International Journal of Engineering & Technology, 8 (1.7) (2019) 56-61

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Proportional Odds Model for Health States Analysis

Shamshimah Samsuddin[1, 2]*, Noriszura Ismail [1]

[1]*School of Mathematical Sciences, Faculty of Sciences and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.*

[2]*Centre for Actuarial Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

**Corresponding Author Email: shamshimah@tmsk.uitm.edu.my*

Abstract

Employees' health status is one of the key issues that should be considered in ensuring the economic growth of a country. Information on employees' health status is useful in social and economic studies, especially in issues related to work-related disabilities and deaths. The estimation of disability probability involves a challenging method as a disabled employee may move from one state to another (from temporary to permanent, or vice versa), or from one event to another (from disabled to active, or to death), implying that repeated responses may be obtained at different time points in the relevant longitudinal studies. Markov Chain Model can be used to analyze repeated measurements, or ordinal responses in a longitudinal data, and to compare between one health states to another. The main objective of this study is to estimate the Markov transition probabilities between health states using the Proportional Odds Model (POM) based on the dataset obtained from Social Security Organization, Malaysia (SOCSO). The results show that female employees in age group 55-59 have the highest probability of remaining in active state (A), while male employees in age group 15-19 have the lowest probability of remaining in active state (A), or have the highest risk of transitioning from healthy state to disability or death states.

Keywords: *health status, disability, repeated responses, Markov Chain Model, Proportional Odds Model*

1. Introduction

Healthy and productive workers is an important factor that should be considered in ensuring the economic growth of a country. Work-related injuries and disabilities have become an increasing burden to workers' compensation because the estimation of transition probabilities concerning disability involves a challenging method. A disabled employee may move from one state to another (from temporary to permanent, or vice versa), or from one event to another (from disabled to active, or to death), implying that repeated responses may be obtained at different time points in the relevant longitudinal studies.

Health-related and social science studies often involved longitudinal or panel data where repeated ordinal responses commonly occur, [1]. For this type of data, a sequence of ordinal responses for the same individual is considered, and the main focus is to estimate the degree of movement, or the transition, between consecutive time points. Several models were suggested to analyze simultaneously the effect of explanatory variables on a response variable, such as multiple linear regression for continuous response, logistic regression for categorical response, and Cox's proportional hazards for censored response. All of these models are categorized under multiple linear regression [2, 3].

Markov Chain Model (MCM) is often used for repeated responses where the effect of previous responses on the current response is taken into account. Many authors have used Markov models to analyze the problems. As examples, [4] using Markov Chain to modified the current No Claim Discount system in Malaysia, [5] fitted the various orders to the daily rainfall and [6] to determine the occurrence of polluted and non-polluted for air pollution. Several methods have been suggested to estimate the Markov transition probabilities for ordinal response data. [7] proposed the Proportional Odds Model (POM) for analyzing categorical outcomes from quality-of-life measures which are subjected to non-ignorable missingness, and [1] also suggested the POM to analyze complete and incomplete longitudinal ordinal responses. An application of the MCM for disability data can be found in [8] who estimated the Markov transition probabilities using several methods such as the counting method, the ordered logit regression, and the ordered probit regression.

In this paper, we model the transition probabilities between health states using a MCM based on the dataset obtained from the Employment Injury Scheme (EIS) which is covered under the Social Security Organization, Malaysia (SOCSO). The Markov transition probabilities are estimated using the POM, which is a popular method for ordinal responses [9].

Of all existing regression models, POM requires the proportional assumption regarding the nature of relationship between the response variable and the prediction factors [10]. The result of a POM can be misleading, or have no meaning at all, if the proportional odds assumption is not fulfilled. The POM is a sequential model that is used for ordinal responses, and is also known as the cumulative logit model [11, 12]. According to [11], the POM and the Partial POM are the best models for sequential ordinal responses compared to other logistic regression models due to the nature of data, and the easy interpretation of the results. The POM is also recommended to medical researchers by [2] as it helps doctors to easily understand the results of ordinal logistic responses. In our study, the estimates obtained

from the POM enable us to determine the age group and gender that has a negative (or positive) relationship with the transition probabilities, and thus, allowing us to identify the group of workers that have a more (or less) risky health status.
In the next section, information on the employees' injury data from SOCSO are provided using a descriptive analysis. The methodology for Markov transition model is also presented in the same section. The results are discussed in Section 3, and the conclusions are provided in Section 4.

