

## Chapter IX

# ESTIMATION OF ADULT MORTALITY USING SUCCESSIVE CENSUS AGE DISTRIBUTIONS

### A. BACKGROUND OF METHODS

#### 1. *Use of a sequence of population age distributions*

The value for demographic estimation purposes of having basically similar information about a population for two points in time has already been stressed, for example, in the discussion of the use of overlapping fertility estimates based on reverse-survival techniques (chapter VIII) and in the sections dealing with hypothetical-cohort methods (see chapters II, III and IV). The age and sex distributions from successive enumerations of a population also provide a basis for estimating intercensal mortality. In a closed population with two accurate censuses  $t$  years apart, the population aged  $x + t$  at the second census represents the survivors of the population aged  $x$  at the first census, so that the intercensal survivorship probability from age  $x$  to age  $x + t$  can be calculated. Traditional mortality measures can then be obtained from the sequence of survivorship probabilities for successive initial ages  $x$ .

This method of mortality estimation from intercensal survival is appealingly simple and straightforward, requiring only the most basic of census information, making no assumptions about the age pattern of mortality, and providing estimates of mortality for a clearly defined time period. The trouble is that these advantages are nullified by the requirements that the censuses be accurate and that the population be closed. In practice, the application of this method very often gives disappointing results. Migration can affect a population as much as mortality; and, in particular, at young adult ages, its influence on population size may be more important than that of mortality. Age-misreporting can also distort the results severely; a marked preference for certain digit endings when declaring age, which will introduce considerable variability into the estimated survivorship ratios, can be reduced to some extent by the use of grouped data, but systematic overreporting or underreporting of age can cause insuperable problems. Changes in enumeration completeness from one census to the other can, if no adjustment is feasible, completely swamp the effects of mortality, giving rise to very misleading results; indeed, it may be stated without exaggeration that, in many cases, intercensal survival estimates are better indicators of the comparability of two census enumerations than of the level of intercensal mortality.

Despite these problems, however, it is worth applying the method where possible, because if the errors in the basic data are not overwhelming, one can obtain useful

estimates of mortality by using suitable age groups and certain smoothing techniques; and even if the data errors are severe, the calculated survivorship probabilities may be useful indicators of the nature of the errors involved.

#### 2. *Organization of this chapter*

The methods described in this chapter are all based essentially on the use of two successive age (and preferably sex) distributions of a population. The age distributions should be obtained from complete enumerations, because sampling errors would greatly distort the results were data from sample surveys used; and the population should ideally be closed to migration, failing which, one population or the other should be adjusted for the effects of net migration. The calculations are greatly simplified if the length of the intercensal period is an exact multiple of five years, though other periods can be accommodated; the period should not, in general, exceed 15 years. The main features of the methods available are described below (for data requirements and parameters estimated, see table 171):

*Section B. Estimation of mortality from intercensal survivorship probabilities.* The traditional procedure for estimating adult mortality from two successive census age distributions by calculating cohort survivorship probabilities for the intercensal period is described. Different procedures for smoothing<sup>1</sup> the calculated probabilities, using the Coale-Demeny model life tables or the logit life-table system in order to estimate a single mortality parameter, are also presented. Variants based on cumulated and uncumulated age distributions are included;

*Section C. Intercensal survival with additional information on the age pattern of mortality.* In a closed population, the proportionate reduction in cohort size from one census to another can be compared with the proportionate reduction expected on the basis of cohort deaths as recorded by a vital registration system or retrospective survey question. Since the age patterns of the popula-

<sup>1</sup> The term "to smooth" is used in this *Manual* in its most general sense to mean the elimination or minimization of irregularities often present in reported data or in preliminary estimates obtained from them. In this sense, the set of possible "smoothing techniques" encompasses a wide variety of procedures, ranging from the fitting of models to simple averaging. The traditional smoothing techniques applied to age distributions and to observed age-specific mortality rates are part of this set, but they do not exhaust it. The somewhat rougher procedures described in this *Manual* are necessary because the basic data available are both deficient and incomplete.

TABLE 171. SCHEMATIC GUIDE TO CONTENTS OF CHAPTER IX

Section	Subsection and method	Type of input data	Parameters estimated
B. Estimation of mortality from intercensal survivorship probabilities	B.2 Survival ratios for five-year age cohorts smoothed by using the Coale-Demeny life tables	Population classified by age and sex from two censuses 15 years apart or less. If the intercensal period is not an exact multiple of five, at least one age distribution must be by single years. An estimate of net intercensal migration	A life table for the intercensal period from age 10 or so onward
	B.3 Survival ratios for five-year age cohorts smoothed by using the logit system	Population classified by age and sex from two censuses 15 years apart or less. If the intercensal period is not an exact multiple of five, at least one age distribution must be by single years. An estimate of child survivorship An estimate of net intercensal migration	A life table for the intercensal period from age 10 or so onward
	B.4 Survival ratios smoothed by cumulation and by using the Coale-Demeny life tables	Population classified by age and sex from two censuses 15 years apart or less. If the intercensal period is not a multiple of five, at least one age distribution must be by single years. An estimate of net intercensal migration	A life table for the intercensal period from age 10 or so onward
C. Intercensal survival with additional information on the age pattern of mortality		Population classified by age and sex from two censuses 15 years apart or less. If the intercensal period is not an exact multiple of five, at least one age distribution must be by single years. Registered deaths during the intercensal period classified by age and sex Estimates of net intercensal migration (to adjust raw data, if necessary)	Completeness of coverage of recorded deaths with respect to the completeness of the first census Completeness of coverage of the first census in relation to that of the second
D. Estimation of a post-childhood life table from an age distribution and intercensal growth rates		Population classified by age and sex from two censuses 15 years apart or less. The same age classification must be used for both populations Estimates of net intercensal migration (to adjust raw data, if necessary)	Estimates of ${}_5L_x$ from age 10 or 15 onward

tion and of deaths are very different, it is possible to disentangle the effects of changes in enumeration completeness from undercoverage of deaths. Specifically, if enumeration completeness and coverage of deaths do not vary with age, at least after childhood, the enumeration completeness of one census in relation to the other, and the completeness of death recording in relation to either, can be estimated. The procedure is severely affected, however, by systematic age errors;

*Section D. Estimation of a post-childhood life table from an age distribution and intercensal growth rates.* Two census enumerations provide the data necessary to calculate intercensal growth rates for five-year age groups. These growth rates can then be used to convert the average age distribution from the two censuses into stationary population form, that is, a life-table  ${}_5L_x$  function. This method is computationally attractive if the length of the intercensal period is not a multiple of five years.

**B. ESTIMATION OF MORTALITY FROM INTERCENSAL SURVIVORSHIP PROBABILITIES**

*1. Basis of methods and their rationale*

The methods described in this section are all based on the same, very simple information, namely, the change in size of successive age cohorts of a population from one census to the next. The methods differ only in the ways in which this basic information is smoothed to reduce the effects of errors and converted into a mortality parameter. Therefore subsections B.3 and B.4 only cover those steps which are different from the steps described in subsection B.2.

*2. Intercensal survivorship ratios for five-year age cohorts smoothed using the Coale-Demeny life tables*

*(a) Data required*

The data required for this method are two census

enumerations with populations classified by age and sex. (Classification by sex is not necessary, but since it is generally available, it is useful to consider it whenever possible.) Although  $\omega$ , the highest age reached, is the theoretical limit for the value of the intercensal interval  $t$  (measured in years), in practice, with intervals longer than 15 years the information available is scanty and the calculations are more likely to be affected by changes other than those caused by mortality. If  $t$  is divisible by five, both age distributions can be classified by five-year age groups and cohorts from the first census can be identified at the second census; if  $t$  is not a multiple of five, it is convenient to have one of the age distributions by single year of age, so that comparable cohorts may be constructed.

(b) *Computational procedure*

The steps of the computational procedure are described below.

