INTERPOLATION OF MORBIDITY RATES

F.G. Reynolds

The estimation of morbidity rates presents some particularly perplexing problems for the actuary. My purpose this morning is to acquaint you with the CIA Morbidity study and explain the techniques used to interpolate from the 1964 CDT table the expected incidence rates we are using.

Introduction

Because of the Crown's pre-emption of the health care field through universal hospital and medical care plans, prior to 1970, actuaries have been primarily concerned with income replacement plans in recent years.

Prior to 1930, most of the disability insurance in Canada consisted of a premium waiver benefit or a $10 of monthly income benefit attached to a basic life insurance policy, either individual or group. Here, as abroad, poor claim experience in the 1930's lead to a contraction in coverages offered and a stiffening of underwriting requirements and policy provisions. The post war decade saw renewed interest in disability income coverages, particularly in the group area and the development of coverages independent of life insurance coverages.

In the mid-1960's, the CIA instituted a committee under the Chairmanship of George Dinney of the Great-West Life. This Committee was interested not only in the problem of morbidity cost but also with incidence and continuance costs, particularly the later. This later interest was spurred by a discovery of the odd history of closing continuance tables. The CDT table, which was the most modern table in the 1960's, used experience data for the first two years and then the data from the 1929 tables for the longer durations. Unfortunately, this table's rates beyond 2 years were based on an even earlier table whose construction methods are obscure. Since this very old and obscure data were the primary basis available for the then rapidly expanding long term disability insurance business [characterized by 6, 12 and 24 month elimination periods and coverage to age 65], a lot of people were extremely concerned with developing long term continuance data and it became an important objective of the study to obtain better continuance experience at durations beyond 2 years, particularly, since the Society was not then studying such long duration continuance rates.

Data collection was started in 1972 and was plagued with two particular problems, both related to the source of the data in the companies' files. For most companies, this source was the accounting records which merely recorded the fact that money was paid to a particular individual in a particular year. The problems were to separate the payments for a year into one or more incidents of disability and, secondly, to connect payments made over a two or more years period. Most companies recognized the crucialness of the first problem but had a great deal of difficulty with the second. Unfortunately, the people at the compiling company were very interested in adding contributors and took the attitude "The problem is unimportant. We will take anything". The result was "Garbage in, garbage out" and the study was allowed to languish to the point that data compilation ceased in 1979. At the present time, work is underway to see if this body of data can be salvaged and to correct the attitude and compilation problems which plagued the data compilation.

The Problem

The 1969 Commissions Disability Tables were very limited in the ages (about 9 groups) and elimination period (about 5) presented while the ingenuity of actuaries in developing new elimination periods and age ranges knows few bounds.

To obtain the incidence rates from the 1964 CDT Table, my predecessors as chairmen began with the costs published in the table and divided by the appropriate disabled life annuities to obtain an estimate of the incidence rates. These were extrapolated and interpolated to additional age groups (particularly 65-69 and 70-74) and additional durations.

As part of the job of revitalizing the study, it was necessary to greatly extend the available incidence rates to new elimination periods, particularly, in the range of 90 to 365 days. What I wish to present is the methodology used and indicate the success achieved.

The Underlying Data

Unfortunately, I have been unable to obtain a copy of the actuarial report on the construction of the 1964 CDT. It was published in the Proceedings of the Health Insurance Association of America and as far as I can tell is practically unavailable. If anyone could help, any assistance would be appreciated.

Certain problems exist with the initial data. For example, the rate for an elimination period of 7 days is .11280 REGARDLESS OF AGE. Other evidence of smoothing and tampering with the data are certainly present. This may go a long way towards explaining some of the irregularities found in this investigation.

Traditional Methods

My predecessors had extended the table to age group 70-74 and obtained a few intermediate points. Some preliminary testing showed that at least two and possibly three or four methods had been used. To obtain these figures there seemed to be no consistency (and no one had documented what they had done) and so I decided to leave the data for elimination periods of less than 90 days alone and, secondly, to accept the basic points extrapolated for age groups 65-69 and 70-74.

