THE USE OF A PIECEWISE CONSTANT PROPORTIONAL HAZARDS MODEL IN ISSUES RELATED TO STANDARDISATION : A REVIEW AND SOME RESULTS

M. Rafiq Shah

Vrije Universiteit, Brussels

IPD-Working Paper 1987-2

Acknowledgements

I thank Prof. Dr. R. Lesthaeghe and the A.B.O.S. officials for financial assistance in the completion of this work. Thanks are due to Dr. S. Wijewickrema and Mr. C. Vanderhoeft for helpful comments and suggestions. Finally, I thank Mr. C. Vanderhoeft for his help in skilful typing and in the pubication of this paper.

ORDERING INFORMATION

Extra copies of this working paper are available from :

Mr. E. Vanden Balck
Secretary/Librarian
Interuniversity Programme in Demography
c/o Centrum voor Sociologie
Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussels
BELGIUM

at the cost of US $ 6.50 or BF 250 . Please attach check or international money order.

ORDER FORM

Name ....
Address ....
....
....
....
....

wishes to receive ...... copies of Working Paper Nr..........
and includes payment of US $ .... or BF........

THE USE OF A PIECEWISE CONSTANT PROPORTIONAL HAZARDS MODEL IN ISSUES RELATED TO STANDARDISATION : A REVIEW AND SOME RESULTS

Dr. M. Rafiq Shah *

1. INTRODUCTION

Standardisation is an important tool of comparison in demographic analysis. It is applicable in connection with both longitudinal and cross-sectional types of data; and in dealing with subjects like mortality, fertility, labour force analysis, evaluation of family planning programmes, to quote a few examples.

Due to problems associated with conventional standardisation in the comparison of more than two groups in question (Clogg, 1978 : 527), mathematical models could be used as an alternative for the purpose of finding adjusted or standardised rates. Clogg (1978) proposed a saturated multiplicative model which like Teachman (1977) and others could be regarded as solving the problems related to a summary measure used in direct standardisation (Clogg, 1978 : 523).

Many others have used multiplicative models for the purpose of indirect standardisation. For instance, in the analysis of vital rates in connection with rare events, or in situations where the age-specific data were not available, multiplicative models were suggested by Stouffer (1951), Kilpatrick (1962), Mantel and Stark (1968), Osborn (1975), Breslow and Day (1975), Gail (1978) and others.

Hoem (1979) has used a special kind of Cox' (1972) model, which will be called the piecewise constant proportional hazards model. At first sight, the model seems to be applicable in both direct and indirect standardisation (Hoem, 1979 : 12).

* This matter was prepared while M. Rafiq Shah was working on a doctoral dissertation in the Interuniversity Programme in Demography at the Vrije Universiteit, Brussels, Belgium.

Seemingly interesting and perspicuous application of statistical theory to demographic events and rates, a thorough examination of Hoem's work is the subject matter of this note. Some objectives that could be achieved after such an examination are as follows.

a. We will be able to demonstrate that the analysis made according to the constant proportional hazards model used by Hoem (1979) is in fact identical to the use of a multiplicative model, where the numbers of events are assumed to have a Poisson error structure.

b. After having shown the equivalence of results (Relative risks) obtained by using the classical multiplicative model, and the one used by Hoem (1979), it will be possible to reinterpret the parameters of the multiplicative model in terms of conventional standardised rates.

c. A critical appreciation of the model will be made to understand if the model used by Hoem could really be used for the purpose of both direct and indirect standardisation.

Since the model proposed by Hoem (1979) is a special case of the proportional hazards model, we shall begin with the piecewise constant proportional hazards model. The present discussion as a whole could be divided into four parts as follows.

First, the definition of a piecewise constant proportional hazards model will be followed by the derivation of a piecewise log-likelihood function (log L). Then it will be shown that (log L) is equivalent to the one obtained from observations having a Poisson error structure in a contingency table perspective. Later on, a log-linear model will be fitted on the data used by Hoem (1979) by using GLIM (Generalized Linear Interactive Modelling; Release 3, Baker and Nelder, 1978) for the purpose of illustration and new interpretation of the parameters of a classical multiplicative model. Finally, the use of the proposed model will be commented in the light of various issues of standardisation; in particular to examine the claim that the proposed model solves the issues of both direct and indirect standardisation.

2. THE PIECEWISE CONSTANT PROPORTIONAL HAZARDS MODEL

2.1. Definition

The proportional hazards model is based on the pioneering work of Cox (1972). The model has been developed further by Breslow (1972, 1974), Holford (1976), Gross and Clark (1975), Kalbfleisch and Prentice (1980) among others.

The application of piecewise constant proportional hazards models in the design and analysis of clinical trials can be found in Peto *et al.* (1976,1977). Laird and Olivier (1980) have demonstrated that piecewise constant proportional hazards models which are used basically for survival data analysis are identical to the models used in the analysis of contingency tables. Menken *et al.* (1981) have used the model for the analysis of socio-economic influence on marriage dissolutions in the USA. More recently, Trussel *et al.* (1983) have used the model for the analysis of covariates related to infant and child mortality in Sri Lanka.

To begin with, let us denote by $U(\underline{t}_i, Z_i)$ the hazard rate at duration $\underline{t}_i$ for an individual i with a known set of covariates Z_i. According to the proportional hazards model, the hazard rate $U(\underline{t}_i, Z_i)$ can be written as the product of two functions, one depending merely on time $\underline{t}_i$, another depending merely on covariates Z_i :

$$U(\underline{t}_i, Z_i) = U(\underline{t}_i). \exp(\beta' Z_i). \qquad (1)$$

$U(\underline{t}_i)$ is called the base-line hazard at duration $\underline{t}_i$; β' is a vector of parameters to be estimated.

Unlike the constant hazard model, where $U(\underline{t}_i)$ is assumed to remain constant over the whole time range, the piecewise constant hazard model assumes that the hazard rate is constant in time intervals $[t_{j-1}, t_j)$ ($j=1,2,\ldots\ldots J$). It is thus assumed that there exists a set of constants a_j ($j=1,\ldots J$) such that

$$U(\underline{t}_i) = \exp(a_j) \qquad \text{if } \underline{t}_i \in [t_{j-1}, t_j). \qquad (2)$$

The constants a_j are unknown parameters and should thus be estimated.

Substitution of (2) in (1) gives

$$U(\underline{t}_i, Z_i) = \exp(a_j). \exp(\beta' Z_i) \qquad \text{if } \underline{t}_i \in [t_{j-1}, t_j). \tag{3}$$

Using the notations of conventional life table analysis, the probability that an individual with covariates Z_i survives upto time $\underline{t}_i$ is :

$$\begin{aligned} S(\underline{t}_i, Z_i) &= \exp(-\textstyle\int_0 U(s, Z_i)ds) \\ &= \exp[-\Lambda(\underline{t}_i). \exp(\beta' Z_i)] \end{aligned} \tag{4}$$

where $\Lambda(\underline{t}_i) = \int_0 U(s)ds$ is the cumulative base-line hazard. Note that the probability that an individual dies in a small interval $[\underline{t}_i, \underline{t}_i+ds)$ approximately equals

$$S(\underline{t}_i, Z_i).U(\underline{t}_i, Z_i).ds. \tag{5}$$

2.2. The (log-)likelihood function

One of the procedures for estimating the parameters of a specified model is the method of maximum likelihood. According to this method, the partial derivatives of the log-likelihood function are set equal to zero and the so obtained system of equations is solved. We derive here the log-likelihood function for the piecewise proportional hazards model.

Let us assume that we have n individuals under observation. Let δ_i be an indicator variable, with $\delta_i=1$ if the i-th individual dies (or fails) at time $\underline{t}_i$, and $\delta_i=0$ if the i-th individual is censored at time $\underline{t}_i$. The contribution to the likelihood of the individuals who die is

$$S(\underline{t}_i, Z_i).U(\underline{t}_i, Z_i),$$

and the contribution of the censored individuals is

$S(\underline{t}_i, Z_i)$.

