TOWARDS COMPUTERIZED UNDERWRITING --

A BIOLOGICAL AGE MODEL

Dr. K.S. Brown
Professor R.L. Brown
Department of Statistics
University of Waterloo
Waterloo, Ontario
N2L 3G1

August 1982

ABSTRACT

This paper presents a theoretical basis for a computerized methodology for estimating biological age, a measure of time to death. The authors then discuss their attempts to reproduce the human underwriting of one life insurance company using a computerized methodology.

INTRODUCTION

Competition within the life insurance industry seems to increase consistently with time. In order to achieve the lowest possible rates for their products, life insurance companies must pursue diligently all possible avenues for reducing expenses.

In the past thirty years, the advent and advancement of computer technology has assisted greatly in reducing costs in many insurance company operations.

In this paper, we will argue that it is now feasible to computerize much of the underwriting process, and with continuous monitoring of the computerized system, one should not have to sacrifice any significant level of underwriting accuracy.

This would have two immediate and obvious advantages. First, costs will be lowered significantly and, second, the time from application to issue of the policy will be shortened significantly.

A BIOLOGICAL AGE MODEL

In determining the premium to be paid for life insurance, or a life annuity, the actuary is concerned with time to death. This is clear from the mathematical formulation of the net single premium functions:

FOR LIFE INSURANCE: $$\bar{A}_x = \int_0^\infty v^t \, {}_tP_x \, \mu_{x+t} \, dt \qquad (1)$$

FOR LIFE ANNUITIES: $$\bar{a}_x = \int_0^\infty \bar{a}_{\overline{t}|} \, {}_tP_x \, \mu_{x+t} \, dt \qquad (2)$$

It is also interesting to note that a good approximation for these values can be determined as follows:

$$\bar{A}_x \simeq v^{\mathring{e}_x} \qquad (3)$$

$$\bar{a}_x \simeq \bar{a}_{\overline{\mathring{e}_x}|} \qquad (4)$$

Hence, our concern is with time to death.

It may seem strange, therefore, that the single most important factor used to determine the size of these premiums is time since birth or chronological age. Of course, we realize that chronological age is an important

factor in predicting time to death, but it has been argued [Brown and Forbes, 1976] that a high correlation with chronological age is neither a necessary nor sufficient condition for an index which accurately measures time to death.

Instead, several biostatisticians have turned to a concept called biological age as a superior way to present, in a single parameter, the best estimate for time to death.

If a risk is assigned a biological age x , this means that the risk has a life expectancy, $\mathring{e}_x$, which corresponds to the life expectancy of the average member of the defined group whose chronological age is x . That is, if you are assigned biological age 45, this means you have the life expectancy of an average person whose chronological age is 45.

Therefore we can see that a biological age model is nothing more than a disguised statement of life expectancy. The reason for this method of statement will be explained later.

There have been several studies on the estimation of an index of biological age (see bibliography).

One study that has been used in the development of the biological age model is the Framingham Study of 1973 which was an epidemiological study of cardiovascular disease. The general framework of this study can be described as follows.

At the start of the study period, measurements on k variables are taken on n individuals deemed free from disease. After a period of m years, the individuals are re-examined and the n_1 individuals who have developed the disease in the interim are noted. Based on these data, it is desired to estimate $P(x_1, x_2, \ldots, x_k)$, the probability that an individual with measurements $(x_1, x_2, \ldots x_k)$ will develop the disease in m years. A model to estimate this probability, suggested by Cornfield (1962), has been used extensively in the analysis of such data:

If $P(x_1, \ldots x_k)$ represents the probability of developing the disease given measurements $(x_1, \ldots, x_k)$ then

$$P(x_1, \ldots, x_k) = \frac{p\delta_1(x_1, \ldots, x_k)}{\delta(x_1, \ldots, x_k)} \tag{5}$$

and

$$1-P(x_1,\ldots,x_k) = \frac{(1-p)\delta_0(x_1,\ldots,x_k)}{\delta(x_1,\ldots,x_k)} \qquad (6)$$

where $\delta_0(x_1,\ldots,x_k)$ and $\delta_1(x_1,\ldots,x_k)$ represent the distributions of $(x_1,\ldots,x_k)$ in the healthy and diseased populations respectively, $\delta(x_1,\ldots,x_k)$ represents the unconditional distribution, and p represents the unconditional probability of developing the disease. Thus from equations (5) and (6)

$$P(x_1,\ldots,x_k) = \left\{1+\frac{(1-p)\delta_0(x_1,\ldots,x_k)}{p\delta_1(x_1,\ldots,x_k)}\right\}^{-1}$$

If δ_0 and δ_1 are assumed to be multivariate normal with the same variance-covariance matrix Σ and means μ_0 and μ_1 respectively, then

$$P(x_1,\ldots,x_k) = \left\{1+\exp\left[-\alpha-\sum_{i=1}^{k}\beta_i x_i\right]\right\}^{-1} \qquad (8)$$

where

$$\alpha = -\frac{1}{2}(\mu_1-\mu_0)'\Sigma^{-1}(\mu_0+\mu_1)-\log\{(1-p)/p\}, \qquad (9a)$$

$$(\beta_1,\beta_2,\ldots,\beta_k) = (\mu_1-\mu_0)'\Sigma^{-1} \qquad (9b)$$

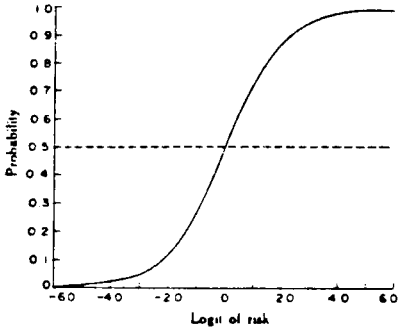


Fig. 1 Graph of the logistic function used to estimate the risk of developing cardiovascular disease, i.e. the probability P is given by $P = \{1+\exp(-LR)\}^{-1}$, where LR is the logit of risk.

