0.4191) = 1.5206$, of the intensity for DP1.

(iv) The intensities shown in Figure 6 are for a life aged 40 at the start of sickness.

For *cs*36 Age is represented in the model by the single factor $\alpha_{age} = -0.8054$, with no interactions. To calculate the intensities for, say, a life aged 50 at the start of sickness we need to multiply the intensities shown by a factor

$$\exp\left(-0.8054\left(\frac{50-43}{26}-\frac{40-43}{26}\right)\right) = 0.7336.$$

For *cs*38 Age is represented in the model by two factors, $\alpha_{age} = -0.6092$ and $\alpha_{age2} = -0.2881$, with no interactions. To calculate the intensities for, say, a life aged 50 at the start of sickness we need to multiply the intensities shown by a factor

$$\exp\left(-0.6092\left(\frac{50-40}{26}\right)-0.2885\left(2\left(\frac{50-43}{26}\right)^2-2\left(\frac{40-43}{26}\right)^2\right)\right) = 0.7646.$$

(v) All the intensities in Figure 6 relate to the calendar year 1988, the mid-point of the years contributing to our data. However, Year is a significant covariate in the models for both *cs*36 and *cs*38.

It is interesting to note that the models in CMIR12 (1991) and CMIWP5 (2004) did not include a time trend — in each case the data came from periods too short to identify any trends. However, indications of a decrease in the recovery intensity from 1975-78 to 1999-2002 can be found in CMIR18 (2000, Table A4) and CMIR22 (2005, Table A4). Renshaw & Haberman (2000) investigated the presence of any significant time trend in UK IP data from 1975 to 1994, making use of results in an earlier paper, Renshaw & Haberman (1995), which used a Generalised Linear Model (GLM) based approach to the modelling of the transition intensities.

For *cs*36 we can adjust the intensities so that they relate to other years by multiplying by a factor $\exp(-0.4584(\text{Year} - 1988)/13)$. For example, intensities for the calendar year 1975 would be a multiple 1.5815 and intensities for 2002 would be a multiple 0.6323 of those shown in Figure 6. This deterioration in recovery rates over the range of our data is consistent with the values in CMIR18 (2000, Table 4A) and CMIR22 (2005, Table 4A) for all causes combined and for deferred periods greater than one week. The recovery intensities for 2011, which is well beyond the range of our data, would be a multiple 0.4444 of

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

those shown in Figure 6. Whether this continued deterioration is realistic is beyond the scope of this paper. A pragmatic approach for anyone requiring recovery intensities applicable to years later than 2002 might be to use the 2002 rates as produced by our models, but, given the deterioration in these rates over the period 1975 to 2002, caution would need to be exercised.

For *cs*38 the Year adjustment is a function of the duration of sickness z (in weeks), given by

$$\exp\left(\left(0.5774 - 9.4564\frac{z}{1+6.7z}\right)\left(\frac{\text{Year} - 1988}{13}\right)\right).$$

Hence, for $z = 26$ weeks and for calendar year 2002, the recovery intensities for both DP1 and DP4 would be a multiple 0.4109 of those shown in Figure 6.

5.2 *Figure 7*

Figure 7 shows the recovery intensities, on a log scale, for *cs*20, malignant neoplasms, and *cs*21, benign and unspecified neoplasms, for DP1, for both males and females, for a non-rated life who is aged 50 at the start of sickness

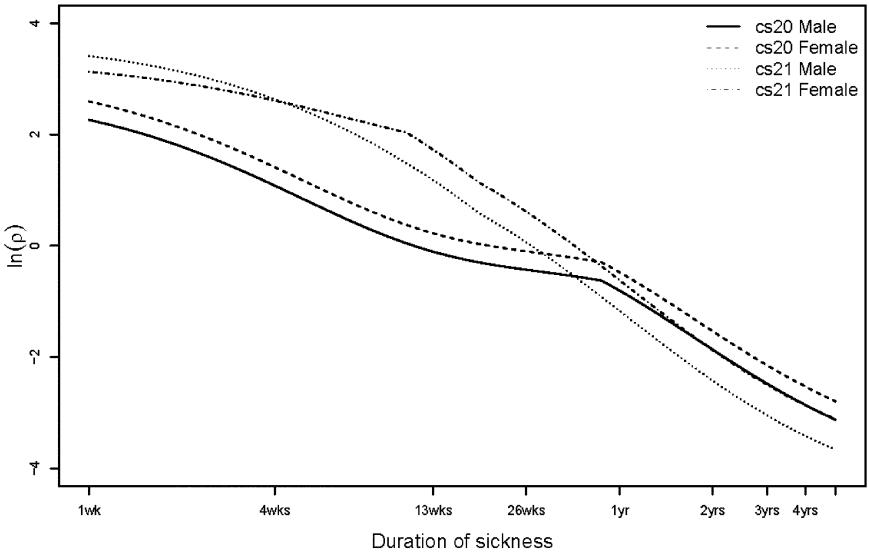


Figure 7. The recovery intensities for *cs*20 and *cs*21, males and females, age 50, DP1, non-rated, 2002

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

and for the calendar year 2002. The duration of sickness axis has been transformed using $t_k(z) = z/(1 + kz)$, where $k = 2.3$. The parameter values for the recovery intensities for these two causes are given in Table 9. Features of these intensities are as follows.