2. Data

The sample data is obtained from SOCSO, which is an organization in Malaysia that acts as a safety net for an employee who becomes unemployed, temporarily disabled, permanently disabled, or dead. The data is based on secondary information of the EIS under SOCSO, covering the years of 2009 until 2013. The EIS is a suitable dataset for the construction of health transition probabilities as it reports four possible outcomes at yearly intervals for the health status: 1. Active/work (A), 2. Temporary disability (TD), 3. Permanent disability (PD), and 4. Death (D). The health status represent the levels of disabilities based on the decisions of panel doctors appointed by SOCSO [13]. A more detailed discussion on the issues of disability among workers who contribute under SOCSO can be found in [14, 15].
The demographic characteristics of claimants at their entering states in year 2009 (or at T_0) are shown in Table 1. The descriptive analysis show that a total of 263,971 contributors under the EIS submitted their claims in years 2009-2013, and can be categorized into gender and age groups. The majority of claimants are male employees (212,416, or 80% of the total claimants), while the remaining are female employees. The average age for male claimants is 34 with a standard deviation of 11.65, while the average age of female claimants is 35 with a standard deviation of 11.59.

Table 1: Descriptive Statistics at Entering States, 2009 (T_0)

N:263,971						
Gender		Male	212,416	Gender	Female	51,555
Age		Min	34	Age	Min	35
		Median	33		Median	34
		Mode	23		Mode	21
	Std.dev.		11.65		Std.dev.	11.59

Table 2: Number of Claimants in Each Year (Age: 15-65)

	State	Year				
		2009	2010	2011	2012	2013
M	1	169,609	166,825	163,869	162,521	161,526
	2	31,329	31,881	33,014	33,338	33,764
	3	11,478	12,683	13,424	13,609	13,443
	4	-	1,027	1,082	839	735
F	1	42,188	41,330	40,484	39,716	39,517
	2	7,193	7,674	8,224	8,730	8,981
	3	2,174	2,448	2,640	2,805	2,664
	4	-	103	104	97	89

Table 2 shows the number of claimants according to health status and gender in years 2009 to 2013. State 1 (A) is the state where the employees do not make any reports of accidents at work or death. For state 2 (TD), the number of claimants increase from year to year, increasing from 31,329 in year 2009 to 33,764 in year 2013 for male employees, and from 7,193 in year 2009 to 8,981 in year 2013 for female employees. The number of claimants for state 3 (PD) also increases throughout the years, except for female employees whose number declines in year 2013. For state 4 (D), the number of deaths for both male and female employees decreases from years 2010 to 2013. It should be noted that information on the number of deaths in year 2009 is not available as this year is considered as the beginning year of the study.

3. Methodology

Our study is based on a longitudinal data where the transitional frequencies are classified into an ordered categorical response variable. The responses are obtained from repeated measures at different time points, and take the values of a set of several ordered categories (values of 1 until 4). The covariates are discrete variables (age group and gender), and their values stratify the population, where each population is homogeneous and is subject to the same temporal changes.
We use Markov chain model (MCM) to estimate the transition probabilities between health states as time or age progresses. Let

$$P_{ij} = P(X_{n+1} = j \,|\, X_n = i) \tag{1}$$

be the one-step transition probability of a response being in state j at time $n+1$, given that it is in state i at time n. Since the transition probabilities are annual probabilities, the values satisfy the following conditions: $P_{ij} \geq 0$ and $\sum_{j=0}^{\infty} P_{ij} = 1$ for $i, j \in \Omega$. The fundamental assumption for the one-step transition probability in this study is that the health state of an individual in a specific year depends only on the health state in the previous year. Therefore, the Markov transition probabilities are easily inferred because the complete sequences of health states from years 2009 to 2013 are observed and known for the entire population. In other words, the transition probabilities can be obtained by observing the changes of health status of each employee in each pair of years (2009-2010, 2010-2011, 2011-2012, 2012-2013) [16].