When the unknown parameters are replaced by their estimates, the resulting estimates of α and β_i are

$$\hat{\alpha} = -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(\bar{X}_{1j}-\bar{X}_{0j})S^{-1}_{ij}(\bar{X}_{1i}+\bar{X}_{0i}) - \log(n_0/n_1). \quad (10a)$$

$$\hat{\beta}_i = \sum_{j=1}^{k}(X_{1j}-X_{0j})S^{-1}_{ij} \quad (10b)$$

These may be recognized as the estimated linear discriminant function coefficients with $\alpha = c-\log(n_0/n_1)$ (see, for example, Rao 1968). Thus, under the assumptions of multivariate normality with equal variance-covariance structure in both populations, the model and estimates of the parameters are well defined. Of course, the imposed structure is rarely justified since the measured variables will include binary type

data (e.g. positive family history of heart disease) and markedly non-normal data (e.g. number of cigarettes smoked per day). However, the form of the model (see Fig.1) is intuitively a reasonable one for estimating risk, since it ranges between zero and one and increases rapidly over the middle portion of the range. Hence, the model attempts to find the linear function of $(x_1,\ldots,x_k)$ which places the healthy individuals at the 'zero' end of the curve and the diseased individuals at the 'one' end.

Table 1. Probability per 1000 of developing cardiovascular disease[A] in next 8 years by specified characteristics

The characteristics are as shown together with glucose intolerance (GI) absent and left ventricular hypertrophy by ECG (LVH-ECG) negative. The probabilities are given for levels of serum cholesterol (SC) in mg/100 ml and systolic blood pressure (SBP) in mmHg(130 Pa)

(From the Framingham Study 1973)

SC	Probability[B] per 1000 for 45 year old male													
	Non-smoker							Smoker						
	SBP = 105	120	135	150	165	180	195	105	120	135	150	165	180	195
185	22	27	35	43	54	68	84	38	47	59	73	91	112	138
210	28	35	43	54	68	84	104	47	59	73	91	113	138	169
235	35	44	54	68	84	104	129	59	74	91	113	139	169	205
260	44	55	68	85	105	129	158	74	92	113	139	170	206	247
285	55	68	85	105	129	158	192	92	113	139	170	206	247	293
310	68	85	105	130	158	192	232	114	140	170	206	248	294	345
335	85	105	130	159	193	232	277	140	171	207	248	295	346	401

[A] In the Framingham definition, cardiovascular disease is considered to have developed if there is a definite manifestation of coronary heart disease, intermittent claudication, congestive heart failure or cerebrovascular accident in the absence of a previous manifestation of any of these diseases or rheumatic heart disease.

[B] Probabilities are estimated using the model

$$P(x_1,\ldots,x_8) = \{1+\exp\ (-\hat{\alpha}-\sum_{i=1}^{8}\hat{\beta}_i x_i)\}^{-1}$$

where $\alpha = -19.7709560$ and

i	$\hat{\beta}_i$	x_i	i	$\hat{\beta}_i$	x_i	
1	0.3743307	age	5	0.5583013	cigarettes	(0 nonsmoker;1smoker)
2	-0.0021165	(age)x(age)	6	1.0529656	LVH-ECG	(0 none; 1 present)
3	0.0258102	SC	7	0.6020336	GI	(0 absent; 1 present)
4	0.0156953	SPB	8	-0.0003619	SC x age	

Estimates of α and $(\beta_1,\ldots,\beta_k)$ when assumptions of normality are not made (but the form of the model is assumed) may be obtained by maximizing the likelihood function

$$L(\alpha,\beta_1,\ldots,\beta_k) = \prod_{j\in U_0}\{1-P_j(x_1,\ldots,x_k)\}\ \prod_{j\in U_1}P_j(x_1,\ldots,x_k)$$

for α and $(\beta_1,\ldots,\beta_k)$, where U_0 and U_1 represent the sets of individuals in the healthy and diseased populations respectively and $P_j(x_1,\ldots,x_k)$ represents equation (8) evaluated at the values of $(x_1,\ldots,x_k)$ observed for the jth individual. The maximum likelihood estimates often do not differ significantly from the linear discriminant coefficients and the latter are

often used because their calculation does not involve iteration.

In such problems, there are often a large number of independent variables available for entry into a model. It is a common problem to attempt to determine an "optimal" subset of these variables; that is a relatively small number of the independent variables which may predict risk nearly as well as the entire set.

The importance of any variable, say the ℓth (e.g. systolic blood pressure), for predicting risk may be investigated by maximizing $L_1(\alpha, \beta_1, \beta_2, \ldots, \beta_\ell = 0, \ldots \beta_k)$ The ratio $-2\log\{L_1(\tilde{\alpha}, \tilde{\beta})/L(\hat{\alpha}, \hat{\beta})\}$, where $(\tilde{\alpha}, \tilde{\beta})$ is the vector of estimates under the second model and $(\hat{\alpha}, \hat{\beta})$ is the vector of estimates under the full model, is asymptotically χ^2_1 under the hypothesis $\beta_\ell = 0$ and large values of this quantity indicate evidence against the hypothesis that the ℓth variable is unimportant in predicting the development of the disease, after adjusting for other variables.