*Step 1: adjustment for net intercensal migration and territorial coverage.* Substantial net migration during the intercensal period will generally render the method of intercensal survival unusable. Nevertheless, if it is possible to adjust one age distribution or the other on an age-specific basis for the effects of migration, the method may be applied after such adjustment has been made. However, it is most unusual for adequate information about migration to be available, and no general procedures for carrying out an adjustment can be expounded here. Problems introduced by changes in territorial coverage may not be quite so serious. By the judicious aggregation of subnational information from one census or the other, it is usually possible to arrive at age distributions for comparable populations. No general procedures for so doing need to be stated, however, beyond pointing out the necessity of making suitable adjustments if changes in territorial coverage have occurred.

*Step 2: grouping of data from the two censuses by cohort.* Simple intercensal survival techniques generally disregard the effects of age distribution within cohort groupings, assuming in effect that the population is distributed within each age group in the same way as if it were a stationary or life-table population. As a result of this simplifying assumption, the width of the cohorts should not be too large (probably not more than five years). Groupings that are five years in size are also convenient because most model life-table systems are tabulated for five-year age groups, though other intervals can be used if necessary. If the intercensal interval  $t$  is divisible by five, conventional five-year age groups from  $x$  to  $x+4$  at the first census will survive to become conventional five-year age groups from  $x+t$  to  $x+t+4$  at the second census; and no regrouping is required. If  $t$  is not divisible by five, a single-year age distribution from either the first or the second census can be used in order to create groups corresponding to conventional five-year cohorts at the other census. In cases where there is substantial age-heaping and a danger that it may introduce systematic age exaggeration, unconventional five-year age

groups (such as 3-7, 8-12 and 13-17) centred on preferred-digit endings may be used.

*Step 3: adjustment for intercensal interval that is not an exact number of years.* When the intercensal interval is not an exact number of years, a small adjustment should be made to one population or the other, by moving it forward or backward in order to approximate the population corresponding to the nearest date defining an interval with an exact number of years and thus to remove the slight effect that normal population growth would have on the intercensal survivorship estimates. The intercensal growth rate  $r$  can be calculated as

$$r = [\ln(N_2) - \ln(N_1)]/t$$

where  $N_2$  is the total population recorded by the second census;  $N_1$  is the total population recorded by the first census; and  $t$  is the intercensal period measured in years. This growth rate can then be used to move either the first or second age distribution over the required length of time. If the decimal portion of  $t$  is less than 0.5, the interval should be shortened to  $t$  exact years, whereas if it is greater than 0.5, the interval should be lengthened to  $t+1$  exact years. If the decimal portion of  $t$  is denoted by  $z$ , the interval can be shortened to  $t$  years either by multiplying each age group at the first census by a factor  $\exp[rz]$  or by multiplying each age group at the second census by a factor  $\exp[-rz]$ . The interval can be lengthened either by multiplying each age group at the first census by a factor  $\exp[r(z-1.0)]$  or by multiplying each age group at the second census by a factor  $\exp[r(1.0-z)]$ .

*Step 4: calculation of cohort survivorship ratios.* Cohort survivorship probabilities or ratios during the intercensal period, denoted by  ${}_tS_{x,x+5}$ , can now be calculated by dividing the cohort size at the second census by its size at the first census. These survivorship ratios approximate life-table (or stationary-population) survivorship probabilities, provided the effects of deviations of the actual age distribution within cohorts from that corresponding to the stationary population are small (as is usually the case). Thus,

$${}_tS_{x,x+5} = {}_5N_{x+t}^2 / {}_5N_x^1 = {}_5L_{x+t} / {}_5L_x$$

where  $t$  is the adjusted length of the intercensal interval after applying step 3;  ${}_5N_x^1$  is the population aged from  $x$  to  $x+4$  enumerated by the first census; and  ${}_5N_{x+t}^2$  is the population aged from  $x+t$  to  $x+t+4$  enumerated by the second census.

*Step 5: fitting of a Coale-Demeny model life table.* The consistency of the cohort survivorship ratios calculated in step 4 may be conveniently examined by finding the mortality level, in the Coale-Demeny model life tables, to which each ratio corresponds. A best estimate of mortality level can then be obtained by discarding any detectable outliers and basing the estimate on the remaining levels (by taking their average, for example). If the adjusted intercensal period  $t$  is divisible by five, stationary-population ratios of the type  ${}_5L_{x+t} / {}_5L_x$  can

be calculated directly for relevant levels of the selected regional family of model life tables. If  $t$  is not divisible by five, however, additional steps become necessary, since the Coale and Demeny life tables do not provide stationary-population age distributions for non-standard age groups. The simplest procedure is to calculate stationary-population distributions for non-standard age groups simply by weighting adjacent standard five-year values by the proportions of the age groups covered. Thus,  ${}_5L_{19}$ , the stationary population aged from 19 to 23, covers one fifth of age group 15-19 and four fifths of age group 20-24; it can be approximated as

$${}_5L_{19} = \left(\frac{1}{5}\right) {}_5L_{15} + \left(\frac{4}{5}\right) {}_5L_{20}$$

In fact, if the  $l(x)$  function is linear with age, the approximation is exact.

If somewhat more precision is required,  ${}_5L_{x+n}$  values can be estimated from tabulated values of  $l(x)$ ,  $l(x+5)$  and  $l(x+10)$  using equation (B.1):

$${}_5L_{x+n} = a(n)l(x) + b(n)l(x+5) + c(n)l(x+10) \quad (\text{B.1})$$

where the coefficients  $a(n)$ ,  $b(n)$  and  $c(n)$ , for  $n$  ranging from 0 to 4, are calculated by fitting a second-order polynomial to the  $l(x)$  values. Values of these coefficients are shown in table 172.

**Step 6: completion of the life table.** Intercensal survival provides no information about the mortality experience

TABLE 172. COEFFICIENTS FOR ESTIMATION OF STATIONARY-POPULATION AGE DISTRIBUTION,  ${}_5L_x$ , FOR UNCONVENTIONAL AGE GROUPS

Index $n$ (1)	Coefficients		
	$a(n)$ (2)	$b(n)$ (3)	$c(n)$ (4)
0.....	2.083	3.333	-0.417
1.....	1.183	4.133	-0.317
2.....	0.483	4.533	-0.017
3.....	-0.017	4.533	0.483
4.....	-0.317	4.133	1.183
5.....	-0.417	3.333	2.083

Estimation equation:

$${}_5L_{x+n} = a(n)l(x) + b(n)l(x+5) + c(n)l(x+10)$$

where  $x$  is exactly divisible by five.

of those born between the censuses, since the first census does not provide their initial number (accurate birth registration could supply this want, but where births are completely registered, better estimates of mortality would probably be available from other sources). In order to obtain a complete life table, therefore, it is necessary to supply further information about child mortality. The most satisfactory source of such estimates is information about children ever born and surviving (see chapter III). If estimates of this type are available, the methods described in chapter VI for linking estimates of child and adult mortality can be used to obtain a complete life table.

A problem remains, however, if no independent esti-

mate of child mortality is available. Since the Coale-Demeny life-table system has been used in selecting a model life table (see step 5), the mortality pattern of the model used can be adopted by taking the life table associated with the average mortality level of the intercensal survival probabilities as representative.

(c) *First detailed example: Panama, 1960-1970*

The first detailed example illustrates a fairly simple case of a 10-year interval and reasonably good age-reporting. Population censuses were held in Panama on 11 December 1960 and 10 May 1970. This example examines only the intercensal survival of the female population, though in a complete study, this analysis should be carried out for both males and females. Table 173 shows the female population enumerated by the two censuses classified by five-year age group.

The computational procedure for this example is given below.

**Step 1: adjustment for net intercensal migration and territorial coverage.** As no information on intercensal migration by age is available, no adjustment can be made. No change in territorial coverage occurred between 1960 and 1970; therefore, no adjustment for coverage is needed.