The traditional actuarial methods had several drawbacks:

(1) The traditional methods were developed for equally spaced data. The best available points seemed to be for elimination periods of 90, 180, 365 and 730 days. Hence, it would have been necessary to develop the divided difference form of these equations.

(2) The points for 365 and 730 day elimination periods would have supplied one half of the data. Since these values were probably based on sparse data and well smoothed, it did not seem desirable to put too much reliance on these points.

(3) It is desirable that the data be consistent by both age groups and elimination period. Trials working with elimination period and checking across age groups seemed to show a lack of consistency i.e. results that were smooth by elimination period were not smooth across age groups.

(4) The values produced seemed to suffer from the defect that they did not decrease sharply enough by length of elimination period.

Each of these factors contributed to a decision to see if a better method could be found.

Methods Explored

The rapid decrease in incidence rates with increase in the elimination period suggested an exponential model. This lead to models of the type $A+Bc^n$. These proved to be disasterous i.e. the parameters did not seem reasonable.

Further exploration lead to the model $\ell n\ q = \ell n\ B + \frac{1}{x+h}\ \ell n\ c$. The values of B, c and h were estimated using the values for 90, 180 and 365 days and checked against the values for 730 days. As can be seen in Table I the error is in the range of 4-6% through to the 60-64 age group when it begins to break down. Yet, this is the point at which someone had extrapolated the data by age group from the younger ages and, hence, at which least confidence could be put in the starting point.

Further support for the method came from examining the values of B, c and h and looking at their pattern by age group. This is shown in Table II.

Omitting the age group 20-24 for the moment, we see a steady increase in the B parameter by age group. Similarly, $\ell n\ c$ shows a pattern of smooth decrease to the age group 60-64, at which point it increases. The pattern for h is similar, except that the minimum is reached at about age group 40-45 after which it increases.

The one jarring point in the nice smooth transition occurs in the 40-44 and 45-49 age groups. The parameter B moves reasonably in this range but the other two parameters

TABLE I

	90	180	365	730 A	730 E
20-24	.00664	.00162	.00075	.00051	.00051
25-29	.00657	.00161	.00074	.00052	.00050
30-34	.00778	.00192	.00091	.00065	.00063
35-39	.00981	.00245	.00119	.00087	.00082
40-44	.01257	.00343	.00172	.00129	.00122
45-49	.01676	.00515	.00283	.00219	.00213
50-54	.02239	.00793	.00463	.00372	.00356
55-59	.03110	.01327	.00842	.00707	.00670
60-64	.04427	.02264	.01491	.01286	.01188
65-69	.06324	.04022	.02865	.02497	.02327
70-74	.09629	.07731	.05789	.05100	.04320

TABLE II

Age Group	$\ell n\ c$	B $(X10^7)$	k
20-24	293.0	3434	9
25-29	301.5	3320	11
30-34	278.0	4299	6
35-39	262.5	5832	3
40-44	257.6	8590	7
45-49	212.0	15900	0
50-54	196.0	27210	3
55-59	171.5	53080	7
60-64	187.5	92800	30
65-69	197.0	181760	68
70-74	693.5	233840	400

do not. Yet, if one looks across the four or six surrounding groups, the suggestion of an irregularity in the underlying data is extremely strong.

For instance the difference table for $\ell n\ c$ is

	Δ	Δ^2
301.5		
	23.5	
278		8
	15.5	
262.5		10.4
	5.1	
257.6		-40.5
	45.6	
212.0		29.6
	16.0	
196.0		

Either the underlying factors shift strongly in the .40's or else the value for the 40-44 age group is considerably off and should be in the range 235-240. Since the value of h is also unusual, it points strongly to an unusual value for one or more of the rates with which I began.

Conclusion

It would appear that the use of estimators for incidence of morbidity rates for durations between 90 and 730 days of the form $\ell n\ q_n = \ell n\ B + \frac{1}{x+k}\ \ell n\ c$.