The Markov transition probabilities are estimated using the Proportional Odds Model (POM). Suppose a response variable takes the ordinal values of 1 until J. The cumulative probabilities can be defined as:

$$\begin{aligned}\Pr(y \le j) &= \Pr(y=1)+\ldots \Pr(y=j)\\ &= p_1+\ldots+p_j\\ &= w_j\end{aligned} \tag{2}$$

Assuming logistic regression, the cumulative probability, w_j, can be parameterized as:

$$\log\left(\frac{w_j}{1-w_j}\right)=\alpha_j+x'\beta,\ \ j=1,2,\ldots J-1 \tag{3}$$

where x is the vector of explanatory variables, β is the vector of regression coefficients, J is the maximum ordinal value, and $\alpha_1<\alpha_2<\ldots<\alpha_{J-1}$ are the intercepts. The POM is also known as a grouped continuous model, as it groups the observations into intervals with cut-points $\alpha_j, (j=1,\ldots,J-1)$. Therefore, the transition probability of being in state j, p_j, can be obtained using:

$$p_j = w_j - w_{j-1} = \frac{e^{\alpha_j+x'\beta}}{1+e^{\alpha_j+x'\beta}} - \frac{e^{\alpha_{j-1}+x'\beta}}{1+e^{\alpha_{j-1}+x'\beta}} \tag{4}$$

4. Results and Discussion

In this study, the parameter estimates of POM are obtained using R software, and the results are shown in Table 3. The estimates in Table 3 can be used to obtain the transition probabilities of being in state 1 (A), state 2 (TD), state 3 (PD), and state 4 (D), given that it is previously in state 1 (A), conditioned on two covariates (age group and gender). The transition probabilities of being in states 1,2,3, and 4, given that it is previously in state 2, or in state 3, are also estimated, but not shown here.

The log odds for being in state 1 when the predictor variables (covariates) are evaluated at zero, is shown by intercept 1 (α_1). The log odds for a male employee in age group 15-19 is 0.8062, and the odds is $\frac{e^{0.8062}}{1+e^{0.8062}}=0.6913$. The log odds for being in state 1 and state 2 (shown by intercept 2, α_2), for a male employee in age group 15-19, is 2.2749. Therefore, the odds for being in state 2 is $\frac{e^{2.2749}}{1+e^{2.2749}}-\frac{e^{0.8062}}{1+e^{0.8062}}=0.2155$. The log odds for being in state 1, state 2 and state 3 (shown by intercept 3, α_3), for a male employee in age group 15-19, is 5.15. Therefore, the odds for being in state 3 is $\frac{e^{5.1570}}{1+e^{5.1570}}-\frac{e^{2.2749}}{1+e^{2.2749}}=0.0875$. Finally, the odds for being in state 4 is $1-\frac{e^{5.1570}}{1+e^{5.1570}}=0.0057$. It should be noted that the odds of being in states 1, 2, 3 and 4, given that it is previously in state 2, or in state 3, can be obtained in a similar manner.

It can be seen that the response variable (health state) is treated as an ordinal under the assumption that the state levels have a natural ordering (from low to high). The regression estimates in Table 3 shows the log odds of one unit increases in the age group (or gender) on the health state, given that other variables are held constant. As an example, a subject in age group 20-24 is estimated to have 0.12 unit increases in the log odds while other variables are held constant. The comparison between male and female ($male=0$, $female=1$) shows that there is 0.05 unit increases in the log odds of a female employee, given that other variables are held constant.

The regression estimates for age group and gender in Table 3 indicates that female employees in age group 55-59 have the highest probability of remaining in active state after 5 years (at T_1) compared to other groups, as shown by the higher values (0.05 for female, and 0.29 for age 55-59). On the contrary, male employees in age group 15-19 have the lowest probability of remaining in active state, or equivalently, have the highest probability (or highest risk) of transitioning from healthy state to disability or death states after 5 years.