These types of tests form the basis for stepwise procedures for model building. That is, variables can be entered into models one at a time, at each step the variable entered is the one which increases the likelihood function

the most, given the previous variables in the model. The procedure stops when no variable, not in the model, increases the likelihood appreciably. Such procedures can be combined with variable elimination procedures which remove variables one at a time until the removal of any variable in the model would decrease the likelihood appreciably. These stepwise procedures have been extensively used to try to determine "optimal" subsets of independent variables; however, they have recently come into considerable criticism (e.g. Hocking(1976), Lawless and Singhal (1978)) . With the development of high speed computers and numercial methods it is possible to very quickly screen a great many of the models and determine which models might be suitable "candidates" for consideration for an optimal model. The programme SMOD as described in Lawless and Singhal (1978) does this model screening, and was used in the analysis described later.

One advantage of developing adequate models of risk, other than the ability to test quantitatively for significant predictor variables, is that the model may be used to prepare convenient summary tables such as is presented in Table 1. Such tables allow a physician to inform the patient of his personally estimated risk of developing disease and also enable him to determine which factors are elevating this risk. Further, when combined with a table such as Table 2 the physician is

able to present the individual's risk relative to individuals experiencing "average" (or perhaps, with some modification to the table, "ideal") risk.

Table 2. Appraised age of individual with specified characteristics

The appraised or risk age is the age of the average Framingham Study (1973) male with the equivalent risk of developing cardiovascular disease in the next 8 years by the same characteristics as specified in Table 1 (GI absent, LVH-ECG negative)

[From Brown and Forbes 1976]

	Appraised age[A] of 45 year old male													
SC	Non-smoker							Smoker						
	SBP = 105	120	135	150	165	180	195	105	120	135	150	165	180	195
185	36	37	39	40	42	44	46	39	41	43	45	47	50	53
210	37	39	40	42	44	46	49	41	43	45	47	50	53	56
235	39	40	42	44	46	49	52	43	45	47	50	53	56	63
260	40	42	44	46	49	52	55	45	47	50	53	56	63	71
285	42	44	46	49	52	55	60	47	50	53	56	63	71	74
310	44	46	49	52	55	60	70	50	53	56	63	71	74	75
335	46	49	52	55	60	70	73	53	57	63	71	74	75	77

[A] Calculations were based on the following average risks (see Section 28, the Framingham Study)

Age	35	40	45	50	55	60	65	70
Risk/1000	18	41	75	115	159	193	212	229

and risks of approximately 0/1000 and 1000/1000 at ages 0 and 105 respectively. The appraised age was estimated by an average quadratic logistic interpolation procedure; results >70 were obtained by interpolation between the arbitrary values given above, and should be treated with caution.

Table 2 is based on Robbins and Hall's (1970) approach to preventive medicine, entitled 'health hazard appraisal'. In this system, an individual's chances of dying from a number of diseases are computed, based on his physiological measurements, lifestyle etc. These are combined, and the composite risk is compared with the risk of death experienced by an"average"member of the population. The individual's appraised age, the age of the average individual with same risk of death, is obtained in this way together with recommendations aimed at reducing the appraised age. Since most persons appreciate the consequences of being, say, 5 years older than they are chronologically, the message is presented more effectively than if the same information were expressed as a probability. In this way, individuals appraised as being 'older' than their chronological age may be encouraged to reduce factors which are elevating risk, and those appraised as 'younger' may have their positive lifestyles reinforced.

A similar "health hazard appraisal" model is used by Health and Welfare Canada in their "Evalu-Vie" programme. The input to the computer program are the coded answers to the following questions:

HEALTH HAZARD APPRAISAL

RISK FACTORS: DETAIL AND CODES

FCT#	FACTOR	ABBREV.	CODE	DESCRIPTION
1	SEX	SEX	1	Male
			2	Female
2	AGE	AGE		Enter age in years
3	ALCOHOL HABITS	ALCOHOL	1	41 or more drinks per week
			2	25-40 drinks per week
			3	7-24 drinks per week
			4	3-6 drinks per week
			5	1-2 drinks per week
			6	STOPPED: Stopped drinking (Person has stopped before symptoms of cirrhosis) Factor should be given to stopped drinkers regardless of amount.
			7	NON DRINKER: Never been a drinker
4	ARREST RECORD	ARREST RECORD	1	Burglary, Robbery, Assault
			2	Without violence or threat
			3	No Arrests
5	WEAPONS	WEAPONS	1	Carry
			2	Does not carry
6	DEPRESSION	DEPRESSION	1	Often severely depressed
			2	Seldom or never severely depressed

N.B. Number of "drinks" should include aperitifs, wines, beer, etc.

