METHODS OF STUDYING THE RISK PROCESS IN DISABILITY INSURANCE

Ove Lundberg
Stockholm

The problem

The aim of this study is to find suitable methods for utilizing available statistical information on the risk development of terminated insurances. This information may comprise not only data respecting sex, age, insurance tariff, etc., but also the points of time when certain events—disability, injuries or damages—have occurred as well as their durability and cause. The method study here presented is based on statistical material gathered from long-term disability insurances which have ceased to be valid, either in consequence of the expiration of the term of insurance, or the insured's death. We endeavour to study the point of time when the first event (disability) occurred and the relation between subsequent events, searching suitable methods for assessing risks after the occurrence of the primary event.

The influence of passed events upon the future risk process may be explained by either a direct dependence between actual events —actual damage or disability may have an impairment for the future of the risk—or by heterogeneity a priori. In the former case, we may speak of a contagious process (Polya-Eggenberger), and, in the latter case, the occurrence or non-occurrence of an event may, according to Lexis and Newbold, act as a risk differentiating factor providing a theoretical basis for a technique of experience rating.

It is difficult to decide whether the relationship noted in respect of a risk process is of the direct or the indirect kind. A direct dependence between one event and the following ones has no relevance to a study of the incidence of the first event and the time of its occurrence. We avoid in such a study the obstacles created by formal rules and regulations with regard to the problem of whether two or more disability periods occurring at short intervals should be considered as a single disability period or not. In order

https://doi.org/10.1017/S0515036100008114 Published online by Cambridge University Press

to measure the frequency of the first event we have—as in the case of measuring mortality—to limit the period exposed to risk until the occurrence of the first event.

The stochastic model

Different models exist intended to give the characteristic features of the course of events of an individual policy or a group of policies. It has proved useful to treat the random development of a policy risk as a stochastic process according to the general theory of risk.

A risk process of the number of events corresponds to a pure birth process which is characterized by an intensity function $k_n(t)$ depending on the number of events n during the time period $(0,t)$. The probability of the $n + 1$st event occurring during $(t,t + dt)$ is equal to $k_n(t).dt$ for small values of dt. If the intensity function is independent of t the process is called time-homogeneous [1]).

A Poisson process is defined by an intensity function independent of n and t. For a Compound Poisson process (in the narrow sense) the intensity function is uniquely defined by a completely monotonic function $P_0(t)$ corresponding to the probability of no claims during the time interval $(0,t)$. The function $k_n(t)$ is then for each number n a non-increasing function of t [2]).

Statistical studies

Disability policies taken by men in the Swedish sickness insurance company EIR have been observed from the date of entry (here identical with the date of issue or, at the earliest January 1, 1931) until *either* the terminating age z determined in the policy (record Z) *or* death, if the insured has died before that age (record D). In most of these policies the terminating age z is 60 or 65 years. The policies are issued with a waiting period of one or three months, but the material also includes policies with a waiting period of fourteen or seven days. All policies in this statistical material are combined with life insurance in a life insurance company. The records include all such policies—450 of record Z and 380 of record D

[1]) Feller, W. An Introduction to Probability Theory and its Applications. Second Ed. Vol. 1, p. 402.

[2]) Lundberg, O. On Random Processes and their Application to Sickness and Accident Statistics. Uppsala 1940, p. 70.

—as have been registered as terminated during the period 1955-1960.

As a measure of the time of observation we have used the accumulated risk premium (according to EIR's unloaded disability table of 1936) for a single payment of 1 unit as soon as the disability has lasted for three months.

As a function of age this "risk premium" equals the probability (or intensity) of the insured's being disabled for a period of at least three months. The increase of the "risk premium" as an increasing function of age has, in this way, been programmed in the time scale (operational time). Thus, no breakdown of the material by age will be necessary for this preliminary study.

The time of observation (= duration) of each policy has been divided into parts equalling one tenth of the operational time scale [1]). The number of disability periods of at least three months duration has been summed up for each class of duration, and the number recorded has been compared to the calculated number, which—according to the choice of time scale—is equal to the corresponding sum of the lengths of the periods exposed to risk for the duration class involved. The quotient of a duration class $t = t_i$ between the observed and the calculated number of first disability periods (primary events) will here be considered as an estimate of the function $k_0(t)$ for $t = t_i$.

