Hyeoun-Ae Park*

Simulation of A Population-based Model of Coronary Heart Disease Morbidity and Mortality

Ⅰ. Introduction

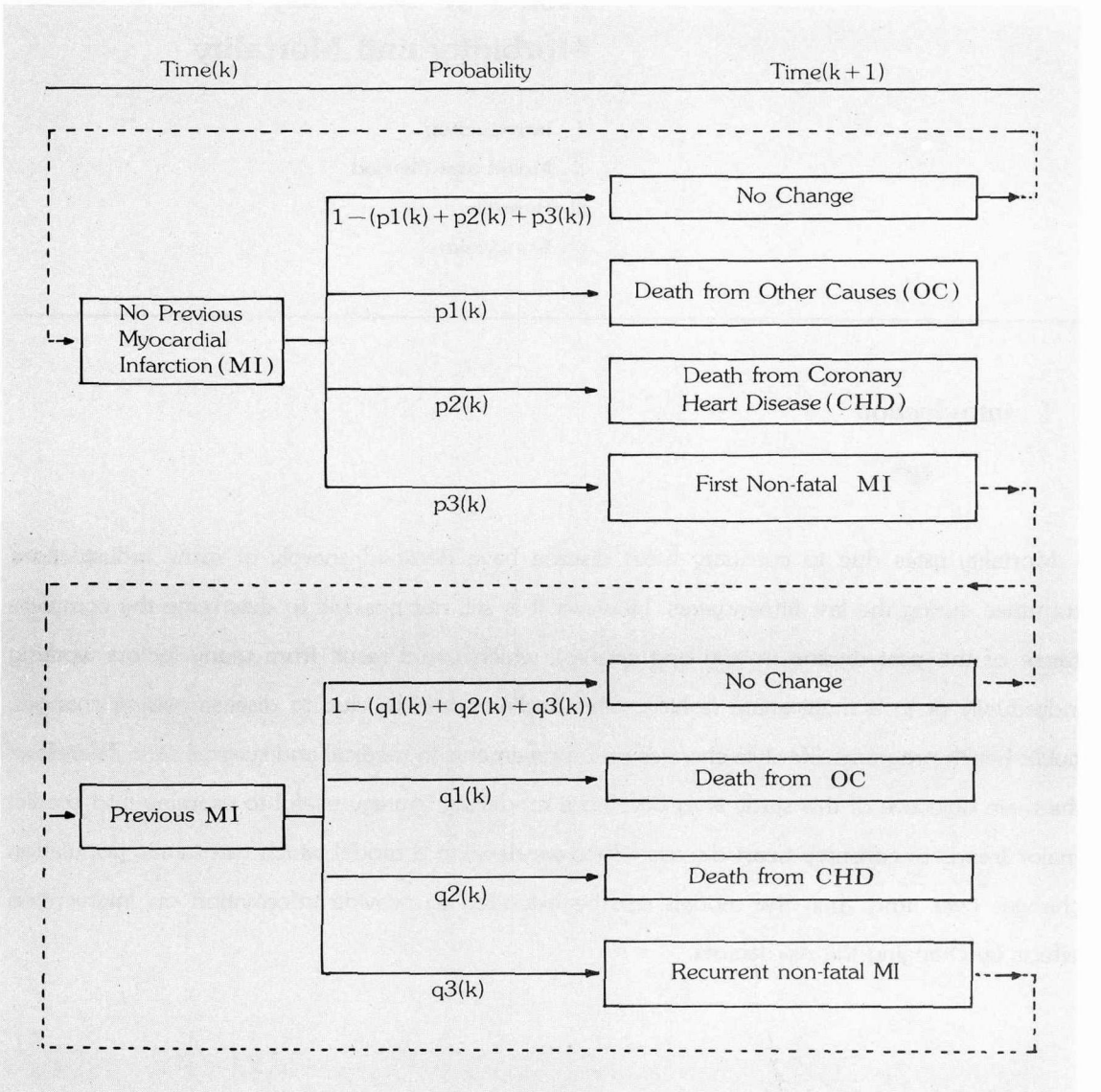
Mortality rates due to coronary heart disease have declined sharply in some industrialized countries during the last fifteen years. However it is still not possible to determine the complete cause of the past decline in any one country, which could result from many factors working individually or in a multivariate fashion. This decline could be due to disease coding changes, public health programs, life style changes or improvements in medical and surgical care. Therefore, the main objective of this study is to develop a modelling strategy useful to examine and predict major trends in coronary heart disease. Once we develop a model which can mimic population changes over time, then the models can be extended to provide information on intervention effects by changing the risk factors.

