II. CONSTRUCTION OF THE MODEL LIFE TABLES

After experimentation with several approaches, a variation of classical principal components analysis was chosen as the analytical model. In this approach the age patterns of mortality which comprised the refined input data set were stratified into clusters by graphical and statistical procedures, each cluster having a distinct average age pattern of mortality. A principal-components model was then fitted to the deviations of each age pattern of mortality from its own cluster average. The age pattern of mortality for each input life table was operationalized as the vector of logit $[{}_nq_x]$ values, the cluster average as the simple averages of the logit $[{}_nq_x]$ values within the cluster and the deviations of each pattern from its cluster average as the arithmetic differences for each age-group. In all cases the age-groups involved were 0-1, 1-4, 5-9, 10-14, ... 80-84.

The details of the model life table construction are as follows. First, profiles of the age patterns of mortality for each life table were constructed by two statistical procedures and one graphical procedure. The statistical procedures were linearly optimal profile construction (based on second and third eigenvectors) and dynamic clustering analysis (maximum linkage, lower ultrametric).[6] The graphical procedure was very simple. For each input life table, the ratios $R(x) = {}_nq_x/{}_nq_x^w$ were calculated, where ${}_nq_x$ is the mortality rate at age x for the given life table and ${}_nq_x^w$ is the mortality rate at age x in the Coale and Demeny West region model life table with the same life expectancy at age 10. The $R(x)$ values were then plotted against age for each life table and the plots were ocularly arranged according to similarity of patterns. All three methods produced essentially the same clusters. There were four clear pattern groups and a few life tables which did not fit together well or easily into any other groups. The four pattern groups or clusters were as follows: The first cluster contains the life tables from the Latin American countries of Colombia, Costa Rica, El Salvador, Guatemala, Honduras, Mexico and Peru, as well as the non-American countries of the Philippines, Sri Lanka and Thailand. The second cluster was the very distinctive pattern of the Chilean life tables. The third cluster was made up of tables from India, Iran, the Matlab area of Bangladesh and Tunisia. The fourth cluster consisted of the tables from Guyana, Hong Kong, the Republic of Korea, Singapore and Trinidad and Tobago among the male populations, and Guyana, Singapore and Trinidad and Tobago among the female populations. The four patterns have been labelled the Latin American pattern, the Chilean pattern, the South Asian pattern, and the Far Eastern pattern, respectively, according to the geographical region which is predominant within each pattern group. Life tables from Israel and Kuwait as well as those for the female populations of Hong Kong and the Republic of Korea did not cohere into any cluster and were therefore omitted from the principal-components analysis and included only in construction of the general pattern of mortality described below.

Within each of these clusters values of ${}_nD_x^{ij}$ were calculated; these are defined, for each age-group, as $(x, x + n)$, the difference between the logit $[{}_nq_x]$ values for life table j of cluster i and the average of the logit values for all the life tables within cluster i, where

$$\text{logit } [{}_nq_x] = 1/2\ ln\left(\frac{{}_nq_x}{1 - {}_nq_x}\right).$$

As expected, a ${}_nD_x^{ij}$ value for one age-group is highly correlated with values at other age-groups. The correlation matrix is presented in tables 3 and 4 for males and

[6] These procedures are explained in J. A. Hartigan, *Clustering Algorithms* (New York, John Wiley and Sons, 1975); and P. H. Sneath and R. R. Sokal, *Numerical Taxonomy—The Principles and Practices of Numerical Classification* (San Francisco, Calif., W. H. Freeman, 1973).