In general the contribution of the i-th individual can thus be written as

$$S(\underline{t}_i, Z_i).U(\underline{t}_i, Z_i)^{\delta_i}$$

and the likelihood function is thus (proportional to)

$$L = \Pi_i\, S(\underline{t}_i, Z_i).U(\underline{t}_i, Z_i)^{\delta_i} \qquad (6)$$

Taking the natural logarithms of both sides of (6) and substituting the relative expressions defined in (3) and (4) yields the log-likelihood

$$\log L = \Sigma_i\, \delta_i.\log U(\underline{t}_i) + \Sigma_i\, \delta_i.(\beta' Z_i) - \Sigma_i\, \Lambda(\underline{t}_i).\, \exp(\beta' Z_i) \qquad (7)$$

Heavy algebraic manipulations reduces (7) to the most commonly used form :

$$\log L = \Sigma_{z \in Z}\, \Sigma_j\, [a_j . D_j(z) + \beta' z . D_j(z) - \exp(a_j + \beta' z).(\Sigma_{i \in \mathfrak{R}_j(z)}\, E_{ij})] \qquad (8)$$

where

Z = the set of different covariate vectors recorded;

$D_j(z)$ = the number of individuals with covariates z who die (fail) in the interval $[t_{j-1}, t_j)$;

$\mathfrak{R}_j(z)$ = the set of individuals with covariates z who are still at risk at time t_{j-1}. Note that $\mathfrak{R}$ denotes a set and R stands for a figure. Formally $\mathfrak{R}_j(z) = \{i \in \mathfrak{R}_j | Z_i = z\}$ (where $\mathfrak{R}_j$ is the set of individuals at risk at the beginning of the jth interval, i.e. at time t_{j-1});

E_{ij} = the exposure time of the ith individual in the jth interval.

The computation of E_{ij} for discrete data requires some assumptions regarding the occurence of the events in the specified interval. For an interval of unit length we will assume, following Menken *et al.* (1981 : 183) :

$$
\begin{aligned}
E_{ij} &= 1 && \text{if } \underline{t}_i \geq t_j \\
&= 1/2 && \text{if } \underline{t}_i \in [t_{j-1}, t_j) \\
&= 0 && \text{if } \underline{t}_i < t_{j-1}.
\end{aligned}
$$

For further simplifications the following notation can be useful :

$$R_j(z) = \Sigma_{i \in R_j(z)} E_{ij}.$$

In the following section, it is shown that the log-likelihood function in (6) is equivalent to the one obtained by assuming a Poisson distribution for the observations in a contingency table.

2.3. The equivalence with log-linear models

Let $D_j(z)$, $(j=1,\dots J;\ z \in Z)$ be a discrete dependent variable having a Poisson distribution with mean $M_j(z)$, and consider the following log-linear model for $M_j(z)$:

$$M_j(z) = R_j(z) \,.\, \exp(a_j + \beta' z). \tag{9}$$

The log-likelihood function for this model is :

$$\log L = \Sigma_{z \in Z}\ \Sigma_j\ [D_j(z).\log(M_j(z)) - M_j(z) - \log(D_j(z)!)].$$

Substitution of the expression for $M_j(z)$ as specified in (9) finally yields the log-likelihood function

$$
\begin{aligned}
\log L = \Sigma_{z \in Z}\ \Sigma_j\ [D_j(z).(a_j + \beta' z) - R_j(z).\exp(a_j + \beta' z) \\
+ D_j(z).\log(R_j(z)) - \log(D_j(z)!)].
\end{aligned} \tag{10}
$$

The log-likelihoods in (8) and (10) differ only by the term $\Sigma_{z \in Z} \Sigma_j D_j(z).\log(R_j(z)) - \log(D_j(z)!)$. This term, however, is independent of the parameters a_j and β and may therefore be dropped. Due to this equivalence, we may use GLIM (or any other statistical package for log-linear models analysis) in order to estimate the unknown parameters of the piecewise constant proportional hazards model. Before such an attempt, however, we describe the use of the proportional hazards model in standardisation.

3. PROPORTIONAL HAZARDS MODELS AND STANDARDISATION

3.1. General remarks

Terms employed in connection with the proportional hazards models used in survival analysis are analogous to the terms used in conventional life tables. The difference is that, in the former, the population concerned is no longer considered as a homogeneous group--the risk of death/failure depends on the characteristics of the subgroups constituting the whole population. A case in point is the risk of death classified by age and the regional place on the marital status of a person as is commonly used in multistate demographic analysis. Instead of "death", the general term "failure" is usually used in view of its extention to other areas such as fertility and migration.

Since proportional hazards models are currently employed for estimating the effects of various covariates where each constitutes a subgroup, it could be used also for the purpose of comparing these groups through what are called standardised rates in conventional sense. Being a statistical tool of demographic comparison, standardisation is thus logically connected with the proportional hazards model for the purpose of comparing the subgroups in a given setting.

3.2. Use of the model in standardisation

For the sake of simplicity let us limit our discussion to mortality analysis and borrow the notation (with some alterations) used by Hoem (1979). The piecewise constant proportional hazards model that could be used for the comparison of mortality in subgroups k of a population is :

$$U_k(x) = \theta_k . U(x) \qquad (11)$$

where $U_k(x)$ is the death rate at age x (in completed years) for the k-th group. ($U_k(x)$ and $U(x)$ are usually called the forces of mortality.) θ_k is the group-specific rate independent of x, and $U(x)$ is the age-specific rate of all the groups combined.

Let

$D_k(x)$ = the number of deaths at age x in the k-th group,

$R_k(x)$ = persons exposed to the risk of death at age x in the kth group ($k = 1,2,\ldots K$; $x = 1, 2, \ldots X$),

$D_k = \Sigma_x D_k(x)$, $D(x) = \Sigma_k D_k(x)$ and $D = \Sigma_x \Sigma_k D_k(x)$.

Similarly,

$R_k = \Sigma_x R_k(x)$, $R(x) = \Sigma_k R_k(x)$ and $R = \Sigma_x \Sigma_k R_k(x)$.

Comparing (11) with (3), we note that

$\theta_k = \exp(\beta' Z)$ and $U(x) = \exp(a_j)$,

where Z are the covariates for the k-th subgroup and x is in the j-th time-interval.

The log-likelihood function of the model (3) is obtained by resetting the subscripts in (8) or (10) (and omitting the constant term). Accordingly,

$$\log L = \Sigma_k \Sigma_x [D_k(x).\log(U(x)) + D_k(x).\log(\theta_k) - U(x).R_k(x).\theta_k] \qquad (12)$$

The maximum likelihood estimators $\hat{\theta}_k$ and $\hat{U}(x)$, obtained by differenciating log L satisfy the following equations :

$$\hat{\theta}_k = \frac{D_k}{\Sigma_x \hat{U}(x).R_k(x)} \qquad (13)$$

and

$$\hat{U}(x) = \frac{D(x)}{\Sigma_k \hat{\theta}_k . R_k(x)}. \qquad (14)$$

Estimates of θ_k and $U(x)$ satisfying (13) and (14) could be obtained by iteration. For instance, starting with $\theta_k = 1$ in the first cycle of the iteration, the estimates of the parameters are :

$$\hat{U}(x)^{(1)} = \frac{D(x)}{\Sigma_k R_k(x)} = \frac{D(x)}{R(x)}$$

and

$$\hat{\theta}_k^{(1)} = \frac{D_k}{\Sigma_x \hat{U}(x)^{(1)} . R_k(x)} .$$

This process is continued untill stable values of the estimates are obtained.