FCT#	FACTOR	ABBREV.	CODE	DESCRIPTION
7	MILES PER YEAR	MILES		Enter miles driven per year and/ or miles as an auto passenger
8	SEAT BELT USE	SEAT BELT	1 2 3 4	Worn less than 10% of the time Worn 10-24% of the time Worn 25-74% of the time Worn 75-100% of the time
9	HISTORY OF BACTERIAL PNEUMONIA	PNEUMONIA	1 2	Has had Has not had
10	BLOOD PRESSURE SYSTOLIC (if unsure enter 120)	BP: SYSTOLIC		Enter systolic blood pressure in mm
11	BLOOD PPESSURE DIASTOLIC (if unsure enter 80)	BP: DIASTOLIC		Enter diastolic blood pressure in mm
12	BLOOD CHOLESTEROL (if unsure use 2)	CHOLESTEROL	1 2 3	Cholesterol Level 280+ " " 220-279 " " 219 and below
13	DIABETIC	DIABETES	1 2 3	Diabetic Diabetic (Controlled) Not Diabetic
14	HEIGHT	HEIGHT		Enter Height in inches with sho (Without shoes: ADD 1 inch for Males 2 inches for Females)
15	WEIGHT	WEIGHT		Enter weight in pounds (In indoor clothing and shoes)
16	FRAME	FRAME	1 2 3	Small Medium Large

FCT.	FACTOR	ABBREV.	CODE	DESCRIPTION
17	DRUGS AND MEDICATION INFLUENCING MOTOR VEHICLE OPERATION	DRUGS/MED	1 2 3	Excess Moderate None
18	EXERCISE	EXERCISE	1	SEDENTARY: Work and Leisure. Under 5 flights of stairs or half mile walking per day
			2	LOW MODERATE: Some Activity Work and Leisure. Between 5 and 15 flights of stairs or 0.5 to 1.5 miles walking or comparable daily activity.
			3	HIGH MODERATE: Programmed exercise 4 times per week or 1.5 to 2 miles of walking or 15 to 20 flights of stairs or comparable daily activity.
			4	VIGOROUS: Greater than moderate.
19	SMOKING HABITS 1) for current smoker heaviest in past 5 years 2) for ex-smoker mark heaviest amount smoked in year before quitting	SMOKING	1 2 3 4 5 6 7	Cigarettes, 40 or more/day Cigarettes, 20-39/day Cigarettes, 10-19/day Cigarettes, less than 10/day Cigars or pipes ONLY;5 or more/day or any amount inhaled Cigars or pipes ONLY;less than 5/day NOT inhaled Nonsmoker: (Never smoked or not smoked for 10 years)
20	CURRENT SMOKING STATUS	STOPSMOK	0 1 2 3 4 5 6 7 8 9	Still smoking or Nonsmoker Years of having stopped smoking

PCT	FACTOR	ABBREV.	CODE	DESCRIPTION
21	FAMILY HISTORY OF ISCHEMIC HEART DISEASE	FH/HEART	1	Both parents died before 60 of Ischemic Heart Disease
			2	One parent died before 60 of Ischemic Heart Disease
			3	One or both parents died before 60 of cause other than Ischemic Heart Disease OR still alive below 60
			4	None of the above
22	FAMILY HISTORY OF DIABETES	FH/DIAB		FAMILY HISTORY: mother, father, sister, brother, child
			1	Yes
			2	No
23	FAMILY HISTORY OF SUICIDE	FH/SUICD	1	Yes
			2	No
24	EMPHYSEMA AND/OR BRONCHITIS	EMPHYSEMA	1	Has Emphysema and/or Bronchitis
			2	Has no signs or symptoms of Emphysema and/or Bronchitis
26	RECTAL POLYP	POLYP	1	Has had
			2	Has not had
27	PROCTO-SIGMOIDOS-COPY	PROCTO	1	Has annually
			2	Does not have annually
29	RECTAL BLEEDING	RCTBLOOD	1	Has had undiagnosed rectal bleeding in the past year
			2	Has not had undiagnosed rectal bleeding in the past year
30	CHRONIC RHEUMATIC HEART DISEASE	RH:FEVER	1.	Rheumatic Heart Murmur, no Chemoprophylaxis
			2	Rheumatic Heart Murmur, on Chemoprophylaxis
			3	History of Rheumatic Fever but no Murmur, no Chemoprophylaxis
			4	History of Rheumatic Fever but no Heart Murmur, on Chemoprophylaxis
			5	No history of Rheumatic Fever and no Rheumatic Heart Murmur

FCT #	FACTOR	ABBREV.	CODE	DESCRIPTION
31	SIGNS OR SYMPTOMS OF CHRONIC RHEUMATIC HEART DISEASE	RH: S/O/S	1	No
			2	Yes
32	ULCERATIVE COLITIS	ULCERCOL	1	Has had Ulcerative Colitis 10 years or more
			2	Has had Ulcerative Colitis less than 10 years
			3	Has no symptoms of Ulcerative Colitis

**THE FOLLOWING FACTORS ARE FOR FEMALES ONLY:

FCT#	FACTOR	ABBREV.	CODE	DESCRIPTION
		FACTORS 33-38 ARE TO BE CODED BY FEMALES ONLY:		
33	VAGINAL BLEEDING	VAGBLOOD	1	Has had undiagnosed vaginal bleeding in past year
			2	Has not had undiagnosed vaginal bleeding in past year
34	AGE AT MARRIAGE OR ONSET OF INTERCOURSE	AGE/MAR	1 2 3	Teenage 20-25 Over 25 or never
35	PAPSMEAR	PAPSMEAR	1 2 3 4	Has not had Negative within 5 years Negative within 1 year 3 Negative within 5 years
36	ECONOMIC AND SOCIAL STATUS	SOCIO/EC	1 2 3	Low Average High
37	JEWISH	JEWISH	1 2	No Yes
38	FAMILY HISTORY OF BREAST CANCER	FH/BREAST	1	Mother or sister had Breast Cancer
			2	Mother or sister had Breast Cancer but patient examines breasts regularly and has periodic examination by physician
			3	Neither mother nor sister had Breast Cancer
			4	Neither had Breast Cancer but patient examines breasts regularly and has periodic examination by physician

Thus, a biological age index may be seen as a function of a set of observable quantities, each of which makes an important contribution to the prediction of age at death, years until death or some other measure stratifying the population by risk subgroups.

