Classification of a member in terms of healthcare risk

A. de Villiers*
Department of Quantitative Manage, University of South Africa,
PO Box 392, Pretoria 0003, Republic of South Africa
dvilla@unisa.ac.za

C.A. van der Merwe
Bureau for Management Information, University of South Africa,
PO Box 392, Pretoria 0003, Republic of South Africa
vdmerca@unisa.ac.za

T.J. van Wyk Kotze
TheoData cc, 27 Haarlem Road, Stellenberg, Durbanville 7550,
Republic of South Africa
theodata@iafrica.com

Received July 2004

Legislation by government has changed the playing fields in the medical scheme industry in South Africa. Medical schemes can no longer choose their members or discriminate against members who claim more than projected amounts. Only those medical schemes that are able to manage their risk optimally, will ultimately survive.

In the research it was established that the number of chronic beneficiaries in a family is an important risk factor if a member is classified into a normal claim category or an above-normal claim category. The medical schemes should make sure that they have systems in place to manage the health of such beneficiaries holistically. This group of individuals is ideally suited for a preventative healthcare programme.

*To whom all correspondence should be addressed.

Introduction

Healthcare costs continue to rise and this means that employees belonging to medical schemes are likely to suffer the consequences (*Business Day*, 1999b). The cost of healthcare provision to employees continues to rise well above the general inflation rate (*Business Day*, 1999a.) Together with the rising costs, political and social change in South Africa has highlighted the need for healthcare reform (Erasmus, 1998).

The Medical Scheme Act 131 of 1998, which came into effect in January 2000, aims to provide more healthcare coverage for more people. According to the Act, membership of medical schemes is open to anyone, except in the case of restricted membership schemes where particular employers have established schemes exclusively for their employees or members.

According to the Act, contributions must be based on a member's income or the number of his/her dependants or both such income and number of dependants. The contributions apply universally to all members and their dependants. Contributions may only vary in respect of the cover provided. Benefit options are priced differently depending on the level of cover afforded. If the rules of the scheme so provide, children may be covered for a reduced contribution.

The Act also imposes a set of minimum benefits in respect of health services prescribed by the regulations under the Act, and rendered by state hospitals according to clinical protocols and criteria.

Both the medical scheme industry and the short-term insurance industry aim to provide an individual with cover in the case of a loss. In the medical scheme industry this cover may be seen as cover for the loss of health. However, one of the differences between the two industries is that the short-term insurance industry is allowed to discriminate against a member, for example, by putting up premiums when a member claims or through the removal of a no-claim bonus. Different sets of premium tables also exist for the different age and demographic groupings of members.

The Medical Scheme Act has changed the playing fields in the medical scheme industry. Under the Act it is no longer possible for medical schemes to charge different premiums for different groupings of members. The evaluation of member profiles, that is, those of current members and prospective members, will play an important role in product development and risk management by the schemes.

This paper discusses an application of a logistic regression model in an attempt to identify a function that can be used by medical schemes to classify an individual member into a certain claim category or risk category.

The data used in the analysis

Risk management and the modelling and forecasting of claims can only be done successfully by medical schemes if they are in a position to identify and measure the correct information consistently over time. For the research, member and claims data for 1999 and 2000 were obtained from a medical scheme for a group of approximately 34 000 principal members belonging to one benefit option. The benefit option is a traditional option where all the benefits are risk benefits to the medical scheme.