Let us assume that the estimators have settled down to their maximum likelihood estimates of $\hat{\theta}_k^{(n)}$ and $\hat{U}(x)^{(n)}$, (the superscipt n shows the n-th cycle of iteration). Accordingly, these estimates satisfy the equations

$$\hat{\theta}_k^{(n)} = \frac{D_k}{\Sigma_x \hat{U}(x)^{(n)} . R_k(x)} \qquad (15)$$

and

$$\hat{U}(x)^{(n)} = \frac{D(x)}{\Sigma_k \hat{\theta}_k^{(n)}.\ R_k(x)} \ . \qquad (16)$$

The values obtained in each step of the iteration are sometimes normalized for the purpose of improving convergence of the procedure (Mentel and Stark, 1968; Ireland and Kulback, 1968). The estimates of the parameters obtained in the n-th cycle of the iteration, $\hat{\theta}_k^{(n)}$ and $\hat{U}(x)^{(n)}$, are normalized by the following rescaling factors, C and D, according to Hoem (1979) :

$$C = \Sigma_x \hat{U}(x)^{(n)}.R(x),$$

and

$$D = \Sigma_x U(x)^o .R(x).$$

(C and D are the expected and observed number of deaths in the pooled groups respectively). The normalized maximum likelihood estimates of the parameters are :

$$\theta_k^* = \hat{\theta}_k^{(n)}.(C/D) \qquad (17)$$

and

$$U(x)^* = \hat{U}(x)^{(n)}.(D/C). \qquad (18)$$

We may examine now the use of the estimates $\hat{\theta}_k^{(n)}$ and $\hat{U}(x)^{(n)}$, and the normalized estimates θ_k^* and $U(x)^*$, which are all indices used in conventional standardisation.

3.3. Model estimates and conventional indices

Similar to other iterative procedures used by Mantel and Stark (1968) , Breslow and Day (1975) among others, the stable values of the estimators, $\hat{\theta}_k^{(n)}$ (see 15), could be used as an index relating to indirect standardisation. For instance, the numerator of $\hat{\theta}_k^{(n)}$ is the number of observed deaths in k-th group, and the denominator is the number of expected deaths in the same group according to the estimated schedule of age-specific death rates $\hat{U}(x)^{(n)}$. Thus $\hat{\theta}_k^{(n)}$ is the estimate of the Standardised Mortality Ratio (SMR).

The normalized estimate of the parameter, θ_k^* (17) could be related to an index used in connection with direct standardisation. Substitution of the values of C and D in (17) yields :

$$\theta_k^* = \hat{\theta}_k^{(n)} \cdot \frac{\Sigma_x \hat{U}(x)^{(n)}.w(x)}{\Sigma_x U(x)^o.w(x)}$$

where $w(x) = R(x)/R$.

Recalling that $\hat{U}_k(x) = \hat{\theta}_k^{(n)}. \hat{U}(x)^{(n)}$, we get

$$\theta_k^* = \frac{\Sigma_x \hat{U}_k(x).w(x)}{\Sigma_x U(x)^o.w(x)} . \qquad (19)$$

θ_k^* is thus the ratio of the estimated number of deaths in the k-th group to the total number of observed deaths in all the groups combined. Note that the age-specific rates of the k-th group, $\hat{U}_k(x)$, are estimated according to the model, and the age-composition of the combined groups is used as a standard. The estimate θ_k^* is thus a "special type" of Comparative Mortality Factor/Figure (CMF); since CMF is defined as the ratio of expected deaths (obtained from the observed age-specific death rates of the k-th group and the population composition of all the groups combined), to the crude rate of the combined groups (Spiegelman, 1968 : 219).

An interesting application of (19) is the measurement of mortality differences of the groups under comarison by the ratio of the "special type" of CMFs which will be called "model" CMF's and denoted by ${}^{m}CMF$. For the comparison of two groups (suffixes 1 and 2) for instance, and with the notation already introduced, this ratio is :

$$\frac{{}^{m}CMF_1}{{}^{m}CMF_2} = \frac{\Sigma_x \hat{U}_1(x).w(x)}{\Sigma_x \hat{U}_2(x).w(x)} = \frac{\theta_1{}^*}{\theta_2{}^*} .$$

4. ILLUSTRATION

For illustrative purposes, let us consider the data used by Hoem (1979), as shown in Table 1. In this example the risk factor is associated with marital status (having two categories: bachelors and married German males), and the demographic phenomena under study is mortality.

As shown in Table 2, both crude and age-specific rates indicate that mortality varies by marital status and age. The crude rates of bachelors and married men are .00517 and .00300 respectively; giving a crude relative risk of death of 1.72, indicating that men who remain bachelors have 72% higher mortality than married men.

Although the trend of age-specific rates of bachelors and married men is similar in Table 2 (the death rates by age descreases), the pattern of the age distribution of their respective populations differ from one to another, i.e. the proportion of married men increases, while that of bachelors decreases by age. Since their crude rates hide the dissimilar pattern of the age structures, a standardised summary measure is required for the purpose of meaningful comparison of mortality between bachelors and married men. Note that such a situation calls for the use of direct standardisation according to conventional analysis.

In the following section we show how the estimates of the parameters of the proposed model (11) could be used as standardised indices for such comparisons.

4.1. Estimation of the parameters

The parameters of the model (11) were estimated by iteration (equations (15) and (16)). It may be noted that being maximum likelihood estimates, they satisfy (13) and (14) too. Of these estimates, $\hat{\theta}_k^{(n)}$ and its normalized form θ_k^*, which could be interpreted as summary measures of group-specific rates, will be the core of our discussion in the following sections.

For bachelors (k=1) and married men (k=2), the values obtained by iteration are : $\hat{\theta}_1^{(n)} = 1.2134$ and $\hat{\theta}_2^{(n)} = .6588$ respectively. This gives a relative risk of $1.842 = \hat{\theta}_1^{(n)} / \hat{\theta}_2^{(n)}$.

Table 1. Constructed number of deaths and person years of german males by age and marital status.

AGE	DEATHS			PERSON YEARS		
	SINGLE	MARRIED	TOTAL	SINGLE	MARRIED	TOTAL
x	$D_1(x)$	$D_2(x)$	$D(x)$	$R_1(x)$	$R_2(x)$	$R(x)$
22	433	24	457	91,444	8,556	100,000
23	412	36	448	86,835	12,708	99,543
24	337	66	439	75,892	23,203	99,095
25	331	102	433	63,241	35,415	98,656
26	287	138	425	52,023	46,207	98,223
27	242	171	413	42,123	55,675	97,798
28	215	185	400	36,915	60,470	97,385
29	192	200	392	32,215	64,770	96,985
TOTAL	2,485 (D_1)	922 (D_2)	3,407 (D)	480,688 (R_1)	307,004 (R_2)	787,685 (R)

Source : Hoem (1978 : 29); Table 3 and Table 4 with subscripts of the notation by marital status reversed.

Table 2. Proportions, crude and age-specific death rates of males by marital status.

AGE	DEATH RATES			PROPORTIONS	
	SINGLE	MARRIED	TOTAL	SINGLE[a]	MARRIED[b]
x	$U_1^o(x)$	$U_2^o(x)$	$U^o(x)$		
22	.004735	.002805	.00457	.190	.028
23	.004745	.002833	.00450	.181	.041
24	.004915	.002844	.00443	.158	.076
25	.005234	.002880	.00439	.132	.115
26	.005517	.002987	.00433	.108	.151
27	.005745	.003071	.00422	.088	.181
28	.005824	.003059	.00411	.077	.197
29	.005960	.003088	.00404	.067	.211
Crude Rates	.005170	.003000	.00433	1.001	1.000

Crude Relative Risk = .00517/.003 = 1.72

Source : Table 1 : a=91444/480688,... ; b=8556/307004,...

After normalization (see (17) and (18)), the estimates are : $\theta_1^* = 1.228$ and $\theta_2^* = .666$, giving a relative risk of $1.844 = \theta_1^*/\theta_2^*$.

It may be noted that the relative risks obtained from $\hat{\theta}_k^{(n)}$ and the normalized estimate θ_k^* should be identical, since the normalization factor is eliminated in the process of comparison. Their difference, accounted for here, may be attributed to rounding errors.

Besides the method of iteration that enables one to find the estimates of the parameters of the model, we may use GLIM for estimating the parameters of a specified model involving particularly large data sets. Other reasons for using GLIM in the data of our illustration are described in the following section.

4.2. The use of GLIM

The establishment of theoretical links between the log-likelihood function of the hazards model and the log-likelihood function for multiplicative models of contingency tables (equations (8) and (10)) encourages us to treate the data of Table 1 as a contingency table and estimate the parameters by using GLIM.