A biological age index, y, may be written then in the form

$$y = f(x,V,I,\Delta)$$

where y represents, for example, age at death, years until death, reduction in optimal lifespan, an individual's appraised or risk age etc; x is the chronological age; V represents the additional contribution (i.e. adjusted for chronological age and other variables in the function) of the absolute level of a set of variables, such as blood pressure, cigarette smoking habits, family history, etc.; I represents the additional contribution of the interaction of variables in V with other variables in V and with chronological age; and Δ represents the additional contribution of past changes in any of the variables in V, and also of changes in variables that are not in V but are important because a change in them implies an elevation or reduction of y. The variables in the sets V and Δ and all constants entering into this function have to be determined from a study of a reasonably large number of individuals followed longitudinally.

A COMPUTERIZED UNDERWRITING EXPERIMENT

Late in 1981, the authors approached the Mutual Life of Canada, a large mutual insurance company located in Waterloo, to see if we might test the biological age model by attempting to computerize their underwriting process.

Mutual Life presently issues about 80,000 individual life insurance policies a year. Approximately 75% of these policies are issued non-medical. Even for these policies, the cost of underwriting and issue averages close to $100 an application and the time needed for underwriting averages around five days.

The Mutual Life allowed us access to their adult non-medical cases. We were sent coded data which included the following information:

- beneficiary relationship
- insurance amount
- the underwriters decision (issue standard, issue rated, or decline)
- and if the decision was other than standard what other information was requested by the underwriter.

We were also provided with the answers given to the non-medical application form as shown on the next page.

The Mutual Life Assurance Company of Canada / Waterloo, Ontario

To be used only if the Life Insured has attained 16 years of age (18 years in Quebec)

APPLICATION PART 2 - EVIDENCE OF INSURABILITY

1 Name in full (print) _____ Date of birth _____ Sex _____

2 **A** Name and address of usual medical adviser (if none so state) _____

B When and why last visited? _____

C What treatment was given or medication prescribed? _____

3 Are you now under observation or taking treatment? If yes, give details _____

	YES	NO
4 Have you ever been treated for or ever had any indication of (please specify which)		
A Disorder of eyes, ears, nose or throat?	☐	☐
B Severe headaches, dizziness, fainting, loss of consciousness, fits, epilepsy, speech disorder, paralysis, stroke, nervous breakdown, mental trouble or other disorder of nervous system?	☐	☐
C High blood pressure, palpitation or pain about the heart or chest, difficult breathing, cardiac asthma, angina or coronary disease, rheumatic fever, heart murmur or other disorder of heart or blood vessels?	☐	☐
D Persistent cough or hoarseness, coughing of blood, asthma, pleurisy, bronchitis, tuberculosis or other disorder of the lungs?	☐	☐
E Ulcer of stomach or duodenum, recurrent indigestion, jaundice, gall stones, colitis, bleeding or other disorder of stomach, gall bladder, liver, intestines or rectum?	☐	☐
F Sugar, albumin or blood in urine, venereal disease, kidney stone or colic or any other disorder of kidney, bladder, genital organs, breasts or disorder of pregnancy	☐	☐
G Arthritis, gout, rheumatism, sciatica, deformity or disorder of joints, limbs or back?	☐	☐
H Cancer or other tumor, enlarged glands or skin disease?	☐	☐
I Diabetes, thyroid or other endocrine disorder?	☐	☐
J Any illness, disease or operation not mentioned above?	☐	☐
K Female life. Pregnant? (If so, give expected delivery date)	☐	☐
5 Do you now or have you ever used alcoholic beverages?	☐	☐

If "yes" complete following questions

A Frequency of use (daily, weekly, monthly)
B Amount consumed on each occasion
C Date last used
D Any treatment for alcohol use (including AA membership)
E Any motor vehicle impaired driving convictions

6 Do you now or have you in the past 3 years used tobacco? ☐ ☐

If "yes" give daily use	1-5	6-10	11-25	over 25
Cigarettes				
Other (specify)				

If discontinued when and why?

7 Have you ever used heroin, morphine, other narcotics, barbiturates, amphetamines or psychoactive (marijuana, LSD, etc.) drugs except as prescribed by a physician? ☐ ☐

8 Other than as stated in above questions specify if you have

A Been a patient or advised to have a diagnostic test, hospitalization or surgery in a clinic, hospital, sanitorium, or medical facility. ☐ ☐

B Used the service of any other physicians in the last 5 years ☐ ☐

Details of yes answers:

Identify question number, circle applicable items. Include diagnosis, treatment, dates, duration and names and addresses of all attending physicians and medical facilities.