The following data variables were supplied for the two years:

Claims file	**Member file**
Practice number	Date of payment
Member number	Member number
Dependant code	Dependant number
Date of payment	Gender (Male or female)
Date of service	Member type (Active or pensioner)
Receiving date of claim	Date of birth
Tariff code	Registration date at scheme
Claim amount	Date of resignation from scheme
Tariff amount	Total number of dependants
Benefit amount	Total number of adult dependants
Member amount/portion	Total number of children (usually 21 years or younger) dependants
Benefit type (hospital, acute medication, etc.)	The number of beneficiaries on the chronic medication programme
Procedure code	Total premium
Diagnosis code	
Nappi code (medicine classification code)	
Admission date (only for hospital)	
Discharge date (only for hospital)	

In an attempt to identify the important risk factors for the scheme, a logistic regression model was fitted to the data provided by the scheme. The dependant variable was based on the claim premium ratio for the principal member with his/her dependants, and the independent variables were age, sex, number of beneficiaries/dependants, member type (active or pensioner) and the number of beneficiaries using chronic medication.

The claim premium ratio is an important statistic for evaluating the performance of a medical scheme down to the lowest level. It is simply a ratio of the benefits paid out over the premiums that are collected for the same period. A claim premium ratio of 100 percent is an indication that the benefits paid out were equal to the premiums collected. If the claim premium ratio is smaller than 100 percent it means a profit to the medical scheme. If it is greater than 100 percent it means a loss to the medical scheme. Bear in mind that administration costs also form part of the expenditure and a preferable claim premium ratio would always be less than 100 percent.

Manipulation of the data

The claims file only contains members for whom there was a claim during the specific year. If a member did not claim during a year there is no claim record for that member. The member file contains the demographic information on a monthly basis for every principal member. In calculating the claim premium ratio it is necessary to link the two files. The members who did not claim during the year will have a claim premium ratio of 0 percent. (These members and the members with a low claim premium ratio are important for the scheme in terms of cross-subsidisation between members.)

Analysis of the data

In order to identify the high-risk combinations of members it is possible to simply calculate the claim premium ratio for different demographic groupings of members, for example, the claim premium ratio according to an age distribution, member type, number of dependants, etcetera. This was all done, but for this paper a logistic regression model was fitted to the data in an attempt to identify the most important risk factor and also to calculate the probability that a member would fall into a certain claim premium ratio category. In an attempt to measure the effect of age in 10-year intervals, the age of the principal members was divided by 10. *Proc Logistic* of SAS (SAS Institute, 1990) was used for the analysis.

Logistic regression

According to (Freeman, 1987), a logistic regression model predicts the probability of Y occurring given values of the independent variables of X_i , $i = 1,2, \dots n$. The logistic regression equation from which the probability of Y is predicted is given by:

$$P(Y) = \frac{1}{1 + e^{-z}}$$

where

$$Z = \ln\left(\frac{P(Y)}{1 - P(Y)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i$$

The resulting value from the equation is a probability value that varies between 0 and 1. A value close to zero means that *Y* is very unlikely to have occurred and a value close to 1 means *Y* is very likely to have occurred.

A logistic regression was used in an attempt to determine a classification function that could be used to classify a member into a normal or above normal claim category. A claim premium ratio of 98 percent was used to split the group into the two categories. (The claim premium ratio for the total group is 98 percent. The mean claim premium ratio per member is 101 percent with a standard deviation of 178 percent. It is also possible to use an alternative cut-off value, for example, 90 percent.)

The following variables were used in the model:

Dependant variable or outcome

The claim premium ratio was split into two categories.
Y = 0 when the claim premium ratio was above 98 percent and *Y = 1* when the claim premium ratio was 98 percent or less.

Independent variables

Gender (male or female)
Age (age was divided by 10 in an event to estimate the effect of age in increments of 10 years.)
Number of beneficiaries on the chronic programme
Number of dependants
Member type (active or pensioner)