GLIM was used also due to the availability of two types of normalization procedures commonly known as the constraints imposed on the parameters of the proposed model. We call these constraints "GLIM" and "Usual" constraints. The purpose of using these constraints is to reinterpret the parameters of the multiplicative models under construction in the present discussion in terms of indices commonly used in standardisation.

According to the "GLIM" constraints, the parameters of the first row and first column of an IxJ table having two factors A and B, say, are unity as follows (Plackett, 1974) :

$$w_1^A = w_1^B = w_{1j}^{AB} = w_{i1}^{AB} = 1,$$

where the w's denote the parameters of a multiplicative model: w_i^A and w_j^B are the main effects of factors A and B, and w_{ij}^{AB} stands for the interaction effects of A and B.

Their counter parts in the log-linear model are :

$$U_1^A = U_1^B = U_{1j}^{AB} = U_{i1}^{AB} = 0,$$

where

$$U_i^A = \log(w_i^A),\ U_j^B = \log(w_j^B) \text{ and } U_{ij}^{AB} = \log(w_{ij}^{AB}).$$

Recent discussion regarding relationships and use of multiplicative and log-linear models could be found in the work of Frans Willekens (1981, 1982).

According to this model specification and the constraints mentioned above, the subgroup having the parameter which has been assigned the value of unity (i.e, w_1=1,...) is considered as the reference or standard group, and other parameters measure the relative risk.

The computer listing of our illustration according to the "GLIM" constraints (Appendix) shows estimates of the parameters of a log-linear model. Since we have two groups under comparison, the value of only one estimate, U_2^B, is shown (the value of the other estimate is zero according to the "GLIM" constraints). Note that in the computer listing U_2^B is denoted as K2.

The value of K2 is -.6111, giving w_2 = .5428. Since w_1 =1 (according to "GLIM" constraints), the standardised relative risk (SRI) is equal to 1.842 = 1/.5428. Comparing SRI with the crude relative risk of 1.72 (Table 2), we see that, after model corrections for age differences, men who remain bachelors have about 7% (i.e., 1.842-1.720) higher risk of death.

Since the relative risk w_1/w_2 is found to be the same as $\hat{\theta}_1^{(n)}/\hat{\theta}_2^{(n)}$, there is reason to interpret w_1 and w_2 as the SMRs of group 1 and group 2 respectively. Note that $\hat{\theta}_k^{(n)}$ is obtained through iteration and equals the SMR of the kth group. This argument may not, however, sound correct, since the value of w_1 or w_2 is not exactly equal to $\hat{\theta}_1^{(n)}$ or $\hat{\theta}_2^{(n)}$

respectively. This might be due to the reason, that the value of w_1 has been arbitrarily set equal to 1 (according to "GLIM" constraints). We therefore now consider the "Usual" constraints, which are slightly less restrictive.

Bishop *et al.* (1975), Goodman (1970, 1971a, and 1971b), Payne (1977) and Everitt (1977) among others use the following constraints which are commonly known as "Usual" constraints :

$$\Pi_i w_i^A = \Pi_j w_j^B = \Pi_i w_{ij}^{AB} = \Pi_i w_{ij}^{AB} = 1$$

for multiplicative models and

$$\Sigma_i U_i^A = \Sigma_j U_j^B = \Sigma_i U_{ij}^{AB} = \Sigma_i U_{ij}^{AB} = 0$$

for log-linear models. The relationship between the w's and the U's is the same as that stated according to "GLIM" constraints.

In the computer listing working with "Usual" constraints (Appendix), the value of U_2^B (i.e., K2) is -.3056, and therefore according to the constraints, U_1^B (i.e., K1) is equal to .3056; giving w_1 and w_2 equal to 1.357 and .737 respectively. The standardised relative risk w_1/w_2 is equal to 1.842 = 1.357/.737, which is similar to the one obtained according to "GLIM" constraints.

The values of w_1 and w_2 according to "Usual" constraints are somewhat similar to the observed SMR's (based on the age-specific mortality rates of both groups combined as a standard, i.e., 1.178 for group 1 and .709 for group 2). We call these estimates somewhat similar to SMR's of the groups concerned, because unlike the "GLIM" constarints w_1 is not arbitrarily equal to unity.

Therefore when required, the parameters related to an additive log-linear model, i.e., w_1 and w_2, under the "Usual" constraints in particular, could be interpreted as estimates of the SMR's of the groups under comparison.

5. SOME CONCLUSIONS

The examination of the piecewise constant proportional hazards model and its equivalent log-linear model, together with the numerical excercise (the present exercise consists of comparing mortality of two groups) suggest the following results and conclusions.

1. Mortality differences between subgroups are measured by standardised ratios. The estimators that are apt for the purpose of such standardised comparisons are :

a. $\hat{\theta}_k^{(n)}$ - we may call it model SMR, and denote it by ${}^{m}SMR$. It is based on the estimated age-specific rates of all the groups combined, $\hat{U}(x)^{(n)}$ - used as a standard. For two groups (k = 1,2), their ratio measures the realtive risk. We may call it the Standardised Relative Risk, and denote it by SRR^1. Using $\hat{\theta}_k^{(n)}$ and the related notation, SRR^1 is :

$$SRR^1 = {}^{m}SMR_1 / {}^{m}SMR_2.$$

b. θ_k^* - A special type of CMF for the k-th group - ${}^{m}CMF$. We call it model CMF since the age-specific rates of each group are estimated through the model, and the proportions of both groups combined is used as a set of standards weights. hence, in this estimator two standards are employed, i.e. $\hat{U}_k(x)$ and $R(x)$. We shall come to this point later on. For two groups under comparison, the ratio of their estimated CMF is the SRR, that is :

$$SRR^2 = {}^{m}CMF_1 / {}^{m}CMF_2.$$

c. By using GLIM we have introduced the k-th column effect parameter, w_k. We have indicated that w_k is somewhat closer to the observed SMR of the k-th group under the "Usual" constraints. For the two groups under comparison in our illustration the ratio of the estimates gives SRR, i.e.,

$$SRR^3 = w_1 / w_2.$$

An important conclusion is that, by whatever name we call the estimators mentioned above, the standardised relative risk (SRR) of these estimators is identical, i.e., measuring identical differences of mortality between the groups under comparison. As we have seen in our illustration the relative risk obtained by using a, b, or c is 1.84, showing that the mortality of group 1 is 84% higher than group 2. We may therefore not conclude that the model could be used for both direct and indirect standardisation.

2. The model provides basically an alternative to indirect standardisation which is recommended when the age-specific data relating to the phenomena under study, are either totally or partially not available. In this respect, the model provides the best alternative to indirect standardisation. Note that for estimating the estimators $\hat{\theta}_k^{(n)}$ and $\hat{U}(x)^{(n)}$ by iteration, we require the data on the total number of deaths in the k-th group, D_k, the population composition of the k-th group $R_k(x)$, and the deaths by age of all the groups combined, $D(x)$, showing that the age-specific death data are not required for the groups under comparison.

An interesting property of the estimator $\hat{\theta}_k^{(n)}$ is that SRR^1 could be used to show the relationship between crude rates and expected rates of the groups under comparison. For instance, as already defined SRR^1 is

$$SRR^1 = \frac{\hat{\theta}_1^{(n)}}{\hat{\theta}_2^{(n)}} = \frac{D_1}{D_2} \cdot \frac{\Sigma_x \hat{U}(x)^{(n)}.R_2(x)}{\Sigma_x \hat{U}(x)^{(n)}.R_1(x)}$$

$$= \frac{\Sigma_x U_1(x)^o.R_1(x)}{\Sigma_x U_2(x)^o.R_2(x)} \cdot \frac{\Sigma_x \hat{U}_2(x).R_2(x)}{\Sigma_x \hat{U}_1(x).R_1(x)} \cdot \frac{\hat{\theta}_1^{(n)}}{\hat{\theta}_2^{(n)}}$$

Where $U_k(x)^o$ is the observed age-specific death rate of the k-th group and $\hat{U}_k(x) = \hat{\theta}_k^{(n)}.\hat{U}(x)^{(n)}$.