Ⅱ. Model and Method

A reduced model of coronary heart disease is shown in Figure 1.
At the beginning of the kth year, each subject in a cohort will have had either a previous myocardial

* Senior Researcher, Korea Institute for Population and Health

Figure 1. Reduced Model for Coronary Heart Disease



infarction (MI) or no previous MI. At the end of the *k*th year, the outcomes for those subjects without a previous MI may be death from other causes with the probability p1, death from Coronary Heart Disease (CHD) with the probability p2, a first non-fatal MI with the probability p3 or no change with the probability $1-(p1+p2+p3)$. Outcomes for those who have had a previous MI are similar except that there will be a recurrent non-fatal MI instead of a first non-

fatal MI. For those subjects with a previous MI, we assumed that their transition probabilities are independent from the number of previous MIs. Thus, two distinct types of conditional probabilities are required in the analyses depending on the previous history of MI. If the individual with no previous MI has no change then the individual will stay in the "No Previous MI" pool at the beginning of the next time interval. If the individual with no previous MI has a first non-fatal MI, then the individual will go to the "Previous MI" pool at the beginning of the next time interval. If the individual with a previous MI has nochange or a recurrent non-fatal MI, then the individual will stay in the "Previous MI" pool at the beginning of the next time interval. We used a year as the unit time interval.

Traditionally it is considered that age, gender, serum cholesterol level, diastolic blood pressure and cigarette smoking are the major risk factors in populations which are associated with the morbidity and the mortality from coronary heart disease.

A data set from the North Karelia, Finland was available for this study. A cohort of eastern Finnish men aged 25-59 years in 1972 was identified and examined as part of the North Karelia project. During the examination, blood pressure, serum cholesterol level, and number of cigarettes smoked per day were recorded. Participants' names were linked to the Finnish death certificate register and the hospital admission records for each year from 1972 through 1978. Figure 2 provides a summary of descriptive statistics of the risk factors measured at 1972 and the number of deaths over 8 years from 1972 through 1978. In general this population had higher cholesterol levels and diastolic blood pressures than the U.S. Framingham population at that time. The Finnish population is very famous for high incidence of and mortality from heart disease. This is the population used for the simulation.

In order to incorporate the information about the risk factors of each individual and more than two endpoints in the model, a polychotomous logistic function was utilized to determine the risk of having an event for an individual i in a population. Risk of the dth endpoint for an individual i at the kth time interval can be written as the following.

$$Pd(k) = \frac{1}{1+\exp(1-f(\underset{\sim}{xi}, \underset{\sim}{\beta d}(k)))}$$

where $f(\underset{\sim}{xi}, \underset{\sim}{\beta d}(k))$

$$= \beta d0(k) + \beta d1(k) \times i1 + \ ,...,\ + \beta d4(k) \times i4$$

$$(d = 1,2,3) \qquad (i = 1,...,n)$$

Figure 2. Descriptive Statistics of Male Population of North Karelia, Finland, 1972-1978.

	Death from CHD	Death from OC	Alive	Total
Chol. at 1972 (Mean±SD)	300.5±59.1	273.5±48.5	261.9±50.3	263.2±50.7
DBP at 1972 (Mean±SD)	98.6±14.6	94.8±14.1	91.2±11.9	91.5±12.1
% of Smokers at 1972	67.6	65.1	49.7	50.6
N =	102	218	5101	5421

where $xi1$ = serum cholesterol level (mg/dl)

$xi2$ = diastolic blood pressure (mmHg)