In a stepwise procedure all the predictor variables enter the model and all the individual coefficients are statistically significant on a 1 percent level of significance. It is interesting to note that the number of chronic beneficiaries was the first variable to enter into the model. Using the model, the probability *P(Y)* was calculated for all the members. The model correctly classified 70 percent of the members as normal or above normal claimers. The 70 percent is based on a cut-off or split probability of 0,5. An individual probability of greater than 0,5 is used to predict that the member will be a high claimer and a probability of 0,5 and less that the member will be a normal claimer. With a cut-off probability value of 0,5; 8 505 (24,5%) members were predicted to be normal claimers, but turned out to be above normal claimers. With a cut-off probability value of 0,3; 4 664 (13,4%) members were predicted to be normal claimers, but turned out to be above normal claimers. The group can be seen as a false/negative group and the medical scheme is making a loss on these members. The cut-off probability value plays an important role in the classification of a member and this classification should in practice be determined in consultation with the client. However, the objective should be to try and fit a model that minimises this group of members, which was predicted to be good and turned out to be bad. Two-way tables for the two different cut-off probability values are in Tables 1 and 2.

Table 1: Predicted probability split at 0,5

Predicted	Observed		
Frequency Percent Row Pct Col Pct	**0 (Cp>98%)**	**1 (Cp<=98%)**	**Total**
0 (Bad)	2 292 6,61 54,78 21,23	1 892 5,46 45,22 7,93	4 184 12,07
1 (Good)	**8 505** 24,54 27,91 78,77	21 972 63,39 72,09 92,07	30 477 87,93
Total	10 797 31,15	23 864 68,85	34 661 100,00

Table 2 Predicted probability split at 0,3

Predicted	Observed		
Frequency Percent Row Pct Col Pct	**0 (Cp>98%)**	**1 (Cp<=98%)**	**Total**
0 (Bad)	6 133 17,69 46,13 56,80	7 162 20,66 53,87 30,01	13 295 38,36
1 (Good)	**4 664** 13,46 21,83 43,20	16 702 48,19 78,17 69,99	21 366 61,64
Total	10 797 31,15	23 864 68,85	34 661 100,00

The equation that results from the analysis is

$$P(Y) = \frac{1}{1+e^{-z}}$$

where

$$Z = -1{,}6509 - 0{,}2343X_1 + 0{,}0769X_2 + 0{,}7230X_3 + 0{,}1934X_4 - 0{,}0627X_5$$

X_1 = Gender
X_2 = Age
X_3 = Number of chronic beneficiaries
X_4 = Number of dependants
X_5 = Member type … (1)

X_1 is equal to 1 when the member is male and –1 when the member is female.

X_5 is equal to 1 when the member is an active member and –1 when the member is a pensioner.

The number of beneficiaries using chronic medication entered the model first and this has the largest positive coefficient of 0,723 followed by gender with a negative coefficient of –0,2343. An increase in the number of chronic beneficiaries will increase the probability that such a member will be classified into an above normal claim category (positive coefficient). For gender, X_1 is equal to 1 when the member is male and –1 when female. A male member will lower the probability of a member being classified into an above normal claim category and a female will increase the probability of a member being classified into an above normal claim category. It is interesting to note that an increase of 10 years in the age will have less effect on a member being classified into an above normal claim category than the number of chronic beneficiaries. (The other coefficients could be interpreted in the same way.)

Using equation (1) it is possible to calculate various probabilities for different types of member, for example:

Example 1: A young and probably healthy member

Age	24 years
Sex	Male
Member type	Active
Number of dependants	0 (single member)
Number of beneficiaries on the chronic programme	0
P(Y) = 0,14	

For such a single, active male member of 24 years not using chronic medication there is only a 14 percent chance that the member's claim premium ratio will be more than 98 percent.

Example 2: An older and probably not-so-healthy member

Age:	65 years
Sex:	Female
Member type:	Pensioner
Number of dependants:	1 (Member plus one)
Number of beneficiaries on the chronic programme:	2
P(Y) = 0,69	

For such a female pensioner member of 65 years of age with one dependant, both of whom are using chronic medication, there is a 69 percent chance that the member's claim premium ratio will be more than 98 percent.

The number of chronic beneficiaries was not only the first variable to enter into the model; it was also calculated that with one additional chronic beneficiary the chances (odds ratio estimate of 2,061) of a member having a claim premium ratio of more than 98 percent were doubled.