**Step 2: grouping of data from the two censuses by cohort.** Because the interval between the two censuses is about nine and one-half years, one of the populations has to be moved slightly to bring cohorts into alignment. The exact interval, 9.41 years, is somewhat closer to nine years than 10, so the adjustment for dates would be minimized by moving the first population forward slightly, or by moving the second one back, to create an intercensal period of nine years. However, there is also an advantage to working with intervals divisible by five, and since the actual interval was only slightly less than nine and one-half years, the convenience factor outweighs that of a marginal gain in accuracy. Thus, the first census will be moved back to approximate the female population on 10 May 1960 (the results would be precisely the same if the second census were moved forward to 11 December 1970). Standard five-year age groups will therefore define cohorts, and no regrouping is required.

**Step 3: adjustment for length of the intercensal interval.** The total female population in 1960 was 529,767, and in 1970 it was 704,333; thus, the exponential rate of population growth during the intercensal period 1960-1970 is

$$r = \frac{\ln [N_2] - \ln [N_1]}{t_2 - t_1} = \frac{13.4650 - 13.1802}{9.41} = 0.0303.$$

The growth factor  $k$  needed to adjust the 1960 population for 0.59 of a year's growth is then obtained as

$$k = \exp\{(0.0303)(9.41 - 10.00)\} = \exp(-0.0179) = 0.9823.$$

Column (3) of table 173 shows the adjusted population.

TABLE 173. ENUMERATED AND ADJUSTED FEMALE POPULATION IN 1960, ENUMERATED POPULATION IN 1970 AND COHORT SURVIVORSHIP RATIOS, PANAMA

Age group (1)	1960 population		1970 population enumerated (4)	Cohort survivorship ratio (cohort age in 1970) (5)
	Enumerated (2)	Adjusted <sup>a</sup> (3)		
0-4	90 071	88 477	114 017	-
5-9	76 598	75 242	106 944	-
10-14	63 635	62 509	85 253	0.9636
15-19	54 431	53 468	73 381	0.9753
20-24	45 634	44 826	63 010	1.0080
25-29	37 818	37 149	50 924	0.9524
30-34	32 179	31 609	40 885	0.9121
35-39	28 724	28 216	36 115	0.9722
40-44	23 974	23 550	29 409	0.9304
45-49	20 618	20 253	25 360	0.8988
50-54	15 068	14 801	21 775	0.9246
55-59	11 999	11 787	17 632	0.8706
60-64	10 283	10 101	13 004	0.8786
65-69	6 737	-	10 061	0.8536
70-74	5 242	18 403	6 690	0.6623
75+	6 756	-	9 873	0.5365
TOTAL	529 767		704 333	

<sup>a</sup>Enumerated population adjusted by a factor of 0.9823 to move it from 11 December 1960 to 10 May 1960.

**Step 4: calculation of cohort survivorship ratios.** Cohort survivorship ratios or probabilities are calculated by dividing the number in each cohort at the second census by the corresponding number in the same cohort at the first census, using, of course, the date-adjusted numbers in columns (3) and (4) of table 173. Thus, for example, the survivors of the cohort aged 20-24 at the first census are aged 30-34 at the second census, and the 10-year survivorship probability for the cohort,  ${}_{10}S_{20,24}$ , is calculated as

$${}_{10}S_{20,24} = \frac{{}_5N_{30}^2}{{}_5N_{20}^1} = \frac{40,885}{44,826} = 0.9121$$

where  ${}_5N_x^1$  and  ${}_5N_{x+10}^2$  are the populations aged from  $x$  to  $x+4$  at the first census and from  $x+10$  to  $x+14$  at the time of the second census, respectively. Results for all age groups are shown in column (5) of table 173. Note that the female population was classified by five-year age group only up to age 74, with an open-ended age group 75 and over. Since those over 75 in 1970 are the survivors of those over 65 in 1960, the last survivorship ratio is a 10-year survivorship probability for those 65 and over in 1960.

**Step 5: fitting of a Coale-Demeny model life table.** The cohort survivorship ratios given in column (5) of table 173 show a certain amount of variability; and one ratio even has a value that, in the absence of migration, would be impossible (greater than 1.0). The fitting of a model life table is therefore desirable. The mortality level associated with each estimate (excluding those which are impossible or out of range) can be found in a family of Coale-Demeny model life tables. In the case of Panama, the West family is selected as the most suitable.

It is assumed that the cohort survivorship probabilities,  ${}_{10}S_{x,x+4}$ , are equivalent to stationary-population survivorship ratios,  ${}_5L_x+10/{}_5L_x$ . These stationary sur-

vivorship probabilities or ratios are not tabulated in the Coale-Demeny tables, but their values are shown in the second half of annex X (tables 271-278). Table 174 shows the cohort survivorship ratios (taken from column (5) of table 173), the stationary-population survivorship ratios (hereafter also called "model ratios") for a range of mortality levels of the West family of model life tables, and the levels implied by the cohort ratios, obtained by interpolating between the model values. The interpolation is straightforward: if the cohort ratio falls between levels  $\nu$  and  $\nu+1$ , the interpolated level  $z$  is found as

$$z(x) = \nu + \frac{{}_{10}S_{x,x+4} - [{}_5L_x^\nu + 10 / {}_5L_x^\nu]}{[{}_5L_x^{\nu+1} + 10 / {}_5L_x^{\nu+1}] - [{}_5L_x^\nu + 10 / {}_5L_x^\nu]} \quad (\text{B.2})$$

where  $\nu$  and  $\nu+1$  indicate the mortality levels of the ratios. If the interval between the levels to which the model ratios correspond is two levels, as shown in table 174, the term to be added to  $\nu$  has to be multiplied by two.

It will be seen that two of the implied levels shown in table 174 are high (above 20) and two are below 14. Discarding these outlying values, an estimate of overall level can be obtained by averaging the remaining estimates; in this case, the estimate obtained is 16.1. This level is then taken as a best estimate of the level of mortality after age 10 for females in Panama, on the basis of the two census enumerations.

**Step 6: completion of the life table.** So far, the level of adult mortality has been estimated, but not the level of child mortality. If no information is available about child mortality, the best that can be done is to assume that the adult level also applies in childhood and to adopt the complete life table of the estimated level; in the case of Panama, this level is 16.1.

If some information on child mortality is available, estimates of adult survivorship obtained from a model life table of the estimated level can be linked with the independent estimate of survivorship to age 5. In chapter III, the level of female child mortality in Panama in the 1960s was estimated to be 18.05 (according to the West model), so a life table at this level is adopted up to age 5. Thus,  $l(1)=0.9405$  and  $l(5)=0.9165$ . Probabilities of survival from age 5 onward are then calculated for level 16.1 and are used to extend the life table from age 5, as shown in table 175.

For comparison purposes, the life-table  $l(x)$  function, which will be generated in subsection B.3 using the logit life-table system, is also included. As can be seen, the two life tables are slightly different in detail but very similar in broad shape and level.

(d) *Second detailed example: Colombia, 1951-1964*

The second detailed example considers a less tidy case, where age-misreporting is more extensive and the intercensal period is not a convenient multiple of five.

The computational procedure is described below.