The estimates of transition probabilities are shown in transition matrices in Tables 4-5, which are arranged according to age group and level of disability starting at any state, with the exception of an absorbing state (death). As an example, for a male employee in age group 20-24 who is initially healthy (state 1), the transition probability for remaining healthy (state 1) after five years (at T_1) is 0.6913, for moving to TD (state 2) is 0.2155, for transitioning to PD (state 3) is 0.0875, and for being in death state (state 4) is 0.057.

Table 3: Result for transition model (from state $i=1$ to states $j=1,2,3,4$)

Parameter	Est.	S.E
Intercept 1, α_1	0.8062	0.0122
Intercept 2, α_2	2.2749	0.0126
Intercept 3, α_3	5.1570	0.0202
Age		
20-24	0.1212	0.0136
25-29	0.2099	0.0136

30-34	0.1886	0.0138
35-39	0.1804	0.0143
40-44	0.1567	0.0141
45-49	0.1658	0.0143
50-54	0.1418	0.0148
55-59	0.2935	0.0171
60-64	0.2210	0.0214
65-69	0.2205	0.0560
Sex		
F	0.0494	0.0060
M	0.0000	0.0000
Log-likelihood		-606018
Degrees of freedom		250

The results in Table 4 indicates that a male employee has a higher probability of transitioning from PD to A at all age groups. For women employees (Table 5), the age groups of 25-29, 35-49 and 40-44 have the highest probability of transitioning from TD to A compared to other age groups. When observing the risks of moving from one state to another, it can be broadly concluded that age and gender are important factors in estimating probability of transition besides level of disability.

5. Conclusion

In this study, we have developed the Markov Chain Model (MCM) for the probability of transition between health states using the Proportional Odds Model (POM), which is a convenient and easy method to model ordered categorical outcomes. The estimates obtained from the POM enable us to determine the age group and gender that has a negative (or positive) relationship with the transition probabilities, and thus, allowing us to identify the group of workers that are exposed to a more (or less) risky health status. The parameter estimates were obtained using R software based on the Employment Injury Scheme (EIS) data from SOCSO, Malaysia. The results indicate that female employees in age group 55-59 have the highest probability of remaining in active state, while male employees in age group 15-19 have the lowest probability of remaining in active state, or have the highest risk of transitioning from healthy state to disability or death states after 5 years (at T_1). For future study, the estimated transition probabilities will be used for projecting the health expectancy and healthcare cost of disabilities and deaths among workers for a longer time horizon and a comparison model can be made in estimating the transition probabilities. As examples, a study conducted by [17] and [18] which comparing between logistic regression, artificial neural network, and Neuro-fuzzy to predict students' academic achievement and survival of cardiac surgery patients accordingly.

Table 4: Probabilities of transitions for male

Age	Health at T_0	Health at T_1			
		1	2	3	4
15-19	1	0.6913	0.2155	0.0875	0.0057
	2	0.9604	0.0316	0.0076	0.0004
	3	0.9771	0.0117	0.0109	0.0002
	4	0.0000	0.0000	0.0000	1.0000
20-24	1	0.7165	0.2000	0.0784	0.0051
	2	0.9579	0.0336	0.0081	0.0004
	3	0.9617	0.0195	0.0184	0.0003
	4	0.0000	0.0000	0.0000	1.0000
25-29	1	0.7342	0.1889	0.0723	0.0046
	2	0.9550	0.0359	0.0087	0.0005
	3	0.9604	0.0201	0.0191	0.0004
	4	0.0000	0.0000	0.0000	1.0000
30-34	1	0.7300	0.1915	0.0737	0.0047
	2	0.9443	0.0443	0.0108	0.0006
	3	0.9548	0.0229	0.0218	0.0004
	4	0.0000	0.0000	0.0000	1.0000
35-39	1	0.7284	0.1925	0.0743	0.0048
	2	0.9425	0.0457	0.0112	0.0006
	3	0.9460	0.0273	0.0262	0.0005
	4	0.0000	0.0000	0.0000	1.0000
40-44	1	0.7237	0.1955	0.0759	0.0049
	2	0.9370	0.0500	0.0123	0.0007
	3	0.9483	0.0262	0.0251	0.0005
	4	0.0000	0.0000	0.0000	1.0000
45-49	1	0.7255	0.1944	0.0753	0.0049
	2	0.9329	0.0532	0.0132	0.0007
	3	0.9541	0.0233	0.0222	0.0004
	4	0.0000	0.0000	0.0000	1.0000
50-54	1	0.7207	0.1974	0.0769	0.0050
	2	0.9276	0.0574	0.0143	0.0008
	3	0.9457	0.0274	0.0264	0.0005
	4	0.0000	0.0000	0.0000	1.0000
55-59	1	0.7502	0.1786	0.0669	0.0043