TABLE 3. CORRELATIONS AMONG ${}_nD_x^{ij}$ VALUES CALCULATED FROM MALE LIFE TABLES

	Age x																	
Age x	*0*	*1*	*5*	*10*	*15*	*20*	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*	*70*	*75*	*80*
0		0.93	0.79	0.82	0.79	0.83	0.84	0.83	0.85	0.83	0.82	0.79	0.77	0.75	0.74	0.68	0.59	0.35
1			0.87	0.85	0.78	0.82	0.84	0.84	0.86	0.85	0.84	0.82	0.81	0.81	0.79	0.72	0.63	0.42
5				0.96	0.89	0.88	0.88	0.89	0.91	0.92	0.93	0.92	0.93	0.92	0.91	0.87	0.85	0.69
10					0.97	0.95	0.93	0.92	0.93	0.94	0.95	0.93	0.94	0.92	0.93	0.90	0.87	0.68
15						0.97	0.95	0.92	0.93	0.93	0.94	0.93	0.93	0.90	0.91	0.87	0.85	0.65
20							0.98	0.96	0.96	0.96	0.95	0.93	0.93	0.89	0.89	0.82	0.79	0.58
25								0.99	0.98	0.98	0.96	0.94	0.94	0.89	0.89	0.80	0.79	0.55
30									0.98	0.98	0.96	0.94	0.93	0.89	0.87	0.77	0.76	0.52
35										0.99	0.99	0.96	0.96	0.91	0.90	0.82	0.80	0.56
40											0.99	0.98	0.97	0.93	0.91	0.84	0.81	0.58
45												0.99	0.98	0.95	0.94	0.87	0.84	0.63
50													0.99	0.98	0.95	0.89	0.86	0.66
55														0.98	0.97	0.92	0.90	0.69
60															0.96	0.93	0.89	0.72
65																0.95	0.93	0.73
70																	0.97	0.83
75																		0.89
80																		

TABLE 4. CORRELATIONS AMONG ${}_nD_x^{ij}$ VALUES CALCULATED FROM FEMALE LIFE TABLES

	Age x																	
Age x	*0*	*1*	*5*	*10*	*15*	*20*	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*	*70*	*75*	*80*
0		0.90	0.73	0.80	0.80	0.77	0.78	0.81	0.83	0.84	0.83	0.80	0.76	0.75	0.68	0.59	0.54	0.25
1			0.89	0.89	0.91	0.90	0.91	0.92	0.92	0.91	0.89	0.85	0.84	0.82	0.79	0.72	0.66	0.40
5				0.94	0.94	0.94	0.94	0.93	0.93	0.92	0.92	0.91	0.92	0.91	0.91	0.88	0.82	0.62
10					0.96	0.94	0.94	0.94	0.95	0.94	0.93	0.91	0.91	0.90	0.89	0.87	0.84	0.63
15						0.99	0.98	0.98	0.98	0.96	0.95	0.92	0.91	0.88	0.87	0.86	0.83	0.65
20							0.99	0.99	0.98	0.95	0.94	0.91	0.90	0.87	0.88	0.87	0.84	0.67
25								0.99	0.99	0.96	0.95	0.92	0.92	0.89	0.89	0.87	0.85	0.68
30									1.00	0.98	0.97	0.94	0.93	0.90	0.89	0.87	0.84	0.65
35										0.99	0.97	0.95	0.93	0.91	0.90	0.87	0.84	0.64
40											0.99	0.98	0.96	0.95	0.91	0.86	0.83	0.61
45												0.99	0.98	0.97	0.93	0.88	0.85	0.62
50													0.99	0.98	0.94	0.88	0.83	0.61
55														0.99	0.97	0.91	0.86	0.63
60															0.97	0.91	0.86	0.61
65																0.97	0.93	0.71
70																	0.98	0.83
75																		0.89
80																		

females. Correlation coefficients are generally above 0.80 with low correlations occurring mainly for the oldest age-groups. Because each cluster consists of a set of consistent life tables with similar age patterns of mortality, the ${}_nD_x^{ij}$ vector for each input life table can be considered an indication of the age pattern of mortality change, i.e., it indicates how mortality changes by age. On the assumption that the age pattern of mortality change is invariant to the cluster pattern,[7] we can express the age structure of mortality in any country (defined by its logit $[{}_nq_x]$ values) as ${}_nY_x^{ij} \doteq {}_n\overline{Y}_x^{i} + a_{1j}U_{1x}$ where ${}_nY_x^{ij}$ equals the logit of the ${}_nq_x$ function for life table j of cluster i and ${}_n\overline{Y}_x^{i}$ equals the average of the ${}_nY_x^{ij}$ within each cluster. The vector U_{1x} then designates the average age pattern of mortality change (some kind of average of the ${}_nD_x^{ij}$ values) and a_{1j} designates the amount of change. This is essentially a 1-component principal-components model, with the vector U_{1x}, called the first principal-component vector, signifying the age pattern of mortality change and its coefficient (a_{1j}), called the loading factor, indicating the extent of the change pertaining to life table j.