9 Have you ever had	YES or NO	DATES	WHY TAKEN?	RESULT	NAME AND ADDRESS OF PHYSICIAN ORDERING INVESTIGATION
A An electrocardiogram?	☐ ☐				
B Any blood tests?	☐ ☐				
C Any X-rays? SPECIFY	☐ ☐				

10 Has an application for insurance or annuity on your life ever been declined, rated or modified in any way? Yes ☐ No ☐ When? _____ Why? _____ Company? _____

11 Have you applied for or received a pension or compensation because of illness or injury? Yes ☐ No ☐ Give Details _____

12 Height (in shoes) _____ ☐ ft. & ins. ☐ cm.
Did you measure? Yes ☐ No ☐

Weight (house clothing) _____ ☐ lbs. ☐ kgs.
Did you weigh? Yes ☐ No ☐

Weight change in past 12 months Gain _____ Loss _____
Reason: _____

DECLARATION: I declare the above answers and statements are full, complete and true and shall form part of the evidence of insurability in respect of my application for insurance (or for reinstatement of or change in my present insurance) in The Mutual Life Assurance Company of Canada.

AUTHORIZATION: I authorize any physician or practitioner who has observed me for diagnosis or treatment, and any hospital, clinic or other medical or medically related facility where I have been a patient, and any insurance company, the Medical Information Bureau or other organization, institution or person, that has any records or knowledge of me or my health, to give full particulars thereof including any prior medical history to the Mutual Life Assurance Company of Canada, or its reinsurers. A photostat of this authorization shall be as valid as the original.

Signed at _____ **Date** _____

Witness _____ **Signature of Life Insured** _____

3-2-80

COMPLETE THIS SECTION FOR PARAMEDICAL AND MEDICAL EXAMINATIONS

12 Height (in shoes)_____ ☐ ft. & ins. ☐ cm. Did you measure? YES ☐ NO ☐

Weight (house clothing)_____ ☐ lbs. ☐ kgs. Did you weigh? YES ☐ NO ☐

Weight change in past 12 months: Gain_____ Loss_____ Reason

13 Girth of bared chest: full inspiration _____ full expiration _____ Girth of abdomen: at umbilicus_____ Males only

14 Blood Pressure (sitting – without rest or exercise)
Repeat at end of examination if over 140/90

READINGS	FIRST	SECOND	FINAL
Systolic	mm	mm	mm
Diastolic (at cessation of sound)	mm	mm	mm

15 **A** Pulse Rate_____

B Effect of exercise (twenty rapid toe touches or equivalent)

	BEFORE EXERCISE	IMMEDIATELY AFTER	TWO MINUTES AFTER
Pulse rate per minute			
Extra systoles per minute			

16 **Urinalysis**
(Haemacombistix Method)
Glucose Neg ☐ Pos ☐
Protein Neg ☐ Pos ☐
Blood Neg ☐ Pos ☐
ONLY Send urine sample to Head Office if any test is positive
Date mailed_____

ADDITIONAL REPORT BY MEDICAL EXAMINER

17 **Any evidence of past or present disease of:** YES NO

(Please comment fully on any abnormal finding.)

		YES	NO
A	Nervous System? (reflexes, coordination, tremors etc.)	☐	☐
B	Head and Neck?	☐	☐
	Ears? – deafness, discharge, hearing aid, etc	☐	☐
	Eyes? – blindness, retina, etc	☐	☐
	Mouth? – including throat	☐	☐
C	Heart and blood vessels? – Examine in erect and recumbent position before and after exercise. Complete section 18 if any abnormality is found	☐	☐
D	Chest and Lungs? Examine on bare chest with expiratory cough	☐	☐
E	Abdomen? Liver, spleen, abnormal masses, tenderness, hernia – reason for surgical scar	☐	☐
F	Genito-Urinary System? (include prostate)	☐	☐
G	Musculoskeletal System? (include spine, joints, deformities)	☐	☐
H	Endocrine System? (include thyroid, breasts)	☐	☐
I	Skin – xanthomas, nevi etc? Lymph nodes?	☐	☐

18 **A** Is there a murmur in the (if more than one, describe below as no. 1 and no. 2)

Erect_____ YES or NO Recumbent_____ YES or NO Left lateral position_____ YES or NO

Timing	Intensity	Quality	Location
☐ Systolic	☐ Faint	☐ Soft	☐ Mitral
☐ Pre-systolic	☐ Moderate	☐ Blowing	☐ Aortic
☐ Diastolic	☐ Loud	☐ Rough	☐ Pulmonic

Transmission beyond valve area
☐ Yes ☐ No

B Does exercise produce
☐ Shortness of breath ☐ Palpitation ☐ Pain ☐ No Distress

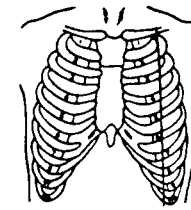
C Does exercise cause the murmur to
☐ Increase ☐ Decrease ☐ Disappear ☐ Remain the same

D Do you consider the heart enlarged? Yes ☐ No ☐

E Please locate

Apex by X
Area of murmur by dotted outline
Point of greatest intensity by O
Transmission by

Midclavicular line



F Is there a history of streptococcal or specific infection?
Yes ☐ No ☐

G What is your diagnosis?

19 How long have you known the person examined? Is the person examined a patient? Yes ☐ No ☐ Is appearance unhealthy or older than stated age? Yes ☐ No ☐

Has applicant to your knowledge ever abused the use of alcohol or been addicted to drugs? Yes ☐ No ☐

Do you know of any facts bearing on the risk which are not brought out by the foregoing questions? Yes ☐ No ☐ If "yes", give details below

Was examination completed in your office? Yes ☐ No ☐ Did you require an interpreter to question the person examined? Yes ☐ No ☐

ADDITIONAL REMARKS

Signed at **Date** , at A.M. P.M.