(i) The recovery intensities for *cs*21 are higher than those for *cs*20 for about one year and then they are broadly comparable. We would expect a higher recovery rate from benign compared to malignant neoplasms. The convergence after one year may be because *cs*21 includes benign *and* unspecified neoplasms; if many claimants suffering from a benign neoplasm have recovered within one year, the remaining cases in *cs*21 may be heavily weighted towards malignant neoplasms.

(ii) For *cs*20 females have a consistently higher recovery intensity than males. This may well be because many cases for females are breast cancer and many cases for males are lung cancer — we cannot check this from our data — and these diseases have different prognoses.

6. Acknowledgements

The authors are grateful to:

(a) The Continuous Mortality Investigation for supplying the data used in this research.

(b) Heriot-Watt University for their financial support for one of the authors, Sing-Yee Ling, while this research was being carried out.

References for Papers I, II and III

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

CMIR2 (1976). *Continuous Mortality Investigation Reports: Number 2.* The Institute of Actuaries and the Faculty of Actuaries.

CMIR6 (1983). *Continuous Mortality Investigation Reports: Number 6.* The Institute of Actuaries and the Faculty of Actuaries.

CMIR8 (1986). *Continuous Mortality Investigation Reports: Number 8.* The Institute of Actuaries and the Faculty of Actuaries.

CMIR12 (1991). *Continuous Mortality Investigation Reports: Number 12.* The Institute of Actuaries and the Faculty of Actuaries.

CMIR15 (1996). *Continuous Mortality Investigation Reports: Number 15.* The Institute of Actuaries and the Faculty of Actuaries.

CMIR17 (1999). *Continuous Mortality Investigation Reports: Number 17.* The Institute of Actuaries and the Faculty of Actuaries.

CMIR18 (2000). *Continuous Mortality Investigation Reports: Number 18.* The Institute of Actuaries and the Faculty of Actuaries.

CMIR22 (2005). *Continuous Mortality Investigation Reports: Number 22.* The Institute of Actuaries and the Faculty of Actuaries.

CMIWP5 (2004). *Continuous Mortality Investigation Working Paper: Number 5.* The Institute of Actuaries and the Faculty of Actuaries.

CMIWP23 (2006). *Continuous Mortality Investigation Working Paper: Number 23.* The Institute of Actuaries and the Faculty of Actuaries.

Cordeiro, I.M.F. (1998). A stochastic model for the analysis of permanent health insurance claims by cause of disability. Ph.D. Thesis, Heriot-Watt University, Edinburgh.

Cordeiro, I.M.F. (2002). A multiple state model for the analysis of permanent health insurance claims by cause of disability. *Insurance: Mathematics and Economics*, **30**, 167-186.

Cox, D.R. (1972). Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, **34(2)**, 187-220.

Devlin, T.F. & Weeks, B.J. (1986). Spline functions for logistic regression modeling. *Proc. 11th Annual SAS Users Group Intnl Conf.* Cary NC: SAS Institute, Inc., 646-651.

Dickman, P.W., Sloggett, A., Hills, M. & Hakulinen, T. (2004). Regression models for relative survival. *Statistics in Medicine*, **23**, 51-64.

Forfar, D.O., McCutcheon, J.J. & Wilkie, A.D. (1988). On graduation by mathematical formula. *Journal of The Institute of Actuaries*, **115(I)**, 1-149.

ICD8 (1967). *Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death, 8th Edition.* World Health Organisation.

Kluwer (2001). *Income protection insurance 2001.* Croner Publications and Kluwer Publishing, United Kingdom.

Lee, R.D. & Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association*, **87**, 659-671.

Ling, S.Y. (2008). Supporting document for Ph.D thesis at http://www.ma.hw.ac.uk/~singyee/.

Ling, S.Y. (2009). Models for income protection insurance incorporating cause of sickness. Ph.D. Thesis, Heriot-Watt University, Edinburgh.

Ling, S.Y., Waters, H.R. & Wilkie, A.D. (2009a). Modelling income protection claim termination rates by cause of sickness I: recoveries. *Annals of Actuarial Science*, **4**, 199-239.

Ling, S.Y., Waters, H.R. & Wilkie, A.D. (2009b). Modelling income protection claim termination rates by cause of sickness II: mortality of UK assured lives. *Annals of Actuarial Science*, **4**, 241-259.

Ling, S.Y., Waters, H.R. & Wilkie, A.D. (2009c). Modelling income protection claim termination rates by cause of sickness III: mortality. *Annals of Actuarial Science*, **4**, 261-286.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models.* Chapman and Hall, United Kingdom.

Pitt, D.G.W. (2007). Modelling the claim duration of income protection insurance policyholders using parametric mixture models. *Annals of Actuarial Science*, **2(I)**, 1-24.

Renshaw, A.E. & Haberman, S. (1995). On the graduation associated with a multiple state model in permanent health insurance. *Insurance: Mathematics and Economics*, **17(1)**, 1-17.

Renshaw, A.E. & Haberman, S. (2000). Modelling the recent time trends in UK permanent health insurance recovery, mortality and claim inception transition intensities. *Insurance: Mathematics and Economics*, **27(3)**, 365-396.

Sanders, A.J. & Silby, N.F. (1988). Actuarial aspects of PHI in the UK. *Journal of the Institute of Actuaries Students' Society*, **31**, 1-57.