$xi3$ = 1 for 1 to 15 cigarettes per day

0 for otherwise

$xi4$ = 1 for more than 15 cigarettes per day

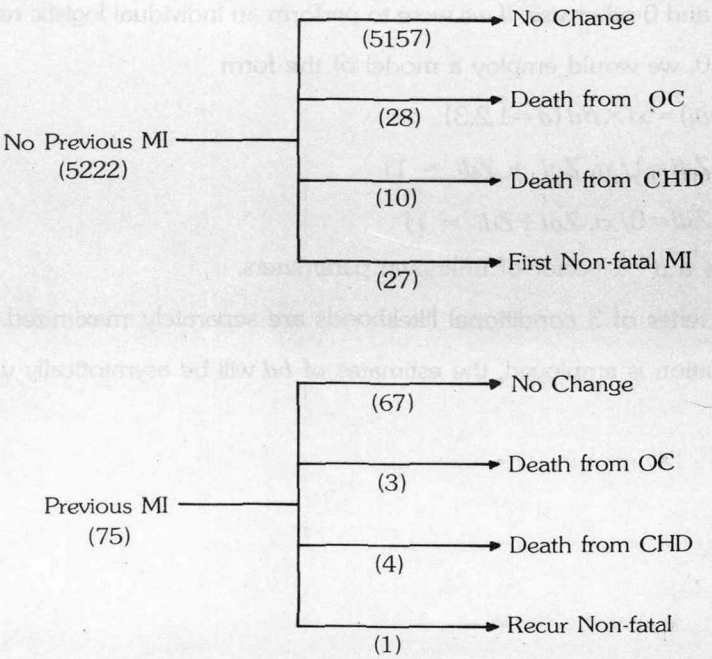
0 for otherwise.

Note that we categorized the smoking into three groups ; non-smokers, below the Finnish average, and above the Finnish average.

There are some practical difficulties in estimating polychotomous logistic regression coefficients. The first is computer storage as a result of the potentially large number of parameters. In our case the number of parameters to be estimated would be a 3*5 matrix for each time interval. Secondly, only one of the major computer packages, SAS, has the polychotomous logistic regression available, the others being restricted to at most simple dichotomous logistic regression. Thirdly, the SAS package does not accomodate sparse data. Figure 3 shows the actual number of events in 1975 in the Finnish population. As you can see, the events are very few. For example, 10 deaths from CHD occured with a denominator of over 5000, and one non-fatal MI has a demominator of 75.

For these reasons, as well as for general analytic simplicity and flexibility, we utilized a series

Figure 3. Actual Transition in North Karelia, Male Population during Year 1975.



of individualized logistic regression as a replacement for the polychotomous regression model. This method of replacement was studied by Begg and Gray (*Biometrika,* 1984). With this simplified method, each category is individually compared with a baseline category using a dichotomous logistic models. Begg and Gray suggested guidelines for selecting the baseline category. If there is a normal category, it will be appropriate to use it as a baseline. If there is no category that is especially suited to be the baseline, then it is desirable to choose the one with the highest prevalence. In our study no change from previous state was treated as the baseline category. The reduced model shown in Figure 1 was used to define the transition states. Thus, for those subjects with no previous MI we compared death from OC with no change, death from CHD with no change, and first non-fatal MI with no change. For those subjects with the previous MI we compared death from OC with no change, death from CHD with no change, and recurrent non-fatal MI with no change.

In this simplified method, we wish to discriminate among 3 categories and the baseline category, denoted as category 0, on the basis of p covariates. Let's define the $\underset{\sim}{xi}$ vector to be a set of

covariates for the *i*th person, $\underset{\sim}{xi}=(xi_a, .., xi_p)'$, where xi_0 is an indicator for the constant term for n cases. Let Z*di* be an indicator variable which takes the value 1 if the *i*th case belongs to the *d*th category and 0 otherwise. If we were to perform an individual logistic regression comparing category *d* with 0, we would employ a model of the form

$$\log(\theta di/\theta 0i) = xi \times \beta d \ (d=1,2,3)$$

where $\theta di = p(Zdi=1/\underset{\sim}{xi}, Z0i + Zdi = 1)$

$\theta oi = p(Zdi=0/\underset{\sim}{xi}, Z0i+Zdi = 1)$

and where $\underset{\sim}{\beta d}$ is a p+1 vector of unknown parameters.

Therefore, a series of 3 conditional likelihoods are separately maximized. Since maximum likelihood estimation is employed, the estimates of *bd* will be asymtotically unbiased.