	2	0.9395	0.0481	0.0118	0.0006
	3	0.9626	0.0190	0.0180	0.0003
	4	0.0000	0.0000	0.0000	1.0000
60-64	1	0.7364	0.1875	0.0715	0.0046
	2	0.9397	0.0479	0.0118	0.0006
	3	0.9750	0.0128	0.0119	0.0002
	4	0.0000	0.0000	0.0000	1.0000
65+	1	0.7363	0.1875	0.0716	0.0046
	2	0.9592	0.0325	0.0078	0.0004
	3	0.9708	0.0149	0.0140	0.0003
	4	0.0000	0.0000	0.0000	1.0000

Acknowledgements

Special thanks go to Universiti Teknologi MARA, Universiti Kebangsaan Malaysia, and Social Security Organization (SOCSO) for supporting this research. The authors gratefully acknowledge the financial support received in the form of research grants (FRGS/1/2015/SG04/UKM/2/2 and GUP-2017-011) from the Ministry of Higher Education (MOHE).

Table 4: Probabilities of transitions for female

Age	Health at T_0	Health at T_1			
		1	2	3	4
15-19	1	0.7018	0.2091	0.0837	0.0055
	2	0.9814	0.0149	0.0035	0.0002
	3	0.9848	0.0078	0.0072	0.0001
	4	0.0000	0.0000	0.0000	1.0000
20-24	1	0.7265	0.1938	0.0749	0.0048
	2	0.9802	0.0159	0.0037	0.0002
	3	0.9745	0.0131	0.0122	0.0002
	4	0.0000	0.0000	0.0000	1.0000
25-29	1	0.7438	0.1828	0.0691	0.0044
	2	0.9788	0.0170	0.0040	0.0002
	3	0.9736	0.0135	0.0126	0.0002
	4	0.0000	0.0000	0.0000	1.0000
30-34	1	0.7397	0.1854	0.0704	0.0045
	2	0.9736	0.0211	0.0050	0.0003
	3	0.9698	0.0154	0.0145	0.0003
	4	0.0000	0.0000	0.0000	1.0000
35-39	1	0.7381	0.1864	0.0710	0.0046
	2	0.9727	0.0218	0.0052	0.0003
	3	0.9638	0.0185	0.0174	0.0003
	4	0.0000	0.0000	0.0000	1.0000
40-44	1	0.7335	0.1893	0.0725	0.0047
	2	0.9701	0.0239	0.0057	0.0003
	3	0.9653	0.0177	0.0167	0.0003
	4	0.0000	0.0000	0.0000	1.0000
45-49	1	0.7353	0.1882	0.0719	0.0046
	2	0.9680	0.0256	0.0061	0.0003
	3	0.9693	0.0157	0.0147	0.0003
	4	0.0000	0.0000	0.0000	1.0000
50-54	1	0.7306	0.1912	0.0735	0.0047
	2	0.9654	0.0276	0.0066	0.0004
	3	0.9636	0.0186	0.0175	0.0003
	4	0.0000	0.0000	0.0000	1.0000
55-59	1	0.7594	0.1726	0.0639	0.0041
	2	0.9713	0.0230	0.0055	0.0003
	3	0.9751	0.0128	0.0119	0.0002
	4	0.0000	0.0000	0.0000	1.0000
60-64	1	0.7459	0.1814	0.0684	0.0044
	2	0.9714	0.0229	0.0054	0.0003
	3	0.9834	0.0085	0.0079	0.0001
	4	0.0000	0.0000	0.0000	1.0000
65+	1	0.7458	0.1815	0.0684	0.0044
	2	0.9809	0.0153	0.0036	0.0002
	3	0.9806	0.0100	0.0093	0.0002
	4	0.0000	0.0000	0.0000	1.0000

References

[1] Ganjali, M. *Fitting Transition Models to Longitudinal Ordinal Response Data Using Available Software*. In *Proceedings of The 8th International ConferenceoOn Teaching Statistics (Icots8)*. 2010.