TABLE 174. DETERMINATION OF THE MORTALITY LEVEL IMPLIED BY EACH COHORT SURVIVORSHIP RATIO, PANAMA, 1960-1970

Age $x$ (1)	Cohort survivorship ratio $10^5 S_{x, x+4}$ (2)	Stationary-population survivorship ratio, $5L_x + 10/5L_x$ , for West model mortality levels					Implied mortality level (8)
		14 (3)	16 (4)	18 (5)	20 (6)	22 (7)	
0.....	0.9636	0.9382	0.9568	0.9723	0.9854	0.9941	16.9
5.....	0.9753	0.9668	0.9758	0.9836	0.9905	0.9956	14.9
10.....	1.0080	0.9602	0.9705	0.9796	0.9875	0.9941	<sup>a</sup>
15.....	0.9524	0.9497	0.9623	0.9734	0.9831	0.9919	14.4
20.....	0.9121	0.9415	0.9556	0.9683	0.9794	0.9897	<14.0
25.....	0.9722	0.9341	0.9493	0.9631	0.9754	0.9868	18.5
30.....	0.9304	0.9260	0.9417	0.9563	0.9697	0.9822	14.6
35.....	0.8988	0.9160	0.9316	0.9465	0.9604	0.9743	<14.0
40.....	0.9246	0.8987	0.9147	0.9303	0.9454	0.9613	17.3
45.....	0.8706	0.8688	0.8866	0.9045	0.9220	0.9414	14.2
50.....	0.8786	0.8210	0.8422	0.8639	0.8855	0.9101	19.4
55.....	0.8536	0.7489	0.7741	0.8002	0.8266	0.8580	21.7
60.....	0.6623	0.6473	0.6750	0.7042	0.7342	0.7719	15.1
65.....	0.5365 <sup>b</sup>	<sup>c</sup>	<sup>c</sup>	<sup>c</sup>	<sup>c</sup>	<sup>c</sup>	<sup>c</sup>

Note: Mean level (excluding two lowest and two highest values) = 16.1.

<sup>a</sup> Survivorship ratio in excess of 1.0.

<sup>b</sup>  $10^5 S_{65+}$ .

<sup>c</sup> Not computed.

TABLE 175. COMPLETION OF AN INTERCENSAL LIFE TABLE USING THE COALE-DEMEYI MODEL LIFE TABLES, PANAMA, 1960-1970

Age $x$ (1)	Values of probability of surviving to age $x$ , $l(x)$ , West female model life tables				Final estimates of probability of surviving to age $x$	
	Level 16 (2)	Level 17 (3)	Level 16.1 (4)	Level 16.1 $l(x+5)/l(x)$ (5)	Coale- Demeny (6)	Logit system (7)
	0.....	<sup>a</sup>	<sup>a</sup>	<sup>a</sup>	<sup>a</sup>	1.0000
1.....	<sup>a</sup>	<sup>a</sup>	<sup>a</sup>	<sup>a</sup>	0.9405	0.9323
5.....	0.8804	0.8985	0.8822	0.9870	0.9165	0.9046
10.....	0.8687	0.8887	0.8707	0.9899	0.9046	0.8954
15.....	0.8598	0.8811	0.8619	0.9854	0.8955	0.8883
20.....	0.8470	0.8701	0.8493	0.9808	0.8824	0.8777
25.....	0.8305	0.8558	0.8330	0.9780	0.8655	0.8638
30.....	0.8119	0.8394	0.8147	0.9748	0.8465	0.8478
35.....	0.7912	0.8211	0.7942	0.9710	0.8262	0.8297
40.....	0.7680	0.8001	0.7712	0.9663	0.8013	0.8088
45.....	0.7418	0.7760	0.7452	0.9585	0.7743	0.7841
50.....	0.7107	0.7466	0.7143	0.9434	0.7422	0.7533
55.....	0.6702	0.7075	0.6739	0.9222	0.7002	0.7121
60.....	0.6177	0.6560	0.6215	0.8838	0.6457	0.6575
65.....	0.5454	0.5843	0.5493	0.8287	0.5707	0.5817
70.....	0.4515	0.4889	0.4552	0.7399	0.4729	0.4813
75.....	0.3335	0.3663	0.3368	0.6173	0.3499	0.3547
80.....	0.2055	0.2297	0.2079	-	0.2160	0.2179

<sup>a</sup> Not calculated.

*Step 1: adjustment for net intercensal migration and territorial coverage.* Once again, no basis exists for adjusting for migration, and no adjustment for territorial coverage is necessary.

*Step 2: grouping of data from the two censuses by cohort.* Population censuses were held in Colombia on 9 May 1951 and on 15 July 1964, the intercensal interval thus being 13.185 years. Since the population distribution by single year of age is available from the 1964 census, standard five-year age groups from  $x$  to  $x + 4$  identified at the first census can be reidentified at the second census as age groups from  $x + 13$  to  $x + 17$ . Thus, survivors of those aged 0-4 at the first census are aged 13-17 years at the second census. Suitably grouped data for the female population, based on a single-year age distribution for the population enumerated in 1964, are shown in columns (2) and (4) of table 176.

*Step 3: adjustment for length of the intercensal interval.* As the intercensal interval was 13.185 years, the population enumerated at the second census can be moved back 0.185 of a year to improve comparability. The total female population in 1964 was 8,869,856, whereas that in 1951 was 5,649,250. Thus, the overall growth rate of the female population was

$$r = \frac{\ln(8,869,856) - \ln(5,649,250)}{13.185} = 0.0342.$$

The adjustment factor  $k$  for the second census is thus

$$k = \exp[(0.0342)(-0.185)] = 0.9937.$$

Column (5) of table 176 shows the 1964 population systematically multiplied by the factor  $k$  (constant with respect to age).

*Step 4: calculation of cohort survivorship ratios.* Survivorship ratios for each cohort are calculated by divid-

ing the number of survivors at the second census (after adjustment) by the corresponding number at the time of the first census. Thus, in table 176, the numbers in column (5) are divided by the numbers in column (2). The results are shown in column (6). Note that the survivorship ratios for the first three age groups in 1951 exceed 1.0, indicating the existence of problems related to coverage or to age-reporting.

*Step 5: fitting of a Coale-Demeny model life table.* The cohort survivorship ratios given in column (6) of table 176 are for five-year cohorts over a period of 13 years. Comparable ratios for stationary populations are not published in the Coale-Demeny life tables, nor can they be calculated directly from information which is published. It is possible to estimate them, but the calculations necessary are rather heavy, particularly if the variations in level are substantial. Time can be saved by finding the approximate mortality level for each cohort survivorship ratio and then estimating more accurately the model survivorship ratios (referring to stationary populations) for adjacent mortality levels in order to perform the final interpolation.

The first step is to find the approximate mortality level to which each cohort ratio corresponds. In the stationary populations, a 13-year survivorship ratio should lie almost half-way between the 10-year and 15-year ratios, both of which can be calculated directly from the published tables. To give an example, consider the cohort aged 40-44 in 1951 whose 13-year survivorship ratio is estimated to be 0.7885. By trial and error, one can find the approximate level to which this cohort ratio corresponds in the West family of model life tables. At level 10, the 10-year female survivorship ratio,  ${}_5L_{50}/{}_5L_{40}$ , has a value of 235,666/273,796 or 0.8607, whereas the 15-year ratio,  ${}_5L_{55}/{}_5L_{40}$ , has a value of 0.7719. The average of the two, 0.8163, is higher than the cohort survivorship ratio, so the approximate mor-

TABLE 176. FEMALE POPULATION BY AGE GROUP, 1951 AND 1964; ADJUSTED FEMALE POPULATION, 1964; AND COHORT SURVIVORSHIP RATIOS, COLOMBIA

Age group (1)	1951 census		1964 census		Cohort survivorship ratio (6)
	female population (2)	Age group (3)	Female population (4)	Adjusted female population <sup>a</sup> (5)	
0-4	921 747	13-17	988 879	982 649	1.0661
5-9	768 958	18-22	810 294	805 189	1.0471
10-14	657 753	23-27	676 158	671 898	1.0215
15-19	605 411	28-32	566 038	562 472	0.9291
20-24	550 555	33-37	499 191	496 046	0.9010
25-29	447 242	38-42	419 842	417 197	0.9328
30-34	337 311	43-47	307 196	305 261	0.9050
35-39	334 197	48-52	284 416	282 624	0.8457
40-44	239 771	53-57	190 266	189 067	0.7885
45-49	196 659	58-62	175 311	174 207	0.8858
50-54	175 580	63-67	124 648	123 863	0.7055
55-59	105 721	68-72	91 170	90 596	0.8569
60-64	116 939	73-77	49 108	48 799	0.4173
65-69	63 339	78-82	39 882	39 631	0.6257
70-74	57 175	83-87	16 949	16 842	0.2946
75-79	27 398	88-92	10 213	10 149	0.3704
80-84	24 807	93-97	3 642	3 619	0.1459
85+	18 687	98+	3 199	3 179	0.1701

<sup>a</sup> Adjusted for the difference between the observed intercensal period, 13.185 years, and 13 exact years.

tality level of this ratio should be lower. At level 8, the equivalent 10-year, 15-year and average survivorship ratios are 0.8377, 0.7380 and 0.7879, the last value being very close to the cohort ratio; hence, level 8 is selected as the approximate level for the cohort ratio. Proceeding in this way, approximate levels for all the cohort ratios are estimated. They are shown in column (3) of table 177.