Therneau, T. & Grambsch, P. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81(3)**, 515-526.

Willetts, R.C., Gallop, A.P., Leandro, P.A., Lu, J.L.C., Macdonald, A.S., Miller, K.A., Richards, S.J., Robjohns, N., Ryan, J.P. & Waters, H.R. (2004). Longevity in the 21st century. *British Actuarial Journal*, **10**, 685-832.

APPENDIX

The list below shows for each sickness category, other than G6, the individual causes in that category, together with their ICD8 codes, some notes on the modelling of the recovery intensities and the parameters of the fitted models.

G1 Infections and Acute Respiratory

ICD8 code	Cause of sickness
1	Typhoid, paratyphoid fever, other salmonella infections
2	Bacillary dysentery and amoebiasis
3	Enteritis and Other diarrhoeal diseases
4	Tuberculosis of respiratory system
5	Other tuberculosis, including late effect
6	Brucellosis
7	Diphteria
8	Whooping cough
9	Streptococcal sore throat and scarlet fever
10	Small pox
11	Measles
12	Viral encephalitis
13	Infectious hepatitis
14	Typhus and other rickettsioses
15	Malaria
16	Syphilis and its sequelae
17	Gonococcal infections
18	Helminthiases
19	All other infective and parasitic diseases

Notes: The data for causes *cs*7, *cs*14, *cs*16, *cs*17 and *cs*18 were amalgamated with the data for *cs*19 because of lack of data; their exposed-to-risk in days (number of recoveries) are 195(1), 329(4), 2,558(3), 57(3) and 769(3), respectively. Of the remaining causes of sickness, *cs*4, *cs*5, *cs*12 and *cs*13 were modelled together and *cs*1, *cs*2, *cs*3, *cs*6, *cs*8, *cs*9, *cs*10, *cs*11, *cs*15 and *cs*19+ were modelled together.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 7. Parameters for the recovery intensity models for causes *cs*4, *cs*5, *cs*12 and *cs*13

	*cs*4	*cs*5	*cs*12	*cs*13
Exposed to risk (days)	36,361	17,198	73,977	78,112
Number of recoveries	66	30	59	330
k	0.5	0.5	0.5	0.5
b_0	1.8335	1.7434	2.4943	2.2079
b_1	−3.0257	−3.0257	−4.9218	−3.0257
α_{rated}	−0.3231	−0.3231	−0.3231	−0.3231
α_{DP26}	−0.7099	−0.7099	−0.7099	−0.7099
α_{year}	−0.4472	−0.4472	−0.4472	−0.4472
α_{age}	−0.1061	−0.1061	−0.1061	−0.1061
τ_{DP4}	50.5	50.5	50.5	50.5
γ_{DP4}	−0.0329	−0.0329	−0.0329	−0.0329
τ_{DP13}	122.5	122.5	122.5	122.5
γ_{DP13}	−0.0549	−0.0549	−0.0549	−0.0549
τ_{year}	27.5	27.5	27.5	27.5
γ_{year}	0.0403	0.0403	0.0403	0.0403
τ_{age}	27.5	27.5	27.5	27.5
ϕ_{age}	−2.0455	−2.0455	−2.0455	−2.0455

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 8. Parameters for the recovery intensity models for causes *cs*1, *cs*2, *cs*3, *cs*6, *cs*8, *cs*9, *cs*10, *cs*11, *cs*15 and *cs*19+

	*cs*1	*cs*2	*cs*3	*cs*6	*cs*8	*cs*9	*cs*10	*cs*11	*cs*15	*cs*19+
Exposed to risk (days)	14,183	11,683	30,668	4,155	385	1,905	753	3,175	2,240	262,666
Number of recoveries	106	48	947	21	12	58	12	30	29	3,135
k	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3
b_0	29.9825	29.8673	30.5319	30.3101	30.3101	30.3101	30.3101	30.3101	29.9898	30.3101
b_1	−89.6762	−89.6762	−89.6762	−89.6762	−89.6762	−89.6762	−89.6762	−89.6762	−89.6762	−89.6762
b_2	25.9040	25.9040	25.9040	25.9040	25.9040	25.9040	25.9040	25.9040	25.9040	25.9040
b_3	−22.0979	−22.0979	−22.0979	−22.0979	−22.0979	−22.0979	−22.0979	−22.0979	−22.0979	−22.0979
α_{sex}	−0.1842	−0.1842	−0.1842	−0.1842	−0.1842	−0.1842	−0.1842	−0.1842	−0.1842	−0.1842
α_{rated}	−0.2666	−0.2666	−0.2666	−0.2666	−0.2666	−0.2666	−0.2666	−0.2666	−0.2666	−0.2666
α_{DP4}	0.3488	−0.1372	−0.5933	−0.1372	−0.1372	−0.1372	−0.1372	−0.1372	−0.1372	−0.1372
α_{DP13}	−0.3030	−0.3030	−0.3030	−0.3030	−0.3030	−0.3030	−0.3030	−0.3030	−0.3030	−0.3030
α_{DP26}	−0.7125	−0.7125	−0.7125	−0.7125	−0.7125	−0.7125	−0.7125	−0.7125	−0.7125	−0.7125
α_{year}	−0.5801	−0.5801	−0.5801	−0.5801	−0.5801	−1.6387	−0.5801	−0.5801	−0.5801	−0.5801
α_{age}	−0.3495	−0.3495	−0.7117	−0.3495	−0.3495	−0.3495	−0.3495	−0.3495	−0.3495	−0.3495
α_{age2}	−0.1331	−0.1331	−0.1331	−0.1331	−0.1331	−0.1331	−0.1331	−0.1331	−0.1331	−0.1331
τ_{DP4}	47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5
γ_{DP4}	−0.0800	−0.0800	−0.0800	−0.0800	−0.0800	−0.0800	−0.0800	−0.0800	−0.0800	−0.0800
τ_{DP13}	119.5	119.5	119.5	119.5	119.5	119.5	119.5	119.5	119.5	119.5
γ_{DP13}	−0.0421	−0.0421	−0.0421	−0.0421	−0.0421	−0.0421	−0.0421	−0.0421	−0.0421	−0.0421
τ_{year}	47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5
γ_{year}	0.0247	0.0247	0.0247	0.0247	0.0247	0.0247	0.0247	0.0247	0.0247	0.0247