[2] Bender, R. And U. Grouven, *Ordinal Logistic Regression In Medical Research.* Journal of The Royal College of Physicians of London, 1997. 31(5): P. 546-551.

[3] Agresti, A. And M. Kateri, *Categorical Data Analysis*, In *International Encyclopedia of Statistical Science*. 2011, Springer. P. 206-208.
[4] Manan, N.B., H. Hashim, And M.A. Mohd. *Modification of The Current Malaysian No-Claim Discount System Using Markov Chains*. In *Business Engineering and Industrial Applications Colloquium (Beiac), 2013 Ieee*. 2013. IEEE.
[5] Zakaria, N.Z. And S.M. Deni, *Application Of Alternative Geometric Distribution and Markov Chain Models For Fitting Sequences of Wet and Dry Days in Peninsular Malaysia*. International Journal of Engineering and Management Research (Ijemr), 2016. 6(1): P. 110-119.
[6] Mohamad, N., S. Deni, And A. Ul-Saufie, *Application of The First Order of Markov Chain Model in Describing The PM10 Occurrences in Shah Alam And Jerantut, Malaysia*. Pertanika Journal of Science & Technology, 2018. 26(1).
[7] Cole, B.F., et al., *A Multistate Markov Chain Model for Longitudinal, Categorical Quality-Of-Life Data Subject to Non-Ignorable Missingness*. Statistics in Medicine, 2005. 24(15): P. 2317-2334.
[8] Jung, J., *Estimating Markov Transition Probabilities Between Health States in The HRS Dataset*. Indiana University, 2006.
[9] Mccullagh, P., *Regression Models for Ordinal Data*. Journal of The Royal Statistical Society. Series B (Methodological), 1980: P. 109-142.
[10] Steyerberg, E.W. And M.J. Eijkemans, *Prognostic Modeling With Logistic Regression Analysis*. Network, 2000. 10: P. 11.
[11] Abreu, M.N.S., A.L. Siqueira, and W.T. Caiaffa, *Ordinal Logistic Regression in Epidemiological Studies*. Revista De Saude Publica, 2009. 43(1): P. 183-194.
[12] Abreu, M.N.S., Et Al., *Ordinal Logistic Regression Models: Application in Quality of Life Studies*. Cadernos De Saúde Pública, 2008. 24: P. S581-S591.
[13] Malaysia, *Akta Keselamatan Sosial Pekerja 1969*, 1969.
[14] Samsuddin, S. And N. Ismail. *Multi-State Markov Model for Disability: A Case of Malaysia Social Security (Socso)*. In *Innovations Through Mathematical and Statistical Research: Proceedings of the 2nd International Conference On Mathematical Sciences and Statistics (ICMSS2016)*. 2016. AIP Publishing.
[15] Samsuddin, S. And N. Ismail, *Isu Hilang Upaya Dikalangan Pencarum Perkeso di Malaysia*. Malaysia Labour Review, 2015. 11(No. 2): P. 85-94.
[16] Samsuddin, S. And N. Ismail. *Transition Probabilities of Health States for Workers in Malaysia Using A Markov Chain Model*. In AIP *Conference Proceedings*. 2017. AIP Publishing.
[17] Rusli, N.M., Z. Ibrahim, And R.M. Janor. *Predicting Students' Academic Achievement: Comparison Between Logistic Regression, Artificial Neural Network, And Neuro-Fuzzy*. In *Information Technology, 2008. ITSIM2008. International Symposium On*. 2008. IEEE.
[18] Rahman, H.A.A., et al. *Comparison Of Predictive Models to Predict Survival of Cardiac Surgery Patients*. In *Statistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference On*. 2012.IEEE.