Signature of Examiner

Our first and most serious problem was that there existed little or no matching between the non-medical questions asked by Mutual Life and the questions used in the biological age models now in existence. Hence, it was impossible for us to feed the Mutual Life data into any of the pre-packaged biological age models.

Instead, it was decided that we would access two sets of underwriting data. Each set would consist of approximately eight hundred applications. Using the first set of applications (which, in fact, totalled 824) we would build a model that could discriminate between those applicants that were rated or rejected from those applicants who were issued standard without further information. Using the model developed on this first set of applications, we would then run it on the second set of applications (which, in fact, totalled 829) to see if our model could correctly separate out those applicants that could safely be issued insurance at standard rates with no further underwriting, from those who could not.

Using the first 824 apps, then, we built a model of the type described earlier (e.g. see equation 8) where, in this instance, $P(x_1,\ldots,x_k)$ is the probability that an individual with variables $x_1,\ldots,x_k$ corresponding to information from the application, will be judged non-standard or rejected.

Then

$$P(x_1,\ldots,x_k) = \frac{1}{1+e^{-\alpha-\Sigma\beta_i x_i}}$$

where $(x_1,\ldots,x_k)$ represent responses to a set of K items chosen from the questionnaire, and the α's represent the weights attached to the items.

The data on which the model was built consisted of information, of the type described earlier, on 824 applications representing approximately one month's adult non-medical applications. Of these 824 applications, 10 were issued non-standard, and 17 were rejected. For the purposes of these analyses, these latter 27 policies were grouped into one category, called the non standard group hereafter. Further, the analysis emphasized only the health related information from the questionnaire.

Initially, simple summaries of the data comparing the responses of the standard group and the non-standard group on each of the medical based questions from the application were conducted. These analyses indicated certain questions for which the response pattern was different in the two groups. Secondly, as an overall measure of the status of the applicant, a single variable, the number of

"no" answers to certain questions* was created. Finally, guided by the results of these preliminary analyses, the model screening programme (SMOD; see earlier description) was used to identify subsets of the questions, the responses to which would separate the two groups almost as well as the responses to the large number of original questions. This latter procedure had to be carried out in steps. At each stage, different combinations of items from the questionnaire were entered into a full model and those items which did not appear to make any significant contribution to the prediction of the non-standard cases were deleted.

Finally, two models emerged as models which seemed to be able to discriminate between standards and non-standards nearly as well as the full set of independent variables.

The variables in these models were:

---

* Questions: 3, 4A, 4B, 4D, 4E, 4F, 4G, 4I, 4J, 5(all), 6(all), 7, 8A, 8B, 9A, 9B, 9C, 11, were considered in arriving at this index.

MODEL 1	MODEL 2
Response to Question 8A	Alcohol use
Response to question 9B	Cigarette smoking
The number of "no" answers	Response to question 8A
Age	The number of "no" answers
	Age

Both models seemed to predict equally well, however there were some anomalies in the models in that responding "yes" to the questions on alcohol use, cigarette smoking and question 8A increased the estimated probability of being judged standard. (Those patterns were also apparent in the original data, i.e. a higher proportion of standard cases drank, and a higher proportion smoked.)

With the maximum likelihood estimates of the coefficients in the model, it was possible to estimate, for each subject, his or her chance of being judged standard. By choosing a cut-off level and declaring standard anyone with an estimated probability of being standard which exceeded that value, it was possible to assign each individual to one of the four cells in the summary table below.

Table 3. Results of classification based on the model

		ACTUAL RATING: Standard	ACTUAL RATING: Non-Standard	
MODEL CLASSIFICATION	Standard	n_{00}	n_{01}	
	Non-standard	n_{10}	n_{11}	
		797	27	824

Thus, in the table, $n_{00} + n_{11}$ cases are correctly classified, while $n_{10} + n_{01}$ are incorrectly classified. By varying the cut-off level it is possible to increase or decrease the number predicted in the standard class. For the purposes of this illustration, the cut-off level $\frac{795}{824} = .9672$ was chosen as this produced small values of n_{01} with reasonably small values of n_{10}. In practice, one could choose a cut-off level which reduced the size of n_{01} at the expense of increasing n_{10} by assessing the costs of misclassifying a standard as a non-standard against the costs of the other misclassification (i.e. the extra cost of underwriting 'standard' cases vs. missing a substandard case). Alternatively, one could choose to underwrite, by hand, the x% of cases with the lowest estimated probabilities of being standard.

When applied to the original data set, using the cut-off level .9672, the two models were able to discriminate as shown below.

Table 4. Results of classification for Models 1 and 2

MODEL 1

	ACTUAL RATING		
	Standard	Non-standard	
Standard	668	2	670
Non-standard	129	25	154
	797	27	824

MODEL 2

	ACTUAL RATING		
	Standard	Non-standard	
Standard	681	4	685
Non-standard	116	23	139
	797	27	824

Thus, using model 1, 15.9% ($\frac{131}{824}$) of the cases were misclassified but most importantly, only two non-standard cases were misclassified as standard. Increasing the cut-off to .975 resulted in only one non-standard case being misclassified as standard, however 151 standard cases were classfied as non-standard. This one non-standard case could not be correctly classified with a cut-off level as high as .99*. Model 2 did not fare quite so well, misclassifying fewer cases overall, but

---

* Further investigation revealed that this case was rated non-standard on the basis of special information on the applicant's arthritis, gout and rheumatism obtained from X-ray examination.

missing 4 out of 27 non-standard cases(3 out of 27 using a level of .975).