G2 Neoplasms

ICD8 code	Cause of sickness
20	Maglinant neoplasm, including neoplasms of lymphatic and haematopoietic tissue
21	Benign neoplasms and neoplasms of unspecified nature

Notes: Causes *cs*20 and *cs*21 were modelled separately.

Table 9. Parameters for the recovery intensity models for sicknesses in G2 Neoplasms

	*cs*20	*cs*21
Exposed-to-risk (days)	1,919,924	292,164
Number of recoveries	1,621	544
k	2.3	2.3
b_0	50.6322	16.9710
b_1	−162.0055	−75.3337
b_2	48.1491	13.3819
b_3	−46.6650	−20.4126
α_{sex}	0.2402	0.3502
τ_{sex}		74.5
γ_{sex}		−0.0123
α_{rated}	−0.2647	
α_{age}	−0.9899	−0.7273
α_{year}		−0.5433
τ_{year}	319.5	126.5
ζ_{year_1}	−10.2329	4.8752
ζ_{year_2}	44.0975	
α_{DP4}	−0.2776	−0.3993
τ_{DP4}	58.5	52.5
γ_{DP4}	−0.0746	−0.0805
α_{DP13}	−0.3029	−0.4018
τ_{DP13}	185.5	165.5
γ_{DP13}	−0.0068	−0.0198
α_{DP26}	−0.5885	−0.4635
α_{DP52}	−0.4977	
$\alpha_{sex:age}$	0.3381	0.7415
$\alpha_{DP4:age}$	0.4438	

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G3 Endocrine and Metabolic

ICD8 code	Cause of sickness
22	Thyrotoxicosis with or without goitre
23	Diabetes Mellitus
24	Avitaminoses and other nutritional deficiency
25	Other endocrine and metabolic diseases
26	Anaemias

Notes: The recovery intensities for these five causes of sickness were modelled together.

Table 10. Parameters for the recovery intensity models for causes of sickness in G3 Endocrine and metabolic

	*cs*22	*cs*23	*cs*24	*cs*25	*cs*26
Exposed-to-risk (days)	209,164	246,938	6,044	137,380	51,597
Number of recoveries	91	137	9	292	67
k	2.3	2.3	2.3	2.3	2.3
b_0	2.7287	2.9605	3.5392	3.5392	3.1256
b_1	−14.5645	−14.5645	−14.5645	−14.5645	−14.5645
α_{DP4}	0.4594	−0.2807	−0.2807	−0.2807	−0.2807
α_{DP26}	−0.3933	−0.3933	−0.3933	−0.3933	−0.3933
α_{year}	0.0339	0.0339	0.0339	0.0339	0.0339
α_{age}	0.1178	−0.6799	−0.6799	−0.6799	−0.6799
α_{age2}	−0.2940	−0.2940	−0.2940	−0.2940	−0.2940
τ_{DP4}	52.5	52.5	52.5	52.5	52.5
γ_{DP4}	−0.0569	−0.0569	−0.0569	−0.0569	−0.0569
τ_{DP13}	140.5	140.5	140.5	140.5	140.5
γ_{DP13}	−0.0372	−0.0372	−0.0372	−0.0372	−0.0372
τ_{year}	20.5	20.5	20.5	20.5	20.5
γ_{year}	0.0453	0.0453	0.0453	0.0453	0.0453
θ_{year}	−3.0624	−3.0624	−3.0624	−3.0624	−3.0624

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G4 Mental Illness

ICD8 code	Cause of sickness
27	Psychoses and non psychotic mental disorders

Notes: There is only one cause, *cs*27, in this category.