The rates of the first (primary) disability period will be compared to the rates of the secondary disability periods. The influence of a primary disability period upon the future risk of disability will be studied by such a method.

In Table 1 we have calculated the rates of the primary and the secondary disability periods at a minimum duration of three months as well as the average rate of all such disability periods. By comparing records Z and D we may make the following statements:

---

[1]) A duration of 0.1 measured in the operational time being used corresponds to 7 years for an insured who has attained an age up to 33 years, 6 years for an insured who has reached age 41, etc., with approximately one year's reduction for every fifth individual year of age, so that for a 61-year-old person the 0.1 duration will correspond to about 2 years only.

https://doi.org/10.1017/S0515036100008114 Published online by Cambridge University Press

Table 1

Number of disability periods (claims) by the duration

Policies ended by the terminating age (record Z)

Durations class [1]	Number of primary disability periods [2]			Number of secondary disability periods			Number of all disability periods		
	calc [3]	obs	rate	calc [3]	obs	rate	calc [3]	obs	rate
0 00-0 09	43 8	24	0 55	1 2	5	4 2	45 0	29	0 65
10- 19	41 4	19	46	3 4	6	1 8	44 8	25	56
20- 29	37 6	26	69	5 4	10	1 8	43 0	36	84
30- 39	31. 0	20	65	6 8	12	1 8	37.8	32	82
40- 49	21 2	13	61	6 6	5	0 8	27 8	18	65
50-	17 8	11	62	7 1	10	1 4	24 9	21	85
Sum	192 8	113	59	30 5	48	1 6	223 3	161	72
with corr [4]				20 5	48	2 3	213 3	161	76

Policies ended by death (record D)

Durations class [1]	calc [3]	obs	rate	calc [3]	obs	rate	calc [3]	obs	rate
0 00-0 09	30 4	79	2 4	2 8	10	3 6	33 2	89	2 7
10- 19	16 3	52	3 2	4 2	22	5 2	20 5	74	3 7
20- 29	7 4	17	2 3	2 7	12	4 5	10 1	29	2 9
30-	4 9	18	3 7	3 4	13	3 8	8 3	31	3 7
Sum	59 0	166	2 8	13 1	57	4 4	72 1	223	3.1
with corr [1]				7.3	57	7.8	66 3	223	3 3

[1]) in operational time from the entry until the beginning of the disability period

[2]) = number of policies with 3 month-disability

[3]) = sum of unloaded "riskpremiums" (operational time)

[4]) correction due to periods of disability which are not exposed to risk

By primary disability is meant the first disability period beginning after the entry and lasting at least three months By secondary disability is meant each subsequent period of disability of at least three months' duration

1. The rates of record D are substantially higher than those of record Z.
2. The rates of the primary and the secondary disability periods show appreciable differences, more marked for record Z than for record D.
3. There is no marked trend in the rates with increasing duration, either for the primary, or for the secondary disability periods.

The first statement is easily explained. In fact, a strong correlation must be anticipated between the disability risk and the

https://doi.org/10.1017/S0515036100008114 Published online by Cambridge University Press

death risk due to the risk of the insured's death during a disability period. Thus, if we know that the policy period has been terminated by the insured's death, disability costs will generally be greater than if we had known that the insured had reached the terminating age.

The difference in rates between the primary and the secondary disability periods can be explained by heterogeneity a priori but also by a direct relationship between the periods. There is often an aggravated risk, not only of repeated disability periods due to the same illness but also of the insured's falling ill in consequence of other ailments resulting from the impaired condition of his general health. In the statistical material presented here such a positive correlation between the periods will, however, be counterbalanced by another element. In fact, according to the applied system of registration, no new disability period—either from the same, or from a new cause—can arise, as long as the insured is disabled.

Since there is no marked tendency with increasing duration in the rates of the primary disability periods, it is natural as a first approximation to assume $k_0(t)$ independent of t, an assumption corresponding to homogeneity a priori with regard to the time scale chosen. Since the secondary period rates are on a significantly higher level according to the table, the records will have become heterogeneous a posteriori as a result of the dependence between the disability periods.