The next step is to make a more accurate estimate of the 13-year model survivorship ratios for the approximate levels determined above. Thus, for the cohort aged 15-19 in 1951, the approximate level is 13, so a model survivorship ratio,  ${}_5L_{28}/{}_5L_{15}$ , is to be calculated for level 13.

The value of  ${}_5L_{15}$  is tabulated in the published tables (for West females, level 13, it has the value of 3.89403). The constants and equation given in table 172 can be used to estimate  ${}_5L_{28}$ . The age range from 28 to 32 is covered by the tabulated  $l(x)$  values for 25, 30 and 35, and  $n$  is equal to 28 minus 25, that is, 3. In the West female model life table of level 13,  $l(25)=0.74769$ ,  $l(30)=0.72326$ , and  $l(35)=0.69647$ ; thus,

$$\begin{aligned} {}_5L_{28} &= a(3)l(25) + b(3)l(30) + c(3)l(35) \\ &= (-0.017)(0.74769) + (4.533)(0.72326) + \\ &\quad (0.483)(0.69647) \\ &= 3.60222. \end{aligned}$$

The model survivorship ratio is then calculated as

$${}_{13}S_{15,19}^m = {}_5L_{28}/{}_5L_{15} = 3.60222/3.89403 = 0.9251.$$

For the next cohort, aged 20-24 in 1951, the approximate mortality level is 11; therefore, the 13-year model survivorship ratio,  ${}_5L_{33}/{}_5L_{20}$ , is calculated for level 11. The value of  ${}_5L_{20}$  is obtained directly from the model life tables as 3.51543; that of  ${}_5L_{33}$  is estimated from

values of  $l(30)$ ,  $l(35)$  and  $l(40)$  for level 11 as follows:

$$\begin{aligned} {}_5L_{33} &= (-0.017)l(30) + (4.533)l(35) + (0.483)l(40) \\ &= (-0.017)(0.66224) + (4.533)(0.63186) + \\ &\quad (0.483)(0.59963) \\ &= 3.14258. \end{aligned}$$

The model survivorship ratio is then calculated as

$${}_{13}S_{20,24}^m = 3.14258/3.51543 = 0.8939.$$

Results for each cohort are shown in column (4) of table 177. Note that for the open-ended cohort, a cruder simplification is adopted. The model survivorship ratio required can be approximated as

$${}_{13}S_{65+}^m = T_{78}/T_{65},$$

that is, the stationary population over 78 divided by the stationary population over 65. The value for  $T_{65}$  is tabulated in the model tables, and  $T_{78}$  can be estimated by weighting the values of  $T_{75}$  and  $T_{80}$  in the following manner:

$$T_{78} = 0.6 T_{80} + 0.4 T_{75}.$$

The precision of the estimate for the open-ended interval need not be high, since the survivorship ratio for the open-ended cohort is likely to be distorted anyway by age-misreporting and age-distribution effects.

Once 13-year model survivorship ratios have been estimated for each approximate level given in column (3), the model survivorship ratio for an adjacent level has to be computed in order that the level of the cohort survivorship ratio can be found by interpolation. If the cohort survivorship ratio exceeds that computed for the approximate level, a model ratio should be estimated for the next higher mortality level; whereas if the cohort

TABLE 177. STEPS IN ESTIMATION OF THE MORTALITY LEVEL TO WHICH EACH COHORT SURVIVORSHIP RATIO CORRESPONDS IN THE WEST MODEL, COLOMBIA, 1951-1964

Age group in 1951 (1)	Cohort 13-year survivorship ratio (2)	West model, 13-year survivorship ratio				Estimated mortality level (7)
		Approximate level (3)	Estimated ratio (4)	Adjacent level (5)	Estimated ratio (6)	
0-4	1.0661	a	a	a	a	a
5-9	1.0471	a	a	a	a	a
10-14	1.0215	a	a	a	a	a
15-19	0.9291	13	0.9251	14	0.9321	13.6
20-24	0.9010	11	0.8939	12	0.9042	11.7
25-29	0.9328	16	0.9318	17	0.9410	16.1
30-34	0.9050	14	0.9014	15	0.9115	14.4
35-39	0.8457	10	0.8406	11	0.8533	10.4
40-44	0.7885	8	0.7805	9	0.7958	8.5
45-49	0.8858	20	0.8876	19	0.8758	19.8
50-54	0.7055	11	0.6998	12	0.7175	11.3
55-59	0.8569	24	0.8473	a	a	>24.0
60-64	0.4173	8	0.3983	9	0.4209	8.8
65+	0.3860	24	0.3245	a	a	>24.0

a Cannot be calculated.



ratio is lower than that computed for the approximate level, a model ratio should be estimated for the next lower level. Thus, for age group 15-19 the cohort ratio,  ${}_{13}S_{15, 19}$ , is 0.9291, whereas the model ratio,  ${}_{13}S_{15, 19}^m$ , for the approximate level 13 is 0.9251. In order to bracket the cohort ratio between estimated model ratios, it is therefore necessary to compute a model survivorship ratio for the next higher level, 14. (Had the cohort ratio  ${}_{13}S_{15, 19}$  been lower than 0.9251, a model ratio would instead have been computed for the next lower level, 12.) The steps to follow to perform this computation are the same as those followed in calculating the model ratios corresponding to the approximate levels, except that now the adjacent level is used. Thus, in the case at hand, the life table of level 14 is used instead of that of level 13. The model ratio obtained is 0.9321, and it is listed together with all other model ratios corresponding to the selected adjacent levels in column (6) of table 177. It is worth noting that, in this instance, the cohort survivorship ratios are generally higher than the model ratios corresponding to the approximate levels because the method used to select an approximate level approximates a 12.5-year interval rather than the actual 13-year interval. Fortunately, in every case, the cohort ratio was less than one level away from the first approximation selected.

The last step consists of interpolating between the model ratios for the approximate and adjacent levels in order to find the mortality level of the cohort ratio. As always, the amount to be added to the lower mortality level is equal to the difference between the cohort survivorship ratio and the model ratio corresponding to the lower mortality level, divided by the difference between the model ratios corresponding to the higher and lower mortality levels. Thus, if  $v$  is the lower mortality level,  $z(x)$  is the estimated level and  $v+1$  is the upper level,

$$z(x) = v + ({}_{13}S_{x, x+4} - {}_{13}S_{x, x+4}^v) / ({}_{13}S_{x, x+4}^{v+1} - {}_{13}S_{x, x+4}^v),$$

so that, for the cohort aged 15-19 in 1951,

$$\begin{aligned} z(15) &= 13.0 + (0.9291 - 0.9251) / (0.9321 - 0.9251) \\ &= 13.57. \end{aligned}$$

Complete results, rounded to one decimal place, are shown in column (7) of table 177.

*Step 6: completion of the life table.* The life table may be completed by adding information on child mortality in exactly the same way as for the example for Panama. However, the results obtained in step 5 are so erratic that they require some comment. For the first three cohorts, more survivors were recorded in 1964 than had been enumerated in 1951. It is probable that two factors were mainly responsible for this outcome: the general tendency to underenumerate young children or to exaggerate their ages; and the tendency to shift the ages of women into the peak reproductive years from either side. The first factor would reduce the initial numbers in the cohorts, and the second would increase the apparent numbers of survivors. Two older cohorts also show

implausibly high survivorship ratios, the cohort aged 55-59 in 1951, whose survivors include the number heaped on age 70, and the cohort corresponding to the open-ended interval (65 and over), whose survivors are probably inflated by age exaggeration. In order to obtain some overall average estimate of intercensal mortality, these five cases should be excluded; and to balance their exclusion, the five lowest levels should also be disregarded. Such Draconian elimination leaves as the only "acceptable" estimates the set 13.6, 16.1, 14.4 and 19.8, the average of which is 16.0. However, because of the fairly wide range of levels covered by these estimates, their average cannot be considered a reliable indicator of intercensal mortality.