Table 11. Parameters for the recovery intensity model for *cs*27

k	2.3	α_{DP4}	0.1729
b_0	−418.8920	τ_{DP4_1}	45.5
b_1	521.1417	γ_{DP4_1}	−0.0379
b_2	−531.0750	τ_{DP4_2}	94.5
b_3	171.7831	γ_{DP4_2}	−0.0159
b_4	−110.0431	α_{DP13}	
α_{sex}	0.0228	τ_{DP13}	198.5
α_{age}	−1.2484	γ_{DP13}	−0.0094
τ_{age}	106.5	α_{DP26}	−0.2385
ζ_{age_1}	26.3914	τ_{DP26}	250.5
ζ_{age_2}	−162.5144	γ_{DP26}	−0.0105
α_{age_2}	−0.2703	α_{DP52}	−0.5993
α_{rated}	0.0375	$\alpha_{age:DP26}$	−0.4554
α_{year}	−0.5211	$\alpha_{age:DP52}$	−0.7442
τ_{year}	26.5	$\alpha_{age:rated}$	0.2489
γ_{year}	0.0460	$\alpha_{year:rated}$	−0.1474
α_{year2}	0.1221	$\alpha_{sex:year2}$	−0.1631
		$\alpha_{sex:DP4}$	−0.1820

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G5 Nervous System and Sensory Organs

ICD8 code	Cause of sickness
28	Inflammatory diseases of eye
29	Cataract
30	Otitis media and mastoiditis
31	Other diseases of nervous system and sense organs

Notes: The recovery intensities for $cs28, cs30$ and $cs31$ were modelled together; $cs29$ was modelled on its own.

Table 12. Parameters for the recovery intensity models for causes of sickness in G5 Nervous system and sensory organs

	$cs28$	$cs29$	$cs30$	$cs31$
Exposed-to-risk (days)	235,093	92,897	44,719	3,587,559
Number of recoveries	424	285	148	2,008
k	2.3	6.7	2.3	2.3
b_0	4.1016	−308.3070	4.4025	3.6802
b_1	−18.0868	63.1057	−18.0868	−18.0868
b_2		−309.1824		
α_{sex}	−0.0937		−0.0937	−0.0937
α_{rated}	0.1686		0.1686	0.1686
α_{DP4}	−0.1120		−0.1120	−0.1120
α_{DP26}	−0.3538		−0.3538	−0.3538
α_{year}	−0.6030	1.2748	−0.6030	−0.6030
α_{age}	−0.9358		−0.9358	−0.9358
$\alpha_{sex:age}$	0.3825		0.3825	0.3825
$\alpha_{DP4:age}$	0.3385		0.3385	0.3385
$\alpha_{DP26:year}$	−0.4569		−0.4569	−0.4569
τ_{rated}	68.5		68.5	68.5
γ_{rated}	−0.0245		−0.0245	−0.0245
τ_{DP4}	46.5	56.5	46.5	46.5
γ_{DP4}	−0.0698	−0.0342	−0.0698	−0.0698
τ_{DP13}	121.5		121.5	121.5
γ_{DP13}	−0.0541		−0.0541	−0.0541
τ_{year}	45.5		45.5	45.5
γ_{year}	0.0302		0.0302	0.0302
θ_{year}		−12.7439		

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G7 Respiratory

ICD8 code	Cause of sickness
39	Acute respiratory infections
40	Influenza
41	Pneumonia
42	Bronchitis, emphysema and asthma
43	Hypertrophy of tonsils and adenoids
44	Pneumoconioses and related diseases
45	Other diseases of respiratory system

Notes: The recovery intensities for $cs39, cs40, cs42, cs43$ and $cs44$ were modelled together; the recovery intensities for $cs41$ and $cs45$ were fitted separately from the other causes and from each other.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 13. Parameters for the recovery intensity models for causes of sickness in G7 Respiratory

	*cs*39	*cs*40	*cs*41	*cs*42	*cs*43	*cs*44	*cs*45
Exposed-to-risk (days)	127,930	41,123	67,836	505,352	14,379	7,373	210,108
Number of recoveries	2,639	4,504	630	1,419	329	38	881
k	2.3	2.3	6.7	2.3	2.3	2.3	2.3
b_0	4.5644	4.8340	−302.4215	4.3385	4.4458	4.1399	29.1540
b_1	−17.7844	−17.7844	61.1308	−17.7844	−17.7844	−17.7844	−103.2332
b_2			−304.1462				25.1611
b_3							−27.5685
α_{sex}	−0.0796	−0.2212		−0.0796	−0.0796	−0.0796	
α_{rated}			−0.4901				
α_{DP4}	−0.5160	−0.5160	−0.1596	−0.5160	−0.5160	−0.5160	−0.4317
α_{DP13}	−1.1482	−1.1482	−0.6582	−1.1482	−1.1482	−1.1482	
α_{DP26}	−0.3686	−0.3686		−0.3686	−0.3686	−0.3686	−1.0302
α_{DP52}	−1.0029	−1.0029		−1.0029	−1.0029	−1.0029	
α_{year}	−0.0206	−0.0206	0.1820	−0.0206	−0.0206	−0.0206	−0.0037
α_{age}	−0.5603	−0.4438	−0.6164	−0.7953	−0.5603	−0.5603	−0.7969
α_{age2}	−0.0347	−0.0347		−0.0347	−0.0347	−0.0347	−0.2947
$\alpha_{DP4:age}$			0.4928				
$\alpha_{DP26:age}$	−1.2614	−1.2614		−1.2614	−1.2614	−1.2614	
$\alpha_{DP4:year}$	−0.4343	−0.4343	−0.7182	−0.4343	−0.4343	−0.4343	
$\alpha_{DP13:year}$	−0.9366	−0.9366		−0.9366	−0.9366	−0.9366	
$\alpha_{DP4:age2}$	−0.3969	−0.3969		−0.3969	−0.3969	−0.3969	
$\alpha_{DP13:age2}$	−0.8687	−0.8687		−0.8687	−0.8687	−0.8687	
$\alpha_{year:age}$							−0.3893
τ_{DP4_1}	40.5	40.5	59.5	40.5	40.5	40.5	42.5
γ_{DP4_1}	−0.1001	−0.1001	−0.0496	−0.1001	−0.1001	−0.1001	−0.1187
τ_{DP4_2}	70.5	70.5		70.5	70.5	70.5	
γ_{DP4_2}	−0.0156	−0.0156		−0.0156	−0.0156	−0.0156	
τ_{DP13}							220.5
γ_{DP13}							−0.0111
τ_{year_γ}	35.5	35.5		35.5	35.5	35.5	
γ_{year}	0.0114	0.0114		0.0114	0.0114	0.0114	
τ_{year_ϕ}							20.5
ϕ_{year}							−3.2097