Ⅲ. Results

An event modelling program was developed using the University of Minnesota's Micropopulation Simulation Facility. At the start of each simulated time interval, the individual's characteristics

Figure 4. Transition Probabilities at 1973 for an Individual with no Previous MI, Chol-290, DBP-100 and Cig-20.

$$\underset{\sim}{xi} = (290, 100, 0, 1)$$

$$\underset{\sim}{\beta_1} = (-5.544, .0021, .0029, .0330, .041)$$

$$\underset{\sim}{\beta_2} = (-14.609, .0078, .0621, 1.1377, .8805)$$

$$\underset{\sim}{\beta_3} = (-5.163, .0075, .0029, .3640, .1286)$$

$$P_1 = \frac{1}{1+e^{5.544-.0021*290-.0029*100-.033*0-.041*1}} = .0099$$

$$P_2 = .005$$

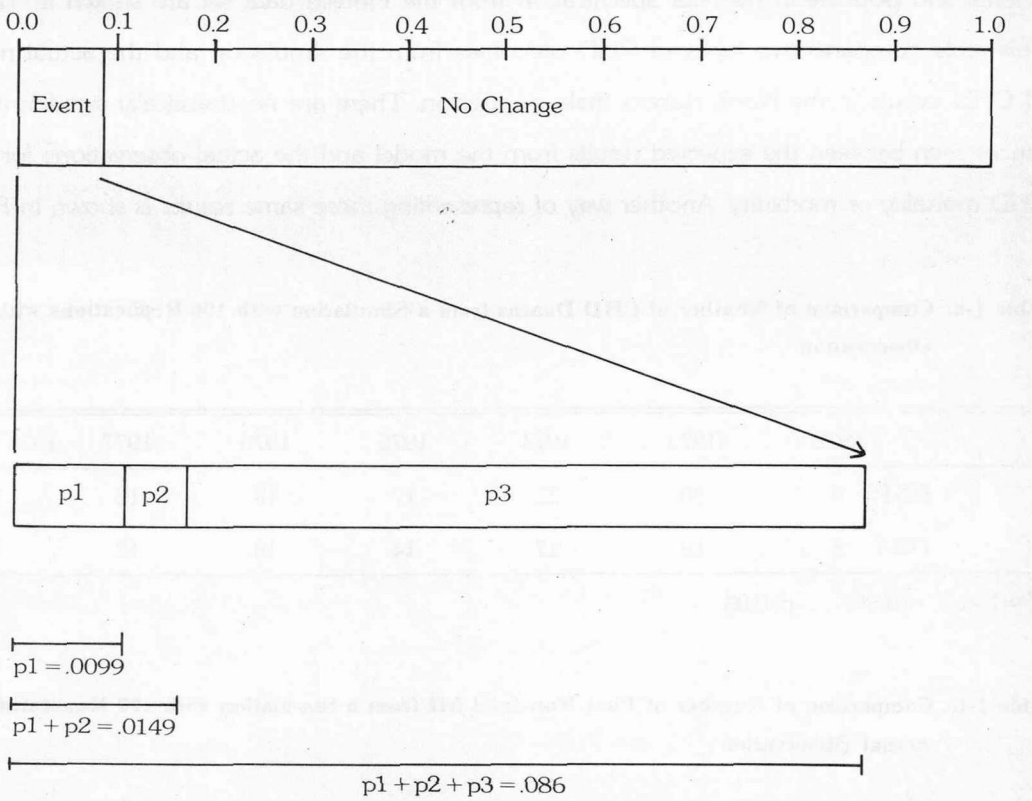
$$P_3 = .071$$

No previous MI
- 0.9141 no change
- 0.0099 death from OC
- 0.005 death from CHD
- 0.071 first non-fatal MI

and the logistics regression coefficients are read from a disc file into the computer memory. A subroutine in the program uses this information to compute the risk of having an event during the simulated interval of one year.

Let us assume that we have an individual with no previous MI, cholesterol level 290, diastolic blood pressure 100 and 20 cigarettes smoked per day. Then the $\underset{\sim}{x}$ vector will have a set of numbers (290,100,0,1) with these values and the two dummy variables for smoking. Figure 4 illustrates three sets of logistic regression coefficients for year 1973. Then p1, p2 and p3 for this year can be computed as shown. Here p1 means the probability of dying of other causes, p2 the probability

Figure 5. Event Decision Scale



if $0 < r \leqq p1$, then death from OC

if $p1 < r \leqq p1$, then death from CHD

if $p1 + p2 < r \leqq p1 + p2 + p3$, then first non-fatal MI

if $p1 + p2 + p3 < r \leqq 1$, then no change

of dying of CHD, and p3 the probability of having of first non-fatal MI.