To conclude, note that no independent information on child mortality is available for the period 1951-1964, so that if the average mortality level estimated above were reliable, a life table could only be completed by assuming that the mortality pattern embodied in the model used (West) adequately represents that experienced by the population being studied (in terms of both adult and child mortality).

### 3. Intercensal survivorship ratios for five-year age cohorts smoothed by use of the logit system

#### (a) Data required

The data required for this method are listed below:

(a) Two census enumerations separated by  $t$  years with populations classified by five-year age group (and sex);

(b) An independent estimate of child mortality. Such estimates are generally derived from information on children ever born and children surviving analysed according to the procedures described in chapter III.

#### (b) Computational procedure

The steps of the computational procedure are described below.

*Steps 1-4.* These steps, by which cohort survivorship ratios analogous to those for a stationary population,  ${}_5L_{x+t}/{}_5L_x$ , are calculated, are identical to those described in subsection B.2(b) and are not repeated here.

*Step 5: smoothing cohort survivorship ratios by use of the logit life-table system.* Somewhat more flexibility in the model pattern of mortality used can be introduced by smoothing through the logit life-table system (see chapter I, subsection B.4). The cohort survivorship ratios, analogous to  ${}_5L_{x+t}/{}_5L_x$ , are transformed into estimates of  ${}_5L_{x+t}$  by multiplying each by the corresponding  ${}_5L_x$ ; the first value or values of  ${}_5L_x$  are estimated on the basis of information about child mortality, and subsequent values are obtained from previous estimates of  ${}_5L_{x+t}$ . The calculations thus form a chain: the first value of  ${}_5L_{x+t}$  is calculated by assuming a value of  ${}_5L_0$ ; and if  $t$  is greater than 5, the second value of  ${}_5L_{x+t}$  can be obtained by assuming a value of  ${}_5L_5$ ; but once  $x$  is greater than  $t$ , the denominators  ${}_5L_x$  will be provided by earlier estimates of  ${}_5L_{x+t}$ .

Once a series of  ${}_5L_{x+t}$  values has been obtained, it is

assumed that the proportion of the stationary population aged from  $x + t$  to  $x + t + 5$  approximates the probability of surviving from birth to age  $y = x + t + 2.5$ ,  $l(x + t + 2.5)$ . The logit transformations of each of these  $l(y)$  estimates can then be calculated and compared with the logit transformations of equivalent values derived from an appropriate standard life table. Then, the  $\alpha$  and  $\beta$  parameters defining the linear relationship between the logit transformations of the estimated and the standard survivorship probabilities can be estimated by using a suitable line-fitting procedure, and a complete  $l(x)$  survivorship function can be generated. It must be noted, however, that for childhood ages the  $l(x)$  values generated in this way will not, in general, coincide with those used as input in applying the method. The magnitude of the differences between the input and output child mortality estimates depends, among other things, upon the appropriateness of the mortality pattern used as standard and upon the quality of the intercensal survivorship estimates. If the magnitude of such differences is unacceptably large, the use of a different standard should be considered. When changes in the standard fail to reduce the differences observed, one may either have to discard entirely the intercensal life table or one may adopt a life table built by linking in a manner similar to that described in step 6 in subsection B.2(b), the child mortality estimates used as input with the estimated  $l(x)$  values over age 10.

(c) *First detailed example: Panama, 1960-1970*

The computational procedure for this example is described below.

*Steps 1-4.* These steps have already been presented in subsection B.2(c).

*Step 5: smoothing cohort survivorship ratios by use of the logit life-table system.* The starting-point of this smoothing procedure is the cohort survivorship ratios shown in column (5) of table 173. It is assumed that a cohort survivorship ratio approximates a life-table survival probability from the central age of the cohort at the first census to its central age at the second census. In the case in hand, therefore, it is assumed that

$${}_{10}S_{x, x+4} = \frac{{}_5L_{x+10}}{{}_5L_x} = \frac{5.0[l(x+12.5)]}{5.0[l(x+2.5)]} \quad (\text{B.3})$$

The one age group for which  $5.0[l(x+2.5)]$  is not an adequate approximation to  ${}_5L_x$  is group 0-4. If estimates of  $l(2)$ ,  $l(3)$  and  $l(5)$  are available from child survival data, the mortality levels associated with these three estimates in a selected family of Coale-Demeny model life tables can be averaged; and the value of  ${}_5L_0$  for that level, sex and family can be read off from the relevant table. If an estimate of  ${}_5L_5$  is required, it should be taken also from this model table.

In the case of Panama, no information relevant to the estimation of child mortality was collected by either the 1960 or the 1970 census. However, child mortality estimates are available from the Retrospective Demographic Survey conducted in 1976, and the reference

period for the estimates derived from information pertaining to women aged from 30-34 to 45-49 effectively covers the intercensal period 1960-1970 (see table 55). The average female mortality level for this period is 18.05 in the West family of model life tables. For level 18,  ${}_1L_0$  is 0.95377 and  ${}_4L_1$  is 3.69742; for level 19, the corresponding values are 0.96004 and 3.75407. Therefore, interpolating linearly,

$${}_1L_0^{18.05} = (0.95)(0.95377) + (0.05)(0.96004) = 0.9541$$

and

$${}_4L_1^{18.05} = (0.95)(3.69742) + (0.05)(3.75407) = 3.7003.$$

Interpolating in a similar fashion for  ${}_5L_5^{18.05}$ ,

$$\begin{aligned} {}_5L_5^{18.05} &= (0.95)({}_5L_5^{18}) + (0.05)({}_5L_5^{19}) \\ &= (0.95)(4.55832) + (0.05)(4.64405) \\ &= 4.5626. \end{aligned}$$

The chaining of survivorship probabilities can then begin. It is assumed that

$${}_{10}S_{0, 4} = \frac{5.0[l(12.5)]}{{}_5L_0}$$

so

$$l(12.5) = \frac{1}{5}({}_{10}S_{0, 4})({}_5L_0)$$

and

$${}_5L_x = 5.0 l(x + 2.5) \quad \text{for } x = 10, 15, \dots, 60.$$

Table 178 shows the full calculations.

The essence of the logit life-table system lies in the comparison of an estimated  $l(x)$  survivorship function with a standard  $l_s(x)$  function on the logit scale (see chapter I, subsection B.4). In this case, a Coale-Demeny West model life table of level 18 for females has been selected as standard on the basis of the child mortality estimates available. The  $l(x)$  function in the Coale-Demeny model life tables is given only for ages 0, 1, 5, 10 and so on, whereas for comparison with the estimated probabilities,  $l(y)$  values are required for ages  $y = 12.5, 17.5, 22.5$  and so on. These values can be obtained by averaging the standard  $l_s(x)$  values for adjacent ages  $x$  (multiples of five) and then calculating the logit transformations of these averages. Thus, to obtain an estimate of  $l_s(12.5)$ ,

$$\begin{aligned} l_s(12.5) &= 0.5(l_s(10) + l_s(15)) = 0.5(0.90762 + \\ &0.90136) = 0.90449. \end{aligned}$$

Therefore, the logit transformation of the standard at 12.5,  $\lambda_s(12.5)$ , is

$$\lambda_s(12.5) = 0.5 \ln \left[ \frac{1.0 - 0.90449}{0.90449} \right] = -1.1241.$$