G8 Digestive

ICD8 code	Cause of sickness
47	Peptic Ulcer
48	Appendicitis
49	Intestinal obstruction and hernia
50	Cholelithiasis and cholecystitis
51	Other diseases of digestive system

Notes: The recovery intensities for *cs*48 and *cs*49 were modelled together; the recovery intensities for *cs*47, *cs*50 and *cs*51 were fitted separately from the other causes and from each other.

Table 14. Parameters for the recovery intensity models for causes of sickness in G8 Digestive

	*cs*47	*cs*48	*cs*49	*cs*50	*cs*51
Exposed-to-risk (days)	91,916	24,189	268,100	70,661	583,935
Number of recoveries	350	608	2,618	612	1,565
k	1.3	6.7	6.7	2.3	2.3
b_0	3.0712	−375.3728	−375.5133	−28.1944	55.5717
b_1	−8.0560	93.3918	93.3918	13.0201	−177.4550
b_2		−375.9880	−375.9880	−30.1205	51.8543
b_3					−50.4236
α_{sex}		−0.4958	−0.4958		−0.1155
α_{rated}	−0.3572	−0.0171	−0.0171	−0.5489	−0.2075
α_{DP4}		−0.2369	−0.4718		0.1674
α_{DP13}		−0.0213	−0.0213	−0.5161	
α_{DP26}		−0.8313	−0.8313		−0.4412
α_{year}	−1.4777	0.2038	0.2038	−0.5034	−0.4041
α_{year2}		0.1512	0.1512	0.2684	
α_{age}	−6.8031	−0.5330	−0.5330	−0.3858	−0.6502
$\alpha_{sex:age}$					0.6943
$\alpha_{DP4:age}$		0.5364	0.5364		
$\alpha_{DP4:year}$		−0.3284	−0.3284		
$\alpha_{DP13:year}$		−0.7464	−0.7464		
$\alpha_{rated:age}$		−0.2655	−0.2655		
$\alpha_{sex:DP4}$		0.5211	0.5211		
τ_{rated}		75.5	75.5		
γ_{rated}		−0.0269	−0.0269		
τ_{DP4_1}	49.5	40.5	40.5		42.5
γ_{DP4_1}	−0.0620	−0.0912	−0.0912	−0.1707	−0.1064
τ_{DP4_2}					90.5
γ_{DP4_2}					−0.0091
τ_{DP13}		128.5	128.5		119.5
γ_{DP13}		−0.0329	−0.0329		−0.0469
τ_{year_γ}		26.5	26.5		32.5
γ_{year}		0.0667	0.0667		0.0351
τ_{year_ζ}	135.5			52.5	
ζ_{year_1}	9.7968			49.9025	
ζ_{year_2}				−460.9677	
θ_{age_1}	4.1772				
θ_{age_2}	−6.3111				

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G9 Genito-Urinary

ICD8 code	Cause of sickness
52	Nephritis and nephrosis
53	Calculus of urinary system
54	Hyperplasia of prostate
55	Other diseases of genito-urinary system

Notes: For *cs*55, the recovery intensities for males and females were fitted separately because their recovery patterns were very different from each other. We refer to *cs*55 for males and females as *cs*55M and *cs*55F, respectively. The recovery intensities for *cs*52, *cs*53 and *cs*55M were modelled together; the recovery intensities for *cs*54 and *cs*55F were fitted separately from the other causes and from each other.

Table 15. Parameters for the recovery intensity models for causes of sickness in G9 Genito-urinary

	*cs*52	*cs*53	*cs*54	*cs*55M	*cs*55F
Exposed-to-risk (days)	102,633	30,929	23,737	211,813	157,067
Number of recoveries	155	163	259	1,211	818
k	1.3	1.3	6.7	1.3	6.7
b_0	3.1425	3.5696	−404.5940	3.6364	1690.9520
b_1	−9.5505	−9.5505	103.1949	−9.5505	−12041.2100
b_2			−404.0155		1684.5420
b_3					−3940.3580
α_{rated}	−0.2545	−0.2545		−0.2545	−0.8554
α_{DP4}	−0.1718	−0.1718		−0.1718	−0.1288
α_{DP13}					−0.5375
α_{year}	0.3411	−0.8032		−0.3135	−0.3416
α_{age}	−0.8187	−0.2099		−0.2099	−1.1964
$\alpha_{DP4:age}$					0.9581
$\alpha_{DP13:age}$					1.5520
$\alpha_{rated:year}$					0.5599
τ_{DP4}	43.5	43.5	47.5	43.5	49.5
γ_{DP4}	−0.1051	−0.1051	−0.0659	−0.1051	−0.0787
τ_{DP13}	155.5	155.5		155.5	
γ_{DP13}	−0.0218	−0.0218		−0.0218	
τ_{year}	50.5	50.5		50.5	
γ_{year}	0.0168	0.0571		0.0168	
τ_{age}	38.5	38.5		38.5	
γ_{age}	−0.0330	−0.0330		−0.0330	

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G10 Musculoskeletal

ICD8 code	Cause of sickness
61	Arthritis and spondylitis
62	Other diseases of musculoskeletal system and connective tissue

Notes: The recovery intensities for *cs*61 and *cs*62 were modelled separately.