Now the cumulative probabilities of p1, p2, and p3 are arranged on a 0 to 1 scale as shown in Figure 5. The event part is expanded below. In order to determine which of the mutually exclusive outcomes occur, a pseudo-random number from a uniform distribution over 0 to 1. 0 is obtained. If the random number falls between 0 and p1, then the individual will die from other causes ; if the random number falls between p1 and the sum of p1 and p2, then the individual will die from CHD ; if the random number falls between the sum of p1 and p2 and the sum of p1, p2, and p3, then the individual will have a first non-fatal MI ; if the random number is greater than this cumulative sum, there will be no change in status of the individual.

Results from a seven-year simulation with 100 replications using estimates of regression coefficients and population member specification from the Finnish data set are shown in Table 1. This table compares two types of CHD outcomes from the simulation and the actual number of CHD events in the North Karelia male population. There are no statistically significant differences seen between the expected results from the model and the actual observations for either CHD mortality or morbidity. Another way of representing these same results is shown in Figures

Table 1-a. Comparison of Number of CHD Deaths from a Simulation with 100 Replications with Actual Observation

	1972	1973	1974	1975	1976	1977	1978
EXP	8	20	22	17	18	12	16
OBS	8	18	17	14	18	12	15

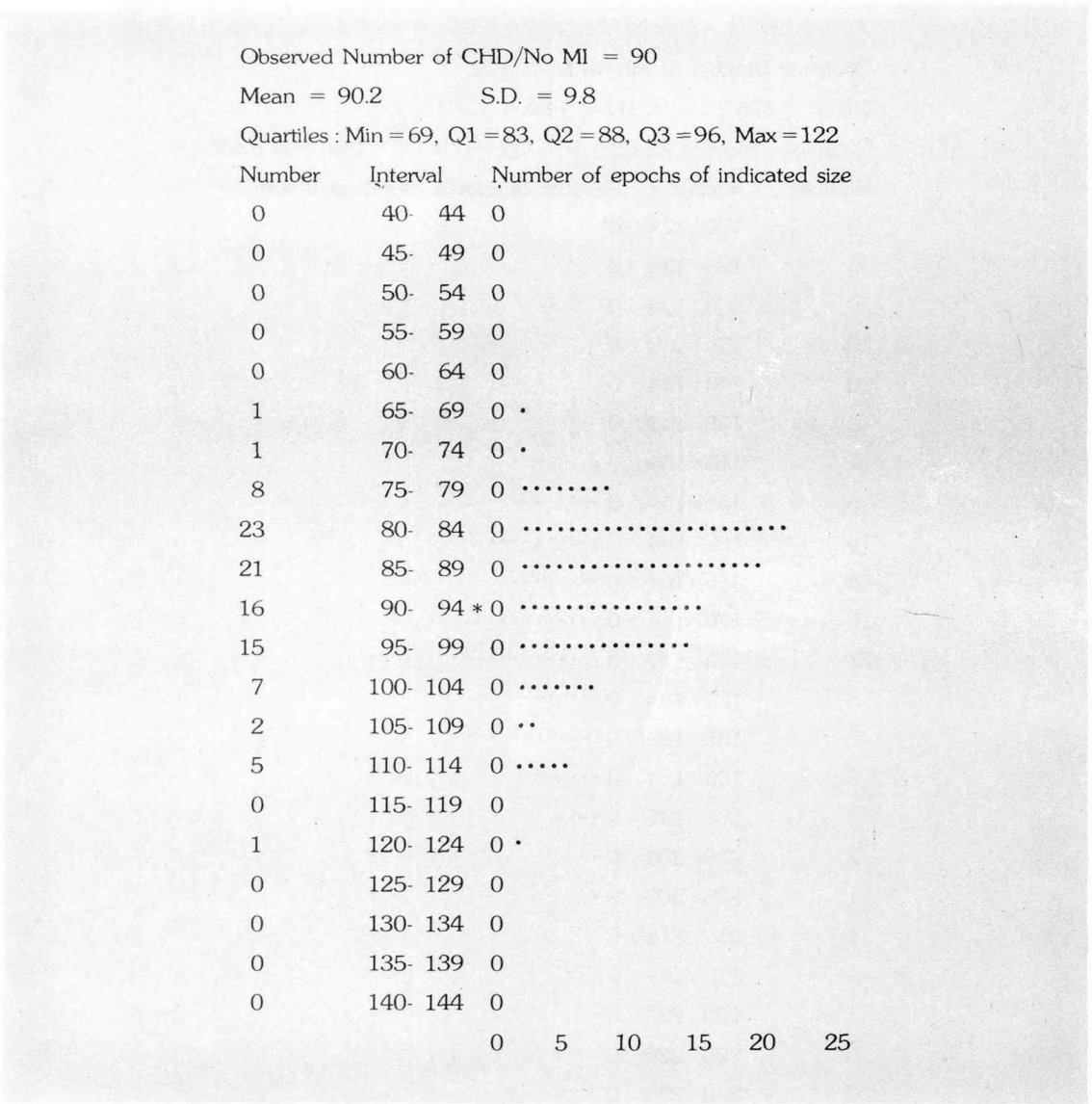
x^2=1.92 df=6 p>0.05

Table 1-b. Comparison of Number of First Non-fatal MI from a Simulation with 100 Replications with Actual Observation

	1972	1973	1974	1975	1976	1977	1978
EXP	21	27	24	29	28	39	34
OBS	21	27	32	28	28	37	34

x^2=2.80 df=6 p>0.05

Figure 6. Distribution of Death from CHD Given No MI over 100 Epochs



Observed Number of CHD/No MI = 90

Mean = 90.2 S.D. = 9.8

Quartiles : Min = 69, Q1 = 83, Q2 = 88, Q3 = 96, Max = 122

Number	Interval	Number of epochs of indicated size
0	40- 44	0
0	45- 49	0
0	50- 54	0
0	55- 59	0
0	60- 64	0
1	65- 69	0 •
1	70- 74	0 •
8	75- 79	0 ••••••••
23	80- 84	0 ••••••••••••••••••••••
21	85- 89	0 ••••••••••••••••••••
16	90- 94 *	0 ••••••••••••••••
15	95- 99	0 ••••••••••••••••
7	100- 104	0 •••••••
2	105- 109	0 ••
5	110- 114	0 •••••
0	115- 119	0
1	120- 124	0 •
0	125- 129	0