Equation (20) holds if

$$\frac{\Sigma_x U_1(x)^o.R_1(x)}{\Sigma_x U_2(x)^o.R_2(x)} \cdot \frac{\Sigma_x \hat{U}_2(x).R_2(x)}{\Sigma_x \hat{U}_1(x).R_1(x)} = 1.$$

One of the conditions for this equality to be true is that the number of observed deaths in group k, $\Sigma_x U_k(x)^o.R_k(x)$, equals the number of expected deaths, $\Sigma_x \hat{U}_k(x).R_k(x)$. This is an important property of the estimates obtained in contingency table analysis and bi-proportional adjustment procedures. Comparison of the results in Table 2 and Table 3 shows that this is indeed the case. Note that in these tables, instead of total number of deaths, the crude and expected death rates are shown to be equal.

For the purpose of comparison, therefore, the estimated age-specific death rates $\hat{U}_k(x)$ have to be weighted by a different set of weights (other than $R_1(x)$ and $R_2(x)$). One set of such weights is the population composition of both groups combined, i.e. $R(x)$ or $w(x) = R(x)/R$. If the weights $w(x)$ are used along with the estimated age-specific rates $\hat{U}_k(x)$, we could easily find SRR^2.

It follows that, when we speak of normalized estimators θ_k^* or the related index SRR^2, we use in fact two standards, namely $U_k(x)$ estimated from the model (11) that is commonly used for indirect standardisation, and afterwards apply the population composition of both groups combined ($w(x)$) as a standard - a procedure that is normally used in direct standardisation.

Therefore, we may not say that the proposed model (11) is applicable to both direct and indirect standardisation. We should rather say that the method proposed uses both types of conventional standardisation procedures when required.

3. It is interesting to note that if we had not used the estimated values ($\hat{U}_k(x)$) - obtained after using the model (11), but instead the observed age-specific death rates of the groups under comparison ($U_k^o(x)$) together with the population composition of the combined groups ($w(x)$), the conventional standardised risk denotes by SRR^c for the two groups in question would have been

Table 3. <u>Estimated age-specific rates and proportions of persons exposed to the risk of death by marital status.</u>

AGE	ESTIMATED RATES [a]		PROPORTIONS [b]	
	SINGLE	MARRIED	SINGLE	MARRIED
x	$\hat{U}_1(x)$	$\hat{U}_2(x)$	$w_1(x)$	$w_2(x)$
22	.00476	.00258	.190	.028
23	.00478	.00259	.181	.041
24	.00496	.00269	.158	.076
25	.00525	.00285	.132	.115
26	.00551	.00299	.108	.151
27	.00571	.00310	.088	.181
28	.00573	.00311	.077	.197
29	.00582	.00316	.067	.211
			1.001	1.000

$$\Sigma_x \, w_1(x) \, . \, \hat{U}_1(x) = .00517497 = .00517$$

$$\Sigma_x \, w_2(x) \, . \, \hat{U}_2(x) = .00300264 = .00300$$

a : From computer listings as shown in the Appendix

b : From Table 1 : $w_1(x) = R_1(x)/ R_1$ and $w_2(x) = R_2(x)/ R_2$

Table 4. Observed age-specific death rates by marital status, and proportions of persons exposed in both groups combined.

AGE	OBSERVED RATES		PROPORTIONS
	SINGLE	MARRIED	COMBINED
x	$U_1^\circ(x)$	$U_2^\circ(x)$	$w(x)$
22	.004735	.002805	.127
23	.004745	.002833	.126
24	.004915	.002844	.126
25	.005234	.002880	.125
26	.005517	.002987	.125
27	.005745	.003071	.124
28	.005824	.003059	.124
29	.005960	.003088	.123
Standardised Rates			1.000

$$SRR^c = \frac{\Sigma_x\, w(x) \,.\, U_1^\circ(x)}{\Sigma_x\, w(x) \,.\, U_2^\circ(x)} = \frac{.005330}{.002945} = 1.8098$$

a : From computer listings as shown in the Appendix

b : From Table 1 : $w_1(x) = R_1(x)/\ R_1$ and $w_2(x) = R_2(x)/\ R_2$

$$SRR^{o} = \frac{\Sigma_x w(x).U_1{}^{o}(x)}{\Sigma_x w(x).U_2{}^{o}(x)} .$$

Using the data of our illustration, the SRR^{o}, i.e. based on observed data, is 1.81 as shown in Table 4. This shows that the mortality of group 1 (bachelors) is 81% higher than the mortality of group 2 (married men). Note that SRR^{o} is in fact the ratio of CMF_1 and CMF_2, both of which are based on observed data.

But CMF like other directly standardised measures depends on the standard used. For instance, when the population distribution of group 2 ($w_2(x) = R_2(x)/R_2$) is used instead of $w(x)$, the relative risk goes up to 1.86, showing that the mortality of group 1 is 86% higher than group 2.

4. The question is why the relative risk based on the model is different from the standardised relative risk based on the observed data? The answer is found immediately by comparing the ratio of model CMF's (19) with SRR^{o} defined above. It is due to the use of estimated rates $U_k(x)$ in the former (i.e., ^{m}CMF's) and the observed rates $U_k(x)^{o}$ in the latter. The estimated rates $\hat{U}_k(x)$ are, however, closer to the observed rates as shown in the output (Appendix) unlike the estimated rates of both the groups combined (i.e. $\hat{U}(x)^{(n)}$) - see graphs in the Appendix.

5. Besides various reasons quoted earlier, we have used GLIM for introducing the parameters of an additive log-linear model that could be easily used for the parameters of its counter-part multiplicative model. The advantage of using such an multiplicative model is that the relative risk obtained from the estimates of the parameters of a multiplicative model, i.e. w's, is identical to the relative risks obtained by using the piecewise constant proportional hazards model (i.e. $w_1/w_2 = \hat{\theta}_1{}^{(n)}/\hat{\theta}_2{}^{(n...)}$).

Another reason for using GLIM is that the variance-covariance matrix of the estimates which are themselves normally distributesd is readily available, and could be used for finding the standard error of the standardised index in question.

The reason of our interest here is, however, that model (11) and the additive log-linear model fitted on the data of Table 1 (our illustration) by using GLIM are identical. The group-specific rates of model (11), i.e. θ_k, are therefore similar to the parameters w_k of the unsaturated multiplicative (additive log-linear model). Since the estimate of θ_k (i.e. $\hat{\theta}_k^{(n)}$) obtained by iteration stands for the SMR of the k-th group, the estimate of w_k may be interpreted as the SMR of the k-th group. This interpretation could be used only in the case of an additive log-linear model or its counter-part multiplicative model, i.e. the model where mortality differences are measured by a ratio estimate.

6. It has been found that the piecewise constant proportional hazards model (in theory as well as in the present illustration) provides estimates of the age-specific rates, and does not deal with the estimation of weights required (wich could be used as a standard; note that the weights used are taken from the observed data, i.e. $w(x)$). Since model (11) is similar to an additive log-linear model (unsaturated model),the saturated log-linear model seems to be a preferable candidate for solving the problems of both direct and indirect standardisation. It may be noted that such models solve the problem of selecting weights (standards) at once, because the estimates of the parameters of a saturated model are independent of the weights used. The search for a standardised index based on a saturated model is therefore required.

REFERENCES

Anderson, S.; A. Auquier; W.W. Hauck; D. Oakes; W. Vandaele; H.I. Weisberg (1980). Statistical Methods for Comparative Studies : Techniques for Bias Reduction. New York : Wiley.

Baker, R.J. and J.A. Nelder (1978). The GLIM System Manual. Release 3. Oxford : The Numerical Algorithms Group, Royal Statistical Society, London.

Bishop, Y.M.M.; S.E. Fienberg and P.W. Holland (1975). Discrete Multivariate Analysis : Theory and Practice. M.I.T. Press.