Table 16. Parameters for the recovery intensity models for causes of sickness in G10 Musculoskeletal

	*cs*61	*cs*62		*cs*61	*cs*62
Exposed-to-risk (days)	3,407,274	6,345,553			
Number of recoveries	1,847	11,284			
k	1.3	1.3	α_{DP4}	0.2852	0.1096
b_0	2.9727	11.3558	τ_{DP4_1}	37.5	40.5
b_1	−9.8003	−24.7647	γ_{DP4_1}	−0.1714	−0.1021
b_2		7.7475	τ_{DP4_2}	73.5	72.5
b_3		−3.6943	γ_{DP4_2}	−0.0296	−0.0242
α_{sex}	−0.4411	−0.2375	α_{DP13}	0.2760	
α_{age}	−0.9037	−0.2574	τ_{DP13}	135.5	131.5
α_{age_2}		0.0704	γ_{DP13}	−0.0355	−0.0362
α_{age_3}		0.1445	α_{DP26}	−12.4658	−0.1652
α_{age_4}		0.1231	θ_{DP26_1}	18.5551	
α_{rated}	−0.6394	−0.2852	θ_{DP26_2}	−7.5355	
θ_{rated}	1.4858		τ_{DP26}		206.5
τ_{rated}		206.5	γ_{DP26}		−0.0430
ϕ_{rated_1}		5.3474	α_{DP52}	0.3218	−0.4208
ϕ_{rated_2}		−11.2222	$\alpha_{sex:age}$	0.6884	
α_{year}	−0.2287	−0.2713	$\alpha_{sex:year}$		0.1988
θ_{year}	−1.5693	−0.9608	$\alpha_{sex:year2}$	−0.3165	
τ_{year}	32.5	30.5	$\alpha_{sex:DP26}$	−0.7751	
γ_{year}	0.0222	0.0345	$\alpha_{DP4:age}$		−0.4048
α_{year2}	0.2026	−0.0220	$\alpha_{DP13:age}$	−0.4698	−0.8077
			$\alpha_{DP26:age}$	−1.2593	−0.7695
			$\alpha_{DP52:age}$	−1.6507	
			$\alpha_{DP4:age3}$		−0.2575
			$\alpha_{DP13:age3}$		−0.2956
			$\alpha_{DP52:age3}$		0.5760
			$\alpha_{rated:year2}$		0.1940

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G11 Injuries

ICD8 code	Cause of sickness
66	Road transport accident
67	All other accidents
68	Attempted suicide and self-inflicted injuries
69	Attempted homicide and injury purposely inflicted by other persons; legal intervention
70	All other external causes

Notes: The recovery intensities for $cs68, cs69$ and $cs70$ were modelled together; the recovery intensities for $cs66$ and $cs67$ were fitted separately from the other causes and from each other.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 17. Parameters for the recovery intensity models for causes of sickness in G11 Injuries

	*cs*66	*cs*67	*cs*68	*cs*69	*cs*70
Exposed-to-risk (days)	1,458,667	2,383,238	18,158	39,193	476,814
Number of recoveries	3,893	7,695	60	117	1,993
k	2.3	2.3	2.3	2.3	2.3
b_0	−220.4817	26.8452	−449.0517	−448.6254	−448.2689
b_1	250.2069	−102.6865	550.8088	550.8088	550.8088
b_2	−280.6401	23.4335	−570.4026	−570.4026	−570.4026
b_3	81.8599	−29.3062	181.6597	181.6597	181.6597
b_4	−57.7592		−118.9095	−118.9095	−118.9095
α_{sex}	−0.1429	0.0511	−0.1819	−0.1819	−0.1819
α_{rated}	−0.2727	−0.1240	−0.2593	−0.2593	−0.2593
α_{DP4}	−0.1762	−0.1896	0.2256	0.2256	0.2256
α_{DP13}	−0.2107	−0.2924			
α_{DP26}	−0.4365	−0.4144			
α_{DP52}	−0.5208	−0.8429	−0.8803	−0.8803	−0.8803
α_{year}	0.2242	−0.1959	−0.3274	−0.3274	−0.3274
α_{year2}		0.1159	0.1697	0.1697	0.1697
α_{age}	−0.3068	−0.5003	−0.4127	−0.4127	−0.4127
α_{age2}			−0.1332	−0.1332	−0.1332
$\alpha_{DP4:age}$		0.2699			
$\alpha_{DP26:age}$	−0.8826				
$\alpha_{DP13:year}$	−0.3122				
$\alpha_{rated:year}$	−0.3346				
$\alpha_{sex:DP4}$		−0.2848			
τ_{DP4_1}	47.5	47.5	44.5	44.5	44.5
γ_{DP4_1}	−0.0803	−0.0869	−0.0775	0.0248	−0.0775
τ_{DP4_2}			92.5	92.5	92.5
γ_{DP4_2}			−0.0157	−0.0157	−0.0157
τ_{DP13}	119.5	128.5	115.5	115.5	115.5
γ_{DP13}	−0.0579	−0.0325	−0.0878	−0.0878	−0.0878
τ_{year_γ}	28.5	28.5	24.5	24.5	24.5
γ_{year}	−0.0163	0.0316	0.0280	0.0280	0.0280
τ_{year_ϕ}	67.5				
ϕ_{year}	−3.0723				
θ_{year}		−1.2328			