0	130- 134	0
0	135- 139	0
0	140- 144	0
		0 5 10 15 20 25

6 and 7. These figures show the distributions of the number of CHD outcomes from a seven year simulation over 100 replications. The asterisks represent the actual number of CHD outcomes; the actual number of both types of CHD events are very close to the mean and median of the number simulated events.

Since the model can mimic the Finnish population results, a next step for us to look at would

Figure 7. Distribution of New MI Given No MI over 100 Epochs

Observed Number of MI/No MI = 172

Mean = 176.2 S.D. = 14.0

Quartiles : Min = 147, Q1 = 167, Q2 = 175, Q3 = 185, Max = 222

Number	Interval	Number of epochs of indicated size
0	120- 124	0
0	125- 129	0
0	130- 134	0
0	135- 139	0
0	140- 144	0
2	145- 149	0 ••
1	150- 154	0 •
8	155- 159	0 ••••••••
10	160- 164	0 ••••••••••
8	165- 169	0 ••••••••
15	170- 174	* 0 •••••••••••••••
21	175- 179	0 ••••••••••••••••••••
8	180- 184	0 ••••••••
12	185- 189	0 ••••••••••••
6	190- 194	0 ••••••
4	195- 199	0 ••••
2	200- 204	0 ••
1	205- 209	0 •
1	210- 214	0 •
0	215- 219	0
1	220- 224	0 •
0	225- 229	0
0	230- 234	0
0	235- 239	0
0	240- 244	0
		0 5 10 15 20 25

be how mortality and morbidit rates vary if we change some of the risk factors. Also more features could be put in the model such as a genetic effect, hospitalization, and other health services.

Ⅳ. Conclusion

A Monte Carlo simulation program uses risk and demographic characteristics of the North Karelia registry to describe the morbidity of myocardial infarction (MI) and the mortality pattern of coronary heart disease and other causes. A stochastic compartmental model has been constructed to describe the temporal dimension of event development. To incorporate the risk factor information and more than two possible end-points, a series of separate simple logistic regressions (Begg and Gray, *Biometrika,* 1984) were performed. The predicted estimates reveal a large effect of cholesterol, and possible risk parameter interaction in females. It is necessary to separate the risk for primary and recurrent MI's in order to match the observed mortality data. Since the model incorporates known distributions of risk factors and individual population characteristics, it will be used to examine the effect of high risk-factor-based and population-based risk factor modification in a community.