It is well known that regression models often predict considerably better for the data set on which they were built than on other similar data sets; but in testing the models developed earlier on the next 829 applications coded, (819 standard, 10 non-standard or rejected) as indicated in Table 5, both models were able to correctly identify all 10 non-standard cases!! Model 1, using a cut-off point of .9672, misclassified 139 standard cases as non-standard, while model 2, with a cut-off level of .9672 did marginally better, misclassifying 135 standard cases.

Table 5. Results of Classifications based on models 1 and 2 for the second data set

MODEL 1

	ACTUAL RATING		
	Standard	Non-standard	
Standard	680	0	680
Non-standard	139	10	149
	819	10	829

MODEL 2

	ACTUAL RATING		
	Standard	Non-standard	
Standard	684	0	684
Non-standard	135	10	145
	819	10	829

Thus, model 1 would have declared 680 applications as standard, leaving only 149 (18%) to be underwritten in the usual way. This could have resulted in a potential saving of about 80% of the cost of underwriting, balanced against the additional expense of approximately $.62, the cost of computer time to use the model to classify the 829 applications.

CONCLUSION

As pointed out earlier, the Mutual life issues close to 80,000 individual life insurance policies a year. 75% of these are issued non-medical with an underwriting and issue cost of around $100 each and delay time of close to five days.

Only 3.5% of the non-medical applications are rate or declined, on average.

We have shown, through the use of a fairly crude model, that a computer can be programmed to produce results very close to those determined by human underwriters. This was done using input data that was not designed for computerization nor was it based on any preconceived model such as the biological age model.

We feel that a great deal of the underwriting that is being done in an expensive and relatively slow manner today could be computerized. This would require using a machine-readable application form and we might wish to use input more suited to the existing biological age models, although we have shown that this may not be essential.

One of the exciting prospects of a computerized underwriting model is that it can continuously reassess itself! For example, every time a policyholder dies, the computer can retrieve the application form for the policyholder and determine what information was provided that might have predicted that early (normal, or late) death. In this manner, the computer can statistically reanalyze the weights that have been given to each input parameter and continuously improve its own programme!

Once one has faith in this computerized model, one can have a much more refined pricing stratification than exists today. No longer will 75-85% of all cases be priced based on sex and biological age only. Rather, the computer will determine the "time to death" in its program and set the proper premium level accordingly.

This same methodology, applied to life annuities, would go far to answering the feminists clamouring for unisex annuity mortality tables since one would no longer be pricing based only on sex and chronological age. Rather, one would be pricing based on an objective statistical prediction as to "time to death".

One can visualize a day, in the not-too-distant future, where an agent will enter the client's home with his portable computer terminal. After using the terminal as a sales aid (e.g. to show cost comparisons and investment attributes) the agent will then ask the client the usual application questions. The answers to these questions will be entered into the computer through the portable terminal and, in a matter of seconds, the computer will tell the agent whether his client has been accepted, rejected, or whether further information is required. (Note -- the term rated will no longer be used).

If the client is accepted, the computer will produce the price level for the policy required based on the statistical analysis of time-to-death. Then, if the client is in agreement, the portable terminal will print out a policy and the process will be complete!

The authors would like to thank the Mutual Life of Canada and, in particular, Barry J. Triller (New Business Executive) and Wilhemina Gould (New Business Analyst) for the assistance with this project. Without their help, this research could not have been completed. We would also like to acknowledge the assistance of Evan Kelly who was responsible for the computer analysis underlying this project.

BIBLIOGRAPHY

Borkan, G. A. (1980). Assessment of Biological Age Using a Profile of Physical Parameters, Journal of Gerontology Vol. 35, No. 2, pp. 177-184.

Brown, K. S. (1977). Mathematical Contributions to the Study of Risk Factors in Cardiovascular Disease. II., Math Scientist, 2, 111-125.

Brown, K. S., Forbes, W. F. (1976) Concerning the Estimation of Biological Age, Journal of Gerontology, Vol. 22, pp. 428-437.

Brown, K. S., Nabert, W. (1977). Evaluation of the Existing Method for Calculating Health Hazard Appraisal Age. Drug Directorate, Health Protection Branch, Health and Welfare Canada.

Comfort, Alex. (1956). The Biology of Senescence. Churchill Livingstone.

Comfort, Alex. (1969). Test-battery to Measure Aging - Rate in Man, The Lancet, Vol. 6, pp. 1411-1414.

Hershey, Daniel. (1980). A New Age-Scale for Humans. Lexington Books.

Hocking, R.R. (1976). The Analysis and Selection of Variables in Linear Regression. Biometrics, 32, pp. 1-50.

Lawless, J.F., Singhal, K. (1978). Efficient Screening of Nonnormal Regression Models. Biometrics, 39, pp. 318-327.

Palmore, E., Jeffers, F. C. (1971). Prediction of Life Span. Heath Lexington Books.

Rao, C. R. (1968). Linear Statistical Inference and Its Applications. Wiley, New York.

Robins, L, Hall, J.R. (1970). How to Practice Prospective Medicine. Public Health Service.

Rockstein, M. (1974). Theoretical Aspects of Aging. Academic Press.

Ross, C. L. (1971). Predicting Longevity. Heath Lexington Books.

Shurtleff, Dewey. (1974). An Epidemiological Investigation of Cardiovascular Disease. U.S. Government.