Of course, this 1-component model will not explain all the variation in the age structures of mortality that appear in the life tables of the refined data set. New sets of deviations, calculated as the difference between the empirical logit $[{}_nq_x]$ values and those predicted from the 1-component model, can be calculated. If we let U_{2x} designate some kind of average age pattern of these second-order deviations and a_{2j} designate the magnitude of this pattern of deviations for any life table j, then a 2-component model can be constructed as:

$${}_nY_x^{ij} \doteq {}_n\overline{Y}_x^{i} + \sum_{m=1}^{2} a_{mj}U_{mx}.$$

In the same way 3-, 4- or up to 18-component models can be estimated. The functional form of the model can therefore be expressed as

$${}_nD_x^{ij} = {}_nY_x^{ij} - {}_n\overline{Y}_x^{i} \doteq \sum_{m=1}^{k} a_{mj}U_{mx} \qquad (1)$$

where:

${}_nY_x^{ij}$ equals the logit of the ${}_nq_x$ function (probability of dying between ages x and $x + n$) for life table j of cluster i; ${}_n\overline{Y}_x^{i}$ equals the average of the ${}_nY_x^{ij}$ within each cluster; a_{mj} equals the factor loading to the m^{th} principal-component vector for country j in the principal-components analysis; U_{mx} equals the element of the m^{th} principal-component vector corresponding to age-group $(x, x + n)$; and k is the number of principal components.

For application purposes it is often more convenient to express the model as

$${}_nY_x^{ij} \doteq {}_n\overline{Y}_x^{i} + \sum_{m=1}^{k} a_{mj}U_{mx}. \qquad (2)$$

When $k=1$, the model is referred to as a 1-component model; when $k=2$, as a 2-component model, and so forth. The principal-components model is similar to more usual linear regression procedures in that the values of parameters are found which minimize sums of squared deviations. In this case, we find the vectors $U_{1x}, U_{2x}, U_{3x}, \ldots, U_{kx}$ which sequentially minimize the sum of squared deviations between actual and predicted ${}_nD_x^{ij}$ values. Distances between actual and predicted values are measured as perpendicular (orthogonal) distances, rather than vertical distances. It can be shown that the U_{mx} vectors are simply the eigenvectors of the matrix of covariances of the ${}_nD_x^{ij}$ values.[8] Although as many components as there are variables (age-groups) are necessary to explain all the variation in the ${}_nD_x^{ij}$ values (in our case 18 components are necessary since there are 18 age-groups), often the first few components account for a sufficient amount of variation to be usable for many purposes. In

[7] The assumption of invariance of age pattern of mortality change to cluster pattern appears very strong at first glance. However, separate application of the principal-components analysis on the Latin American pattern, the Chilean pattern and the Far Eastern pattern showed very similar first-component vectors within each cluster. (Because there was little variation in mortality levels among the life tables included in the South Asian pattern it was impossible to carry out a separate principal-components analysis for that cluster.) It was this empirical finding that permitted the superimposing of a single pattern of mortality change on the four different basic age patterns.

[8] For a more rigorous description of principal-components analysis see, for example, D. F. Morrison, *Multivariate Statistical Methods* (New York, McGraw-Hill, 1976).