Bibliography

Begg, C. and R. Gray, "Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions", *Biometrika,* Vol. 71(1), 1984, pp. 11~18.

Gibbons, R. J., K. L. Lee, D. Pryor, F. E. Harrell, R. E. Coleman, F. R. Cobb, R. A. Rosati, and R. H. Jones, "The Use of Radionuclide Angiography in the Diagnosis of Coronary Artery Disease-A Logistic Regression Analysis", *Circulation,* Vol. 68(4), 1983, pp. 740~746.

Gillum, R. F., R. J. Prineas, R. V. Luepker, H. L. Taylor, D. R. Jacobs, T. E. Kottke, and H. Blackburn, "Decline in Coronary Deaths : A Search for Explanations", *Minnesota Medicine,* Vol. 65, 1982, pp. 235~238.

Kotte, T. E., P. Puska, J. T. Salonen, J. Tuomilehto and A. Nissinen,"Projected Effects of High-risk versus Population-based Prevention Strategies in Coronary Heart Disease", *American Journal of Epidemiology,* Vol. 121(5), 1985, pp. 697~704.

Salonen, J. T., P. Puska, T. Kottke and J. Tuomilehto, "Changes in Smoking, Serum Cholesterol and Blood Pressure Levels during a Community-based Cardiovascular Disease Prevention Program-the North Karelia Project", *American Journal of Epidemiology,* Vol. 114(1), 1981, pp. 81~94.

冠狀動脈疾患의 罹患率과 死亡率 模型의 시뮬레이션

박 현 애*

본 연구의 목적은 冠狀動脈疾患의 罹患率과 死亡率을 이산·다중상태의 確率過程으로 된 慢性病患 시뮬레이션 시스템(Simulation System)으로 模型化하는데 있다. 모형에 포함된 질병의 상태는 1 건강한 경우, 2 관상동맥질환을 앓은 경우, 3 관상동맥질환으로 사망하는 경우, 4 관상동맥질환이 아닌 다른 원인으로 사망하는 경우 등이다. 이들 가능한 질병상태사이의 推移確率(Transition Probabilities)은 性別, 年齡, 관상동맥질환을 앓은 경험의 有無와 콜레스테롤, 혈압, 흡연정도처럼 관상동맥질환에 영향을 미치는 危險要因들을 가지고 로지스틱(Logistic) 回歸分析에서 얻어진 函數에 의해서 推定된다. 本 硏究에서 로지스틱 회귀분석모형들의 係數推定은 핀란드 North Karelia 프로젝트에서 얻은 자료에 최대우도추정법(Method of Maximum Likelihood)을 적용하여 이루어졌다.

로지스틱 회귀분석의 適合度는 위험도의 10分位數方法(The Deciles of Risk Approach)을 사용하여 검증하였고 관상동맥질환의 기본모형의 수행평가는 회귀분석모형의 계수추정시 사용한 자료의 再置換 방법으로 행해졌다.

이 연구를 통해서 얻어진 주요 결과는 다음과 같다. 로지스틱회귀분석을 통하여 관상동맥질환의 罹患率과 死亡率을 豫測하는데 첫째 콜레스테롤値, 둘째 弛緩期 血壓, 세째 吸煙 정도의 순으로 높은 설명력을 보이고 있었다. 硏究에 이용된 여러 모형들의 적합도 검증을 통하여 이들 모형중 콜레스테롤値와 弛緩期血壓으로 이루어진 로지스틱 회귀모형이 North Karelia 프로젝트 자료에서 나타난 이환율과 사망율을 가장 잘 설명하는 모형으로 밝혀졌다.

가장 적합한 로지스틱 모형을 사용한 시뮬레이션에서 얻어진 관상동맥 질환의 罹患者數와 死亡者數는 실제로 관찰된 결과와 통계적으로 일치하였다.

이 모형을 사용하여 관상동맥질환의 가장 효과적인 예방책을 찾기 위해 高危險集團을

*韓國人口保健硏究院 責任硏究員

대상으로 하는 예방 방법과 전체인구를 대상으로 하는 예방 방법을 시뮬레이션을 통하여 罹患率과 死亡率의 감소효과를 비교할 수 있을 것이다. 더 나아가 이 模型은 여러 예방책들의 비용효과 분석에도 기여할 것이다.