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

G12 All Other Known Causes

ICD8 code	Cause of sickness
46	Diseases of teeth and supporting structures
56	Abortion
57	Other complications of pregnancy, childbirth and the puerperium
58	Delivery without mention of complication
59	Infections of skin and subcutaneous tissue
60	Other diseases of skin and subcutaneous tissue
63	Congenital anomalies
64	Certain causes of perinatal morbidity
65	Other specified and ill-defined diseases

Notes: The recovery intensities for *cs*46, *cs*56, *cs*57, *cs*58, *cs*63, *cs*64 and *cs*65 were modelled together, as were the recovery intensities for *cs*59 and *cs*60.

https://doi.org/10.1017/S1748499500000713 Published online by Cambridge University Press

Table 18. Parameters for the recovery intensity models for causes of sickness in G12 All other known causes

	*cs*46	*cs*56	*cs*57	*cs*58	*cs*59	*cs*60	*cs*63	*cs*64	*cs*65
Exposed-to-risk (days)	7,763	1,965	14,971	1,233	126,352	194,363	89,725	36,664	2,452,853
Number of recoveries	201	18	108	2	701	642	102	140	3177
k	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3	2.3
b_0	29.8043	28.8907	28.8907	28.8907	4.1147	3.9567	28.5392	28.8907	28.8907
b_1	−89.4239	−89.4239	−89.4239	−89.4239	−16.4267	−16.4267	−89.4239	−89.4239	−89.4239
b_2	25.1623	25.1623	25.1623	25.1623			25.1623	25.1623	25.1623
b_3	−22.5471	−22.5471	−22.5471	−22.5471			−22.5471	−22.5471	−22.5471
α_{sex}	−0.0563	−0.0563	−0.0563	−0.0563			−0.0563	−0.0563	−0.0563
α_{rated}	−1.2448	−0.0209	−0.0209	−0.0209			−0.0209	−0.0209	−0.0209
α_{DP4}	−0.0552	−0.0552	−0.0552	−0.0552			−0.0552	−0.0552	−0.0552
α_{DP13}	0.0118	0.0118	0.0118	0.0118	0.5062	0.0120	0.0118	0.0118	0.0118
α_{DP26}	−0.2239	−0.2239	−0.2239	−0.2239	−0.4631	−0.4631	0.6170	−0.2239	−0.2239
α_{DP52}	−0.3710	−0.3710	−0.3710	−0.3710			0.8936	−0.3710	−0.3710
α_{year}	−0.4669	−0.7892	−0.7892	−0.7892	0.1197	0.1197	−0.7892	−0.7892	−0.7892
α_{age}	−0.4937	−0.4937	−0.4937	−0.4937	−0.4512	−0.4512	−0.4937	−0.4937	−0.4937
$\alpha_{DP26:age}$	−0.5812	−0.5812	−0.5812	−0.5812			−0.5812	−0.5812	−0.5812
$\alpha_{DP13:year}$	−0.3462	−0.3462	−0.3462	−0.3462			−0.3462	−0.3462	−0.3462
$\alpha_{rated:year}$	0.2946	0.2946	0.2946	0.2946			0.2946	0.2946	0.2946
$\alpha_{sex:year}$	0.2106	0.2106	0.2106	0.2106			0.2106	0.2106	0.2106
$\alpha_{sex:DP52}$	−1.0490	−1.0490	−1.0490	−1.0490			−1.0490	−1.0490	−1.0490
τ_{DP4}	47.5	47.5	47.5	47.5	63.5	63.5	47.5	47.5	47.5
γ_{DP4}	−0.1016	−0.1016	−0.1016	−0.1016	−0.0396	−0.0396	−0.1016	−0.1016	−0.1016
τ_{DP13}	121.5	121.5	121.5	121.5	119.5	119.5	121.5	121.5	121.5
γ_{DP13}	−0.0572	−0.0572	−0.0572	−0.0572	−0.0773	−0.0773	−0.0572	−0.0572	−0.0572
τ_{DP26}	220.5	220.5	220.5	220.5			220.5	220.5	220.5
γ_{DP26}	−0.0512	−0.0512	−0.0512	−0.0512			−0.0512	−0.0512	−0.0512
τ_{year_γ}	34.5	34.5	34.5	34.5			34.5	34.5	34.5
γ_{year}	0.0313	0.0313	0.0313	0.0313			0.0313	0.0313	0.0313
τ_{year_ϕ}					32.5	32.5			
ϕ_{year}					−4.1303	−4.1303			