TABLE 5. AVERAGE PATTERN OF MORTALITY FOR EACH CLUSTER DEFINED BY LOGIT $[{}_nq_x]$ VALUES

	Males				
	Cluster				
Age x	*Latin American*	*Chilean*	*South Asian*	*Far Eastern*	*General*
0	−1.12977	−1.04722	−0.97864	−1.53473	−1.27638
1	−1.49127	−1.81992	−1.24228	−2.15035	−1.78957
5	−2.13005	−2.42430	−2.01695	−2.61442	−2.35607
10	−2.40748	−2.52487	−2.44280	−2.66392	−2.55527
15	−2.21892	−2.24491	−2.35424	−2.42326	−2.34263
20	−2.01157	−2.02821	−2.27012	−2.23095	−2.16193
25	−1.93591	−1.90923	−2.16833	−2.15279	−2.09109
30	−1.86961	−1.78646	−2.05942	−2.05765	−2.00215
35	−1.76133	−1.66679	−1.90053	−1.89129	−1.86781
40	−1.64220	−1.52497	−1.71213	−1.68244	−1.70806
45	−1.49651	−1.37807	−1.51120	−1.47626	−1.52834
50	−1.34160	−1.21929	−1.28493	−1.23020	−1.33100
55	−1.15720	−1.03819	−1.08192	−1.02801	−1.12934
60	−0.96945	−0.84156	−0.84671	−0.77148	−0.91064
65	−0.74708	−0.63201	−0.62964	−0.54696	−0.68454
70	−0.52259	−0.42070	−0.40229	−0.32996	−0.45685
75	−0.29449	−0.21110	−0.19622	−0.11911	−0.23002
80	−0.04031	0.01163	−0.00129	0.10572	0.00844

	Females				
	Cluster				
Age x	*Latin American*	*Chilean*	*South Asian*	*Far Eastern*	*General*
0	−1.22452	−1.12557	−0.97055	−1.42596	−1.35963
1	−1.45667	−1.82378	−1.15424	−1.95200	−1.77385
5	−2.13881	−2.52319	−1.93962	−2.55653	−2.39574
10	−2.46676	−2.63933	−2.36857	−2.68018	−2.64549
15	−2.31810	−2.38847	−2.19082	−2.33095	−2.44766
20	−2.14505	−2.20417	−2.09358	−2.15952	−2.28991
25	−2.03883	−2.09701	−2.04788	−2.03377	−2.18850
30	−1.93924	−1.99128	−1.95922	−1.94554	−2.08535
35	−1.83147	−1.87930	−1.87311	−1.82299	−1.97231
40	−1.74288	−1.75744	−1.76095	−1.69084	−1.84731
45	−1.62385	−1.61558	−1.61425	−1.52189	−1.69291
50	−1.47924	−1.45886	−1.39012	−1.33505	−1.50842
55	−1.28721	−1.26115	−1.15515	−1.13791	−1.30344
60	−1.07443	−1.05224	−0.90816	−0.93765	−1.08323
65	−0.83152	−0.80346	−0.68011	−0.72718	−0.84402
70	−0.59239	−0.58202	−0.43231	−0.50916	−0.59485
75	−0.35970	−0.35093	−0.17489	−0.28389	−0.34158
80	−0.08623	−0.10587	0.05948	−0.01285	−0.06493

the case of the model life table project, one component alone explained about 90 per cent of the variation, whereas three components explained 97 per cent.

In table 5 is presented, by sex, the average pattern of mortality for each cluster as operationalized by the average of the logit [${}_nq_x$] values of the life tables it contains. An over-all pattern, referred to here as the "general pattern", is also shown. This general pattern was estimated by averaging the logit [${}_nq_x$] values of all life tables in the refined data set without regard to cluster. Chapter III below describes in detail the characteristics of the various patterns.

Table 6 presents the first three principal-component vectors by sex. As expected, the first principal component models the age pattern of mortality change. According to this component, as mortality declines, change is greatest during the childhood years and lessens as age increases. Declines during infancy are somewhat smaller than those during childhood, similar to those that take place during the later middle years of life. The phrase, "pattern of mortality change", as used here, of course refers to change in the logit [${}_nq_x$] function of the life table. Since, except when ${}_nq_x$ values are quite high, the logit [${}_nq_x$] is very close to one half of $\ell n\ {}_nq_x$ values, it is possible to think of elements of the first component as representing proportional change in ${}_nq_x$ values.

The second component appears to account mainly for characteristic differences among life tables in the relation between mortality under age 5 and mortality above age 5, differences that were not fully accounted for by either the initial clustering of the mortality patterns into four groups or by the age pattern of mortality change described by the first component. The third component appears to affect mortality during the childbearing years for females and during a diverse group of ages for males.