Breslow, N. (1974). Covariance Analysis of Censored Data. Biometrics, 30.

Breslow, N. (1975). Analysis of Survival Data under the Proportiuonal Hazards Model. International Statistical Review, 43.

Breslow, N. and Day (1975) Indirect Standardisation and Multiplicative Model of Rates, with Reference to the Age Adjustment of Cancer Incidence and Relative Frequency Data. Jour. Chron. Disease, 28.

Clogg, C. (1978). ADjustement of Rates Using Multiplicative Models. Demography, 15 (4).

Cox, D.R. (1972). Regression Models and Life Tables. Jour. Roy. Stastistical Soc. Series B, 34.

Everitt, B.S. (1977). The Analysis of Contingency Tables. Chapman and Hall.

Fienberg, S.E. (1977). The Analysis of Cross-classified Data. M.I.T. Press.

Gail, M. (1978). The Analysis of Heterogeneity for Indirect Standardised Mortality Ratios. Jour. Roy. Statist. Soc. Series A, 141.

Goodman, L.A. (1970). The Multivariate Analysis of Qualitative Data : Interactions among Multiple Classifications. Jour. Amer. Statist. Assoc., 65.

Goodman, L.A. (1971a). The Partitioning of Chi-square, the Analysis of Marginal Contingency Tables, and the Estimation of Expected Frequencies in Multidimensional Contingency Tables. Jour. Amer. Statist. Assoc. 66.

Goodman, L.A. (1971b). The Analysis of Multidimensional Contingency Tables : Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications. Technometrics, 13.

Gross, A.J. and V.A. Clark (1975). Survival Distributions, Reliability Applications in the Bio-Medical Sciences. New York : Wiley.

Hoem, J.M. and A. Ahlbom (1979). Statistical Analysis of Multilpicative Model and its Application to the Standardisation of Vital Rates, Part 1 : Consistency Aspects. University of Copenhagen, Laboratory of Actuarial Mathematics, Working Papere No. 21.

Holford, T.R. (1976). Life Tables with Concomitant Information. Biometrics, 32.

Holt, D. (1979). Log-linear Models for Contingency Tables : On the Interpretation of the Parameters. Sociological Methods and Research. Vol. 7, No. 3.

Ireland, C.T. and S. Kulback (1968). Contingency Tables ith Given Marginals. Biometrica, 55.

Kalbfleisch, J.D. and R.L. Prentice (1980). The Statistical Analysis of Failure Time Data. New York : Wiley.

Kilpatrick, S.J. (1962). Occupational Mortality Indices. Population Studies, 16.

Laird, N. and D. Olivier (1980). Covariance Analysis of Censored Survival Data Using Log-linear Analysis Techniques. Research Report S-59, Harvard University.

Mantel, N. and C.R. Stark (1968). Computation of Indirect Adjusted Rates in the Presence of Confounding. Biometrics, 24.

Menken, J.; J. Trussell; D. Stempel and O. Babakol (1981). Proportional Hazards Life Table Models : An Illustrative Analysis of Socio-Demographic Influence on Marital Dissolution in the U.S.A. Demography, 18.

Nelder, J.A. (1977). A Reformulation of Linear Models (with discussion). Jour. Roy. Statist. Soc. Series A, 140.

Osborn, J. (1975). A Multiplicative Model for the Analysis of Vital Rates. Appl. Statist., 24.

Payne, C. (1977) The log-linear Model for Contingency Tables. In C.A. O'Muircheartaigh and C.W. Payne (eds.) : The Analysis of Survey Data., Vol. 2; Model Fitting.

Peto, R.; M.C. Pike; P. Armitage; N.E. Breslow; D.R. Cox; S.V. Howard; N. Mantel; K. McPherson; J. Peto and P.G. Smith (1976). Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient, I. Introduction and Design. British Jour. of Cancer, 34.

Peto, R.; M.C. Pike; P. Armitage; N.E. Breslow; D.R. Cox; S.V. Howard; N. Mantel; K. McPherson; J. Peto and P.G. Smith (1977). Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient, II. Analysis and Examples. British Jour. of Cancer, 35.

Plackett, R.L. (1974). The Analysis of Categorical Data. New York : Hafner.

Spiegelman, M. (1986). Introduction to Demography. Revised Edition. Cambridge Mass : Harvard Universiry Press.

Stouffer, S.A. (1951). Standardisation of Rates when Specific Rates are Unknown. In Handbook of Statistical Methods for Demographers. Edited by A.J. Jaffe, U.S. Bureau of Census Pubications.

Teachman, J.D. (1977). The Relationship between Schoen's ∇ and a Log-linear Measure. Demography, 14, 2.

Trussell, J. and C. Hammerslough (1983). A Hazards Model Analysis of the Covariates of Infant and Child Mortality in Sri Lanka. Demography, 20.

Vanderhoeft, C. (1984). Constuction of Likelihood Functions and Maximization Through GLM - PH Model. Unpublished Manuscript, Vrije Universiteit, Brussels, Belgium.

Willekens, F. (1981). Multidimensional Population Analysis with Incomplete Data. Working Paper, No. 19. The Netherlands Interuniversity Demogrphic Institute (N.I.D.I.), Voorburg, The Netherlands.

Willekens, F. (1982). Log-linear Modelling of Spatial Interaction. Paper presented at the 2nd European Congress of the Regional Science Association, Groningen, 24-27.

APPENDICES : GLIM programs and outputs

A.1. The GLIM constraints

```
$C RELATIONSHIP BETWEEN HOEM'S MODEL (1978) AND LOG-LINEAR
$C MODELS USED FOR STANDARDISATION
$C**************************************************************
$C    *          GLIM - CONSTRAINTS                            *
$C**************************************************************
$UNITS 16
$DATA    K  X   DKX    RKX
$READ    1 22   433   91444
         1 23   412   86835
         1 24   373   75892
         1 25   331   63241
         1 26   287   52023
         1 27   242   42123
         1 28   215   36915
         1 29   192   32215
         2 22    24    8556
         2 23    36   12708
         2 24    66   23203
         2 25   102   35415
         2 26   138   46207
         2 27   171   55675
         2 28   185   60470
         2 29   200   64770
$C  K := NUMBER OF GROUPS (K=1 FOR BATCHELORS, K=2 FOR MARRIED MAN)
$C  X := AGE
$C  RKX = NUMBER OF PERSONS EXPOSED TO THE RISK OF DEATH AT AGE X
$C         IN GROUP K
$C  DKX = NUMBER OF DEATHS AT AGE X IN GROUP K
$CALC T=X-21
$FACTOR K 2 T 8
$VAR 8 XO AX RX DX MUX UX I : 2 TE DE U
$CALC LKX=%LOG(RKX)
$YVAR DKX
$OFFSET LKX
$ERROR P
$FIT T+K
$DISPLAY MAR
$EXTRACT %PE
$CALC %D=%CU(DKX) : I=%GL(8,1) :
      DX(I)=DKX(I)+DKX(I+8) : RX(I)=RKX(I)+RKX(I+8) :
      MUX=DX/RX :
      AX(I)=%FV(I)/RKX(I) :
      %C=%CU(RX*AX) :
      AX=AX*%D/%C :
      %T=%EXP(-%PE(9)) : TE(1)=%C/%D : TE(2)=TE(1)/%T :
      %A=%CU(DKX(I)) : %B=%CU(MUX(I)*RKX(I)) : DE(1)=%A/%B :
      %A=%CU(DKX(I+8)) : %B=%CU(MUX(I)*RKX(I+8)) : DE(2)=%A/%B :
      %Q=DE(1)/DE(2) :
      U=DE/TE : %U=U(1)/U(2) :
      UX=MUX/AX : XO=I+21
$PR :::
 " AGE-SPECIFIC MORTALITY RATES : " :
 " ****************************** " :
 "          AGE          OBS. RATE    EST. RATE   DISTORTION " :
 "                        K1+K2        K1+K2 "
$LOOK XO MUX AX UX
$PR : : : " PLOT OF OBSERVED (M) AND EXPECTED (A) RATES " :
 " ******************************************* " :
$PLOT MUX AX XO
$PR: : :
 " STANDARDISED MORTALITY RATIOS " :
 " ***************************** " :
 "                    K=1      K=2    (1)/(2) " :
 " HOEM'S SMR    " TE %T :
 " OBSERVED SMR" DE %Q :
 " DISTORTION    " U %U
$STOP
```

```
GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

          SCALED
 CYCLE   DEVIANCE       DF
    2      1.047         7

Y-VARIATE  DKX
ERROR POISSON  LINK LOG
OFFSET LKX

LINEAR PREDICTOR
%GM T K

       ESTIMATE         S.E.        PARAMETER
  1   -5.348         .4682E-01     %GM
  0     ZERO       ALIASED         T(1)
  2    .4930E-02     .6648E-01     T(2)
  3    .4224E-01     .6693E-01     T(3)
  4    .9898E-01     .6750E-01     T(4)
  5    .1476         .6837E-01     T(5)
  6    .1827         .6962E-01     T(6)
  7    .1874         .7070E-01     T(7)
  8    .2017         .7161E-01     T(8)
  0     ZERO       ALIASED         K(1)
  9   -.6111         .4174E-01     K(2)
   SCALE PARAMETER TAKEN AS        1.000

   UNIT  OBSERVED        FITTED        RESIDUAL
     1       433         434.9         -.9181E-01
     2       412         415.0         -.1490
     3       373         376.5         -.1815
     4       331         332.1         -.5890E-01
     5       287         286.8          .1392E-01
     6       242         240.5          .9747E-01
     7       215         211.7          .2235
     8       192         187.5          .3321
     9        24         22.09          .4074
    10        36         32.96          .5286
    11        66         62.48          .4457
    12       102         100.9          .1068
    13       138         138.2         -.2006E-01
    14       171         172.5         -.1151
    15       185         188.3         -.2370
    16       200         204.5         -.3179


 AGE-SPECIFIC MORTALITY RATES :
 ******************************
            AGE           OBS. RATE      EST. RATE    DISTORTION
                            K1+K2          K1+K2
   1       22.00         .4570E-02      .3873E-02      1.180
   2       23.00         .4501E-02      .3892E-02      1.156
   3       24.00         .4430E-02      .4040E-02      1.097
   4       25.00         .4389E-02      .4276E-02      1.026
   5       26.00         .4327E-02      .4489E-02      .9639
   6       27.00         .4223E-02      .4649E-02      .9083
   7       28.00         .4107E-02      .4671E-02      .8793
   8       29.00         .4042E-02      .4738E-02      .8530
```