Due to the limited material available, a study of the secondary disability periods would seem rather futile. As an illustration of the method, however, it may be interesting to study the risk process of the secondary periods as a sub-process starting at the onset of the primary disability period. We thus obtain the rates of Table 2 below.

In this table those policies are excluded for which the primary disability period has lasted until terminating age (record Z) or death (record D). The greater part of the corrections due to periods of disability which are not exposed to risk has thus been made belore calculating the rates. The decreasing tendency of the rates with increasing duration cannot be considered significant with regard to small number of events.

Since there is for increasing duration in the time scale chosen no

https://doi.org/10.1017/S0515036100008114 Published online by Cambridge University Press

Table 2

Number of secondary disability periods (claims) by the duration from the onset of the first disability period

Duration class	Number of secondary disability periods record Z			record D		
	calc.	obs.	ratio	calc.	obs.	ratio
0.00-0.09	9.9	21	2.1	5.0	38	7.6
.10- .19	6.0	10	1.7	2.1	14	6.7
.20- .29	5.7	5	0.9	1.2	5	4.2
.30-	4.9	12	2.4			
Sum	26.5	48	1.8	8.3	57	6.9
with corr. (see table 1)	20.5	48	2.3	7.3	57	7.8

significant trend of the rates of either the primary or the secondary disability periods, attention will be paid only to the sums of the table 1 over all durations.

After a primary disability period there is—according to Table 1—a risk of a new, i.e. secondary, disability period that can be calculated to the following quotient of the primary risk

Record	Uncorrected	Corrected
Z	1.6 : 0.59 = 2.7	2.3 : 0.59 = 3.9
D	4.4 : 2.8 = 1.6	7.8 : 2.8 = 2.6

Hence, the corrected secondary rate has been about three times as high as the primary rate.

The risk process might be described by models with an estimated value of k_0 and another, essentially higher, value of k_n for $n > 0$. Without further knowledge we can estimate all these k_n equal for $n > 0$. The number of events during a time interval $(0, t)$ is for such a pure birth process given by the differential equations or their solutions

$$P_0'(t) = -k_0 P_0(t) \qquad P_0(t) = e^{-k_0 t}$$

$$P_1'(t) = -k P_1(t) + k_0 P_0(t) \qquad P_1(t) = \frac{k_0}{k - k_0}\left(e^{-k_0 t} - e^{-kt}\right)$$

$$P_n'(t) = -k P_n(t) + k P_{m-1}(t) \quad \text{for } n \geqq 2, \qquad P_n(t) = k e^{-kt} \int_0^\infty P_{n-1}(t)\, e^{kt}\, dt$$

https://doi.org/10.1017/S0515036100008114 Published online by Cambridge University Press

The mean number of events as a function of t

$$m(t) = \sum_{n=0}^{\infty} n \, P_n(t)$$

can be deduced from the differential equations in an ordinary way. We get $m'(t) = k\text{-}(k\text{-}k_0)\, e^{-k_0 t}$ and

$$m(t) = k\,t - \frac{k - k_0}{k_0}\,(1\text{-}e^{-k_0 t})$$

We notice that the mean number of events is not linear in t and that the average rate $m'(t)$ increases from k_0 for $t = 0$ to k for $t = \infty$. Such an increase—which agrees with the tendence of the rates for all disability periods of Table 1—is thus a consequence of the aggravation of the average risk by the successive occurrence of primary events to the individual risks.

Summary: We have studied in two records the rate of the first (primary) disability period and the increased risk of disability that follows on the occurrence of a disability period of at least three months. There is no marked trend in the rates with increasing duration, either for the primary or for the secondary disability periods. A simple model of a time-homogeneous pure birth process may be applicable to the studied risk process with regard to the time scale chosen.

Although the rates of our statistics neither can, nor should serve as a basis for an "experience rating", the incidence of a primary disability period and the time of its occurrence might provide a basis for an appraisal of the future development of risks—disability risks as well as death risks.

Such an appraisal can, of course, be improved by paying regard to other details, e.g. the age of the policyholder at the occurrence of the disability period and the cause of disablement. We have, however, in this preliminary study been specially interested in the methods for finding a simple model useful for estimating the risk on the basis of easily obtainable details of the risk experience.