TABLE 178. APPLICATION OF SMOOTHING PROCEDURE BASED ON THE LOGIT SYSTEM TO THE COHORT SURVIVORSHIP RATIOS FOR THE PERIOD 1960-1970, PANAMA

Age $x$ (1)	Cohort survivorship ratio $10^5 l_x, x+4$ (2)	Stationary population		Logit transformation	
		Five-year age groups $5L_x$ (3)	Exact age $l(x+12.5)$ (4)	Estimated $\lambda(x+12.5)$ (5)	Standard <sup>a</sup> $\lambda_s(x+12.5)$ (6)
0.....	0.9636	(4.6544) <sup>b</sup>	0.8970	-1.0822	-1.1241
5.....	0.9753	(4.5626) <sup>b</sup>	0.8900	-1.0454	-1.0807
10.....	1.0080	4.4850	0.9042	-1.1224	-1.0252
15.....	0.9524	4.4499	0.8476	-0.8580	-0.9633
20.....	0.9121	4.5209	0.8247	-0.7743	-0.8988
25.....	0.9722	4.2381	0.8241	-0.7722	-0.8314
30.....	0.9304	4.1235	0.7673	-0.5966	-0.7595
35.....	0.8988	4.1203	0.7407	-0.5248	-0.6795
40.....	0.9246	3.8365	0.7094	-0.4462	-0.5852
45.....	0.8706	3.7033	0.6448	-0.2981	-0.4713
50.....	0.8786	3.5472	0.6233	-0.2518	-0.3302
55.....	0.8536	3.2241	0.5504	-0.1011	-0.1529
60.....	0.6623	3.1166	0.4128	0.1762	0.0715
65.....	(0.5365)	2.7521	c	c	c
70.....	c	2.0641	c	c	c
75.....	c	1.4765	c	c	c

<sup>a</sup> Female, level 18, West model.  
<sup>b</sup> Obtained as described in the text.  
<sup>c</sup> Not computed.

Note that computing  $l_s(x+2.5)$  by directly averaging  $l_s(x)$  and  $l_s(x+5)$  has an advantage over the use of closer approximations, since it is being assumed in the case of the estimated  $l(x)$  values that  $5L_x$  is equal to  $5.0(l(x+2.5))$ . If the  $l(x)$  and  $l_s(x)$  functions are similar, but not linear, between  $x$  and  $x+5$ , the logit transformations of the estimated and the standard  $l(y)$  values calculated as shown above will still be comparable, but will not refer to exact age  $x+2.5$ . Yet, the form of their relationship (its linearity and the parameters that define it) will not be greatly affected.

Columns (5) and (6) of table 178 show the logit transformations of the estimated and standard  $l(x)$  functions, respectively. The points defined by  $[\lambda_s(x), \lambda(x)]$  are plotted in figure 21. It can be seen that these points follow a generally linear trend, though there are outliers, notably the point associated with the cohort aged 10-14 in 1960, which had an apparent survivorship ratio from 1960 to 1970 greater than 1.0. A straight line has been fitted to the points by using group means (see chapter V, subsection C.4). This line has a slope of 1.022 (an estimate of  $\beta$  in the logit system) and an intercept of 0.094 (an estimate of  $\alpha$  in the logit system). The intercensal cohort survivorship probabilities thus indicate an age pattern of mortality similar to that of the standard, since  $\beta$  is roughly equal to one, but an overall level of mortality somewhat heavier than that of the standard, since  $\alpha$  is slightly larger than zero.

A complete life table can now be calculated by inverting the logit transformation estimated by means of  $\alpha$  and  $\beta$ . Thus,

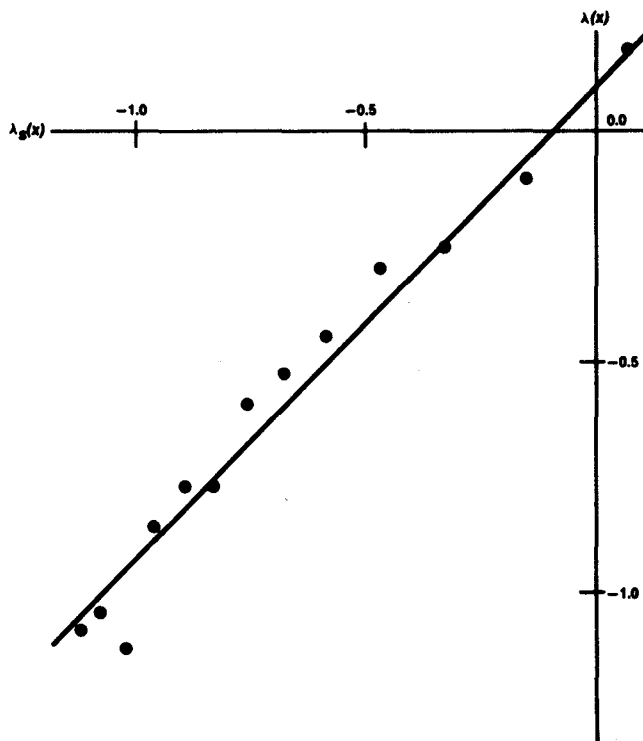
$$\lambda^*(x) = \alpha + \beta \lambda_s(x)$$

and

$$l(x) = [1.0 + \exp(2.0 \lambda^*(x))]^{-1}$$

The results obtained in this case are shown in column (7) of table 175. Note that, as mentioned earlier, the child mortality estimates obtained by this procedure differ from those used as input (the latter are shown as  $l(1)$  and  $l(5)$  in column (6) of table 175). In terms of infant mortality, for example, the logit estimate is  $l(1) = 0.9323$ , while the value used as input is 0.9405

Figure 21. Plot of the logit transformation of the estimated survivorship function,  $l(x)$ , against that of the standard, West model for females, level 18; Panama



(corresponding to level 18.05 in the Coale-Demeny West models for females). Hence, whereas according to the life table generated by using the logit system, some 68 out of every 1,000 births die before reaching age one, according to the estimates derived from data on children ever born and surviving, about 60 deaths per 1,000 births are expected. If the latter estimate were correct, the one obtained through the logit fit overestimates infant mortality by about 13 per cent. This outcome is due to the fact that intercensal adult mortality is substantially higher than the child mortality used as input in terms of the West mortality pattern. In the case of Panama during the period 1960-1970, the fairly low child mortality estimates derived from reports of older women are likely to be biased downward. Therefore, the estimates yielded by the logit fit are probably acceptable.

It should also be pointed out that the chaining of survivorship ratios used in this procedure introduces a substantial element of smoothing into the results even before the smoothing action of the logit system is introduced. Each link in the chain depends upon one or more of the earlier links, and each  $l(x + 2.5)$  estimate is determined both by an earlier estimate of  $l(x - 7.5)$  and by the intercensal cohort survivorship probability  ${}_{10}S_{x-10, x-6}$ . Thus, the final estimates yielded by this procedure are likely to be smoother than those obtained directly from each survivorship probability, as was done in subsection B.2(c).

(d) *Second detailed example: Colombia, 1951-1964*

The computational procedure for this example is described below.

*Steps 1-4.* These steps have already been covered in subsection B.2(d). Therefore, this example begins with the application of this procedure once the 13-year cohort survivorship ratios shown in table 176 are available.

*Step 5: smoothing the cohort survivorship ratios by use of the logit life-table system.* To use the smoothing procedure based on the logit system, it is necessary to have some estimate of child mortality in order to begin the chaining of survivorship probabilities. A recent study by Somoza,<sup>2</sup> based on the results of the Colombian National Fertility Survey (part of the World Fertility Survey), found that mortality among female children born during the period 1941-1959 could be approximated by level 14.5 of the West model life tables; for female children born during the period 1960-1967, the same procedure yielded a level of 16.2. Therefore, it is estimated that, for the intercensal period 1951-1964, female child mortality was, on average, equal to that of level 15.35 (the arithmetic average of the two levels estimated by Somoza). For this level,  ${}_5L_0$  is equal to 4.4699,  ${}_5L_5$  to 4.3076 and  ${}_5L_{10}$  to 4.2508, these values being obtained by interpolating linearly between the tabulated values for levels 15 and 16, respectively.