The claim premium ratio, mean benefit amount per member for the total benefits and proportional utilisation of chronic medication for the different groups (number of beneficiaries using chronic medication) are:

Number of beneficiaries using chronic medication	**Claim premium ratio (%)**	**Mean benefit amount per member**	**Chronic medication as a percentage of the total benefits (%)**
0	77	R 9 687,50	0,0
1	140	R16 660,85	10,3
2	151	R23 186,24	11,8
3 and more	151	R25 517,40	12,0

The claim premium ratio for members with no beneficiaries using chronic medication is 77 percent, compared with 140 percent for members with one beneficiary using chronic medication and 151 percent for members with two or more beneficiaries using chronic medication. Chronic medication may be seen as an additional benefit that a member has to apply for and the utilisation thereof is between 10,3 percent and 12 percent of the total benefits utilised by such members. However, this proportional utilisation is not in line with the increase in the claim premium ratio and the mean benefit amount per member.

One could argue that the number of dependants per member distorts the results above. However, the utilisation for the individual disciplines also differs for the majority of disciplines for members with the same number of dependants. This is evident in the following comparison of the mean benefit amount per member for members with zero and one beneficiaries using chronic medication. In both cases the members have two dependants (member plus one dependant).

Discipline	**Number of beneficiaries using chronic medication**	
	0	**1**
General practitioners	R1 300	R1 726
Medical specialists	R1 342	R2 264
Hospital	R3 266	R6 723
Pathology	R 499	R 823
Radiology	R 479	R 893
Optometry	R 202	R 306

From the results it seems that the members with beneficiaries using chronic medication not only utilise chronic medication, but that their utilisation of overall benefits is higher than members with no chronic beneficiaries on the chronic programme.

Conclusion

In the study it was established that the number of chronic beneficiaries in a family is an important risk factor if a member is classified into a normal claim category or above-normal claim category. A function was also determined by which it is possible to calculate the probability that a member will fall into a certain claim category. Such a classification function may be helpful for a medical scheme in determining a risk profile for their members, especially new members joining the scheme.

With the logistic regression model shown in this article, 70 percent of members are classified correctly. This may seem too low, but with more demographic variables available, for example economic sector and geographic location, it may be possible to improve the fit. With the necessary information it may also be possible to extend the use of logistic regression and other classification functions to individual beneficiaries and disciplines, for example, general practitioner and optometry. In such instances, the age and gender of the beneficiaries will probably be important predictor variables, especially if one considers the benefit utilisation over the life span of an individual. Lately the majority of contribution tables have been priced per beneficiary type (principal member, adult dependant and child dependant). With contribution tables per beneficiary type it would be possible to apply a logistic regression to the claim premium ratio per beneficiary and not per principal member with his/her dependants (family) as it has been shown in this article.

The medical schemes should make sure that they have systems in place to manage the health of beneficiaries using chronic medication holistically. The utilisation of benefits by these individuals does not stop with the utilisation of chronic benefits only. Management should go beyond a chronic programme with a list of approved conditions and a list of approved medicines. When a beneficiary applies to enter the chronic programme of a medical scheme, the medical scheme usually captures a great deal of detailed information on that applicant. This group of individuals is ideally suited for a preventative healthcare programme.

References

Freeman, D.H. 1987. *Applied categorical data analysis*. New York: Marcel Dekker.

Business Day, 1999a. ‘Cost of providing health care rising’, 2 September:3.

Business Day, 1999b. ‘Employees footing the bill’, 3-5 September:8.

Erasmus, B.M. 1998. ‘An employer perspective of managed health care’. MBA thesis, University of the Witwatersrand, Johannesburg.

SAS Institute. 1990. *SAS/STAT user’s guide, version 6*. Volume 2. 4th Edition. Cary, NC: SAS Institute.