PLOT OF OBSERVED (M) AND EXPECTED (A) RATES
****

```
 .480E-02 *
 .475E-02 *                                                  A
 .470E-02 *
 .465E-02 *                                      A     A
 .460E-02 *
 .455E-02 *       M
 .450E-02 *             M                  A
 .445E-02 *                   M
 .440E-02 *                         M
 .435E-02 *                                M
 .430E-02 *                         A
 .425E-02 *
 .420E-02 *                                      M
 .415E-02 *
 .410E-02 *                                            M
 .405E-02 *                   A                              M
 .400E-02 *
 .395E-02 *
 .390E-02 *             A
 .385E-02 *       A
 .380E-02 *
 .........*.........*.........*.........*.........*.........*......
         20.8      22.4      24.0      25.6      27.2      28.8
```

STANDARDISED MORTALITY RATIOS
****

	K=1	K=2	(1)/(2)
HOEM'S SMR	1.228	0.6665	1.843
OBSERVED SMR	1.179	0.7094	1.662
DISTORTION	0.9603	1.064	0.9022

A.2. The usual constraints

```
$C RELATIONSHIP BETWEEN HOEM'S MODEL (1978) AND LOG-LINEAR
$C MODELS USED FOR STANDARDISATION
$C************************************************************
$C    *              USUAL - CONSTRAINTS                      *
$C************************************************************
$UNITS 16
$DATA    K  X   DKX    RKX
$READ    1  22  433   91444
         1  23  412   86835
         1  24  373   75892
         1  25  331   63241
         1  26  287   52023
         1  27  242   42123
         1  28  215   36915
         1  29  192   32215
         2  22   24    8556
         2  23   36   12708
         2  24   66   23203
         2  25  102   35415
         2  26  138   46207
         2  27  171   55675
         2  28  185   60470
         2  29  200   64770
$C   K = NUMBER OF GROUPS (K=1 FOR BATCHELORS, K=2 FOR MARRIED MAN)
$C   X = AGE
$C   RKX = NUMBER OF PERSONS EXPOSED TO THE RISK OF DEATH AT AGE X
$C         IN GROUP K
$C   DKX = NUMBER OF DEATHS AT AGE X IN GROUP K
$CALC T=X-21
$CALC K2=%EQ(K,1)-%EQ(K,2) :
      T2=%EQ(T,1)-%EQ(T,2) :
      T3=%EQ(T,1)-%EQ(T,3) :
      T4=%EQ(T,1)-%EQ(T,4) :
      T5=%EQ(T,1)-%EQ(T,5) :
      T6=%EQ(T,1)-%EQ(T,6) :
      T7=%EQ(T,1)-%EQ(T,7) :
      T8=%EQ(T,1)-%EQ(T,8)
$VAR 8 XO AX RX DX MUX UX I : 2 TE DE U
$CALC LKX=%LOG(RKX)
$YVAR DKX
$OFFSET LKX
$ERROR P
$FIT T2+T3+T4+T5+T6+T7+T8+K2
$DISPLAY MAR
$EXTRACT %PE
$CALC %D=%CU(DKX) : I=%GL(8,1) :
      DX(I)=DKX(I)+DKX(I+8) : RX(I)=RKX(I)+RKX(I+8) :
      MUX=DX/RX :
      AX(I)=%FV(I)/RKX(I)/%EXP(%PE(9)) :
      %T=%EXP(-%PE(9)) : TE(1)=1/%T : TE(2)=%T :
      %A=%CU(DKX(I)) : %B=%CU(MUX(I)*RKX(I)) : DE(1)=%A/%B :
      %A=%CU(DKX(I+8)) : %B=%CU(MUX(I)*RKX(I+8)) : DE(2)=%A/%B :
      %Q=DE(1)/DE(2) : %T=%EXP(2*%PE(9)) :
      U=DE/TE : %U=U(1)/U(2) :
      UX=MUX/AX : XO=I+21
$PR : : :
 " AGE-SPECIFIC MORTALITY RATES : " :
 " ******************************* " :
 "          AGE         OBS. RATE    EST. RATE   DISTORTION " :
 "                        K1+K2        K1+K2 "
$LOOK XO MUX AX UX
$PR : : : " PLOT OF OBSERVED (M) AND EXPECTED (A) RATES " :
 " ******************************************* " :
$PLOT MUX AX XO
$PR: : :
 " STANDARDISED MORTALITY RATIOS " :
 " ***************************** " :
 "                        K=1       K=2    (1)/(2) " :
 " HOEM'S SMR   " TE %T :
 " OBSERVED SMR" DE %Q :
 " DISTORTION   " U %U
$STOP
```

```
GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON

        SCALED
 CYCLE  DEVIANCE       DF
   2     1.047          7

Y-VARIATE   DKX
ERROR POISSON  LINK LOG
OFFSET LKX

LINEAR PREDICTOR
%GM T2 T3 T4 T5 T6 T7 T8 K2

       ESTIMATE        S.E.         PARAMETER
   1   -5.546        .1947E-01     %GM
   2    .1032        .4518E-01     T2
   3    .6593E-01    .4511E-01     T3
   4    .9194E-02    .4505E-01     T4
   5   -.3938E-01    .4542E-01     T5
   6   -.7448E-01    .4630E-01     T6
   7   -.7918E-01    .4725E-01     T7
   8   -.9350E-01    .4806E-01     T8
   9    .3056        .2087E-01     K2
  SCALE PARAMETER TAKEN AS         1.000

  UNIT   OBSERVED        FITTED       RESIDUAL
   1        433          434.9        -.9181E-01
   2        412          415.0        -.1490
   3        373          376.5        -.1815
   4        331          332.1        -.5890E-01
   5        287          286.8         .1392E-01
   6        242          240.5         .9747E-01
   7        215          211.7         .2235
   8        192          187.5         .3321
   9         24          22.09         .4074
  10         36          32.96         .5286
  11         66          62.48         .4457
  12        102          100.9         .1068
  13        138          138.2        -.2006E-01
  14        171          172.5        -.1151
  15        185          188.3        -.2370
  16        200          204.5        -.3179


 AGE-SPECIFIC MORTALITY RATES :
 ******************************
         AGE            OBS. RATE     EST. RATE    DISTORTION
                          K1+K2         K1+K2
  1      22.00         .4570E-02     .3504E-02      1.304
  2      23.00         .4501E-02     .3521E-02      1.278
  3      24.00         .4430E-02     .3655E-02      1.212
  4      25.00         .4389E-02     .3868E-02      1.135
  5      26.00         .4327E-02     .4061E-02      1.065
  6      27.00         .4223E-02     .4206E-02      1.004
  7      28.00         .4107E-02     .4226E-02       .9720
  8      29.00         .4042E-02     .4287E-02       .9429
```