The set of model life tables presented in annex I is a 1-component model, based on the five average patterns (the four distinct pattern groups and the over-all general

TABLE 6. FIRST THREE PRINCIPAL COMPONENTS

	Males			*Females*		
Age x	*1st component* U_{1x}	*2nd component* U_{2x}	*3rd component* U_{3x}	*1st component* U_{1x}	*2nd component* U_{2x}	*3rd component* U_{3x}
0	0.23686	−0.46007	0.09331	0.18289	−0.51009	0.23944
1	0.36077	−0.68813	−0.29269	0.31406	−0.52241	−0.11117
5	0.33445	0.06414	−0.47139	0.31716	0.08947	0.07566
10	0.30540	0.12479	−0.17403	0.30941	0.03525	0.06268
15	0.28931	0.24384	0.10715	0.32317	0.03132	−0.26708
20	0.28678	0.10713	0.28842	0.32626	0.07843	−0.39053
25	0.27950	0.06507	0.33620	0.30801	0.06762	−0.28237
30	0.28023	0.03339	0.33692	0.29047	0.00482	−0.14277
35	0.26073	0.02833	0.21354	0.25933	−0.01409	−0.05923
40	0.23626	0.06473	0.15269	0.22187	−0.02178	0.18909
45	0.20794	0.08705	0.06569	0.19241	0.01870	0.24773
50	0.17804	0.10620	0.00045	0.17244	0.04427	0.33679
55	0.15136	0.11305	−0.03731	0.15729	0.08201	0.34121
60	0.13217	0.09467	−0.10636	0.14282	0.08061	0.38290
65	0.12243	0.10809	−0.11214	0.12711	0.15756	0.26731
70	0.11457	0.14738	−0.22258	0.11815	0.24236	0.14442
75	0.10445	0.21037	−0.19631	0.11591	0.30138	0.09697
80	0.08878	0.30918	−0.38123	0.09772	0.50530	−0.13377

TABLE 7. PROPORTION OF VARIATION IN MORTALITY EXPLAINED BY COMPONENT AND AGE

	Males			*Females*		
	Proportion of variation explained by:			*Proportion of variation explained by:*		
Age x	*One component*	*Two components*	*Three components*	*One component*	*Two components*	*Three components*
0	0.772	0.931	0.935	0.687	0.911	0.932
1	0.815	0.977	0.993	0.870	0.971	0.973
5	0.909	0.911	0.965	0.921	0.924	0.925
10	0.951	0.960	0.969	0.935	0.936	0.937
15	0.914	0.950	0.954	0.974	0.974	0.986
20	0.940	0.947	0.975	0.967	0.969	0.994
25	0.946	0.949	0.990	0.974	0.976	0.991
30	0.933	0.934	0.974	0.982	0.982	0.986
35	0.959	0.960	0.979	0.985	0.986	0.986
40	0.962	0.966	0.978	0.962	0.962	0.975
45	0.964	0.974	0.976	0.954	0.954	0.982
50	0.933	0.951	0.951	0.913	0.915	0.976
55	0.934	0.963	0.964	0.898	0.908	0.982
60	0.884	0.909	0.926	0.866	0.878	0.987
65	0.870	0.907	0.929	0.843	0.897	0.962
70	0.763	0.831	0.918	0.789	0.928	0.949
75	0.700	0.854	0.928	0.725	0.931	0.940
80	0.386	0.641	0.854	0.412	0.874	0.887
All ages combined	0.892	0.940	0.967	0.913	0.952	0.968

pattern) and the age pattern of mortality change defined by the first principal component. As shown in table 7, such a 1-component model explains most of the variation among all the input life tables. Specifically, 89 per cent for males and 91 per cent for females of the variation in the logit $[{}_nq_x]$ values in the input data set are accounted for by the first component after the regional clustering. However, as is clearly observed from table 7, the amount of variation in mortality explained is not identical for all age-groups. Among males, 90 per cent or more of variation is explained only for ages between 5 and 59; for females, this range extends from 5 to 54. The second and third components explain over-all an additional 5 per cent and 3 per cent of variation, respectively, for males; 4 per cent and 2 per cent, respectively, for females. For both sexes, with extension of the model to two components, over 90 per cent of variation is explained for all but a few of the oldest age-groups.

The models presented in annex I are 1-component models. However, advantage can be taken of the availability of the second and third components and the additional variation they explain to form variant model age patterns. These possibilities are explored in chapter IV below.