PLOT OF OBSERVED (M) AND EXPECTED (A) RATES
****

```
 .462E-02 *
 .456E-02 *       M
 .450E-02 *             M
 .444E-02 *                   M
 .438E-02 *                         M
 .432E-02 *                                M
 .426E-02 *                                                  A
 .420E-02 *                                      2     A
 .414E-02 *
 .408E-02 *                                A           M
 .402E-02 *                                                  M
 .396E-02 *
 .390E-02 *
 .384E-02 *                         A
 .378E-02 *
 .372E-02 *
 .366E-02 *                   A
 .360E-02 *
 .354E-02 *             A
 .348E-02 *       A
 .342E-02 *
..........*.........*.........*.........*.........*.........*..
         20.8      22.4      24.0      25.6      27.2      28.8
```

STANDARDISED MORTALITY RATIOS
****

	K=1	K=2	(1)/(2)
HOEM'S SMR	1.357	0.7367	1.843
OBSERVED SMR	1.179	0.7094	1.662
DISTORTION	0.8688	0.9629	0.9022

IPD-WORKING PAPERS 1983

PRICE PER COPY	
BF 230 US$ 6.00	1983-1. R. Lesthaeghe : "A Century of Demographic and Cultural Change in Western Europe : An Exploration of Underlying Dimensions".
BF 300 US$ 8.00	1983-2. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Education on Proximate Determinants and on Life Time Fertility in Kenya".
BF 230 US$ 6.00	1983-3. F. Eelens : "Impact of Breast-feeding on Infant and Child Mortality with varying Incidence of Malaria - Evidence from the Kenya Fertility Survey 1977-78".
BF 230 US$ 6.00	1983-4. S. Becker, A. Chowdhury, S. Huffman : "Determinants of Natural Fertility in Matlab, Bangladesh".
BF 350 US$ 9.00	1983-5. C. Vanderhoeft : "A Unified Approach to Models for Analysis of Zero-One Data with Applications to Intermediate Fertility Variables".
BF 300 US $ 8.00	1983-6. S. Wijewickrema, R. Bulté : "Migration Impact on Population Growth in Belgium : A Multiregional Investigation with Detailed Projection Results for the Period 1976-2001".
BF 230 US$ 6.00	1983-7. S. Wijewickrema et Alii : "Marital Status Trends in Belgium, 1961-77 : Application of Multi-State Analysis".
BF 230 US$ 6.00	1983-8. H.J. Page : "Fertility and Family : New Currents and emerging Emphases in Research and Policy" (Overview of the UN Expert Group Meeting on Fertility and Family (New Delhi, 1983) preparatory to the International Conference on Population).
BF 230 US$ 6.00	1983-9. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Female Education on the Kenyan Marital Fertility Transition" (abbreviated non-technical version of Working Paper 1983-2).

IPD-WORKING PAPERS 1984

PRICE PER COPY	
BF 230 US$ 6.00	1984-1. F. Rajulton : "An Age Dependent Semi-Markov Model of Marital Status in Belgium : An Application of Littman's Algorithm to Period Data, 1970".
BF 400 US$10.00	1984-2. R. Lesthaeghe : "Fertility and its Proximate Determinants in sub-Saharan Africa : The Record of the 1960's & 70's".
BF 300 US$ 8.00	1984-3. C. Vanderhoeft : "Life Tables and Relational Models : A Unified Approach, and Applications in Demography".
BF 230 US$ 6.00	1984-4. R. Lesthaeghe : "Demographic Renewal and Problems of Social Organization".
BF 230 US$ 6.00	1984-5. S. Becker : "A Response Bias in the reporting of Month of Birth in Pregnancy History Surveys".
BF 230 US$ 6.00	1984-6. R. Lesthaeghe : "On the Adaptation of sub-Saharan Systems of Reproduction".

IPD-WORKING PAPERS 1985

PRICE PER COPY	
BF 330 US$ 8.50	1985-1. R. Lesthaeghe and F. Eelens : "Social Organization and Reproductive Regimes : Lessons from sub-Saharan Africa and historical Western Europe".
BF 300 US$ 8.00	1985-2. P. Willems and S. Wijewickrema : "The Evolution of Nuptiality in Belgium from 1954 to 1981".
BF 400 US$10.00	1985-3. F. Eelens and L. Donné : "The Proximate Determinants of Fertility in Sub-Saharan Africa: A Factbook based on the Results of the World Fertility Survey".
BF 130 US$ 3.50	1985-4. Ronald C. Schoenmaeckers : "Current Fertility Behaviour in Africa. Results from a Birth Interval Analysis of WFS Data".
BF 400 US$10.00	1985-5. Camille Vanderhoeft : Stratified Proportional Hazards Models. A GLIM oriented approach, with special reference to the problem of competing risks.
BF 250 US$ 6.50	1985-6. Stanley Wijewickrema : "Childlessness in Belgium and Flanders".
BF 250 US$ 6.50	1985-7. R. Lesthaeghe : "Value Orientations, Economic Growth and Demographic Trends - Toward a Confrontation?
BF 250 US$ 6.50	1985-8. Fernando Rajulton and Stan Wijewickrema : "An Enquiry into the Two Basic Assumptions of Multistate Increment-Decrement Life Tables".

IPD-WORKING PAPERS 1986

PRICE PER COPY	
BF 250 US$ 6.50	1986-1. F. Rajulton : "Marriageability and divorceability : a simulation of the unobservables through the conditional gaussian diffusion process.
BF 180 US$ 4.50	1986-2. H.J. Page : "Child-bearing versus Child-rearing : Co-residence of Mothers and Children in Sub-Saharan Africa".
BF 450 US$11.50	1986-3. R. Lesthaeghe, G. Kaufmann, D. Meekers : "The Nuptiality Regimes in Sub-Saharan Africa".
BF 230 US$ 6.00	1986-4. R. Lesthaeghe, G. Kaufmann, D. Meekers : "The Nuptiality Regimes in Sub-Saharan Africa : Data Files NUPFILE 1 and NUPFILE 2".
BF 200 US$ 5.00	1986-5. Hwa Young LEE and Fernando RAJULTON : "A Semi-Markovian Multistate Analysis of Sexual Union Formation in Flanders, Belgium".
BF 450 US$11.50	1986-6. Zeng Yi : "Fertility and Household Dynamics in China".
BF 300 US$ 8.00	1986-7. R. Lesthaeghe & D. Meekers : "Value Changes and the Dimensions of Familism in the EEC".

IPD-WORKING PAPERS 1987

PRICE PER COPY	
BF 250 US$ 6.50	1987-1. C. Vanderhoeft : "Hidden heterogeneity in composite link models : further developments".
BF 250 US$ 6.50	1987-2. M. Rafiq Shah : "The use of a piecewise constant proportional hazards model in issues related to standardisation : a review and some results".