<sup>2</sup> Jorge L. Somoza, *Illustrative Analysis: Infant and Child Mortality in Colombia*, World Fertility Survey Scientific Report No. 10 (Voorburg, The Hague, International Statistical Institute, 1980).

The first cohort survivorship ratio,  ${}_{13}S_{0,4}$ , is regarded as being equivalent to the ratio  ${}_5L_{13}/{}_5L_0$ . Therefore, multiplying by the assumed value of  ${}_5L_0$ , one obtains an estimate of  ${}_5L_{13}$ :

$${}_5L_{13} = ({}_{13}S_{0,4})({}_5L_0) = (1.0661)(4.4699) = 4.7654.$$

Similar calculations for  ${}_5L_{18}$  and  ${}_5L_{23}$  give results of 4.5105 and 4.3422, using the assumed values of  ${}_5L_5$  and  ${}_5L_{10}$ , respectively. The next survivorship ratio,  ${}_{13}S_{15,19}$ , is regarded as equivalent to  ${}_5L_{28}/{}_5L_{15}$ , but no value has been assumed for  ${}_5L_{15}$ , which therefore has to be estimated. Estimates of  ${}_5L_{13}$  and  ${}_5L_{18}$  have already been obtained, and  ${}_5L_{15}$  can be estimated from them by weighting them suitably. The age interval from 13 to 17 shares three years, or 60 per cent, with age group 15-19; and the interval from 18 to 22 shares two years, or 40 per cent, with age group 15-19. An estimate of  ${}_5L_{15}$  can therefore be obtained by summing 60 per cent of  ${}_5L_{13}$  and 40 per cent of  ${}_5L_{18}$ :

$$\begin{aligned} {}_5L_{15} &= 0.6 {}_5L_{13} + 0.4 {}_5L_{18} = (0.6)(4.7654) + \\ &\quad (0.4)(4.5105) = 4.6634. \end{aligned}$$

An estimate of  ${}_5L_{20}$  can be obtained in a similar way from the estimated values of  ${}_5L_{18}$  and  ${}_5L_{23}$ :

$$\begin{aligned} {}_5L_{20} &= 0.6 {}_5L_{18} + 0.4 {}_5L_{23} = (0.6)(4.5105) + \\ &\quad (0.4)(4.3422) = 4.4432. \end{aligned}$$

These values of  ${}_5L_{15}$  and  ${}_5L_{20}$  can now be used to estimate  ${}_5L_{28}$  and  ${}_5L_{33}$  from the cohort survivorship ratios  ${}_{13}S_{15,19}$  and  ${}_{13}S_{20,24}$ , whereupon the values of  ${}_5L_{23}$ ,  ${}_5L_{28}$  and  ${}_5L_{33}$  can be used to estimate  ${}_5L_{25}$  and  ${}_5L_{30}$ , which can in turn be used to estimate  ${}_5L_{38}$  and  ${}_5L_{43}$  from the cohort survivorship ratios  ${}_{13}S_{25,29}$  and  ${}_{13}S_{30,34}$ , and so on until all but the last of the cohort ratios have been used (the last ratio, for the open-ended cohort aged 65 and over in 1951, cannot be equated with an  ${}_nL_x$  value and therefore cannot be used). The results of the various sets of calculations are shown in table 179.

It is assumed that the probability of surviving from birth to the mid-point of each age group can be approximated by one fifth of the  ${}_5L_x$  values. That is,

$$l(x + 2.5) = (0.2) {}_5L_x.$$

Column (5) of table 179 shows estimates of  ${}_5L_x$  for values of  $x$  of 13, 18, 23, 28 and so on up to 73. Each can be divided by five to estimate values of  $l(x + 2.5)$ , or survival probabilities from birth to ages 15.5, 20.5, 25.5 and so forth up to 75.5. Thus, for example,  $l(20.5)$ , is estimated as

$$l(20.5) = \frac{1}{5} {}_5L_{18} = \frac{1}{5} (4.5105) = 0.9021.$$

Full results are shown in column (6) of table 179.

The final stage of the smoothing process is the comparison of the logit transformations of the estimated sur-

TABLE 179. SMOOTHING OF FEMALE COHORT SURVIVORSHIP RATIOS BY USE OF THE LOGIT LIFE-TABLE SYSTEM, COLOMBIA, 1951-1964

Age group of cohort in 1951 (1)	Age x (2)	Cohort survivorship ratio $13^S_{x,x+4}$ (3)	Stationary population			Logit transformation	
			Standard five-year age group $5^L_x$ (4)	Non-standard five-year age group $5^L_{x+13}$ (5)	At an exact age $l(x+15.5)$ (6)	Estimated $\lambda(x+15.5)$ (7)	Standard $\lambda_s(x+15.5)^a$ (8)
0-4	0	1.0661	(4.4699) <sup>b</sup>	4.7654	0.9531	-1.5059	-0.8142
5-9	5	1.0471	(4.3076) <sup>b</sup>	4.5105	0.9021	-1.1104	-0.7615
10-14	10	1.0215	(4.2508) <sup>b</sup>	4.3422	0.8684	-0.9434	-0.6997
15-19	15	0.9291	4.6634	4.3328	0.8666	-0.9356	-0.6358
20-24	20	0.9010	4.4432	4.0033	0.8007	-0.6953	-0.5702
25-29	25	0.9328	4.3384	4.0469	0.8094	-0.7231	-0.5207
30-34	30	0.9050	4.2010	3.8019	0.7604	-0.7774	-0.4323
35-39	35	0.8457	4.0207	3.4003	0.6801	-0.3771	-0.3549
40-44	40	0.7885	3.9489	3.1137	0.6227	-0.2505	-0.2606
45-49	45	0.8858	3.6413	3.2255	0.6451	-0.2988	-0.1459
50-54	50	0.7055	3.2857	2.3181	0.4636	0.0729	0.0035
55-59	55	0.8569	3.1584	2.7064	0.5413	-0.0828	0.1936
60-64	60	0.4173	2.8625	1.1945	0.2389	0.5794	0.4457
65+	65	0.3860	2.4734	-	-	-	-

<sup>a</sup> Females, West model of level 15.

<sup>b</sup> Obtained independently from child mortality estimates for the period.

ivorship probabilities (equivalent to the stationary population at exact ages) with those of an adequate standard. The logit transformations of the  $l(x)$  values shown in column (6) of table 179 are easily calculated; for instance, for  $l(20.5)$ ,

$$\lambda(20.5) = 0.5 \ln[(1.0 - l(20.5)) / l(20.5)] = 0.5 \ln(0.10852) = -1.1104.$$

The full set of  $\lambda(x)$  values is shown in column (7) of table 179. The selection of a suitable standard and the calculation of  $l_s(x)$  values that match the estimated values present a problem, however, since the life tables published in the Coale-Demeny set do not include the values of  $x$  required in this case. Estimates of  $l_s(x)$  must be obtained for  $x = 15.5, 20.5$  and so on, so that they may be compared with the estimated  $l(x)$  function on the logit scale. In this application, a West model life table of level 15 for females has been adopted as standard. The required estimates of  $l_s(x)$  are obtained by interpolating linearly between the published  $l_s(y)$  values. Thus, for  $x = 15.5$ ,

$$\begin{aligned} l_s(15.5) &= 0.9 l_s(15) + 0.1 l_s(20) \\ &= (0.9)(0.83740) + (0.1)(0.82284) \\ &= 0.83594. \end{aligned}$$

The logit transformations of these  $l_s(x)$  values are calculated next, the results being shown in column (8) of table 179. The logit transformations of the estimated and standard  $l(x)$  functions are plotted one against the other in figure 22. The resulting points fluctuate substantially, that for the cohort aged 0-4 in 1951 being particularly divergent and the remainder following a broadly linear trend. The straight line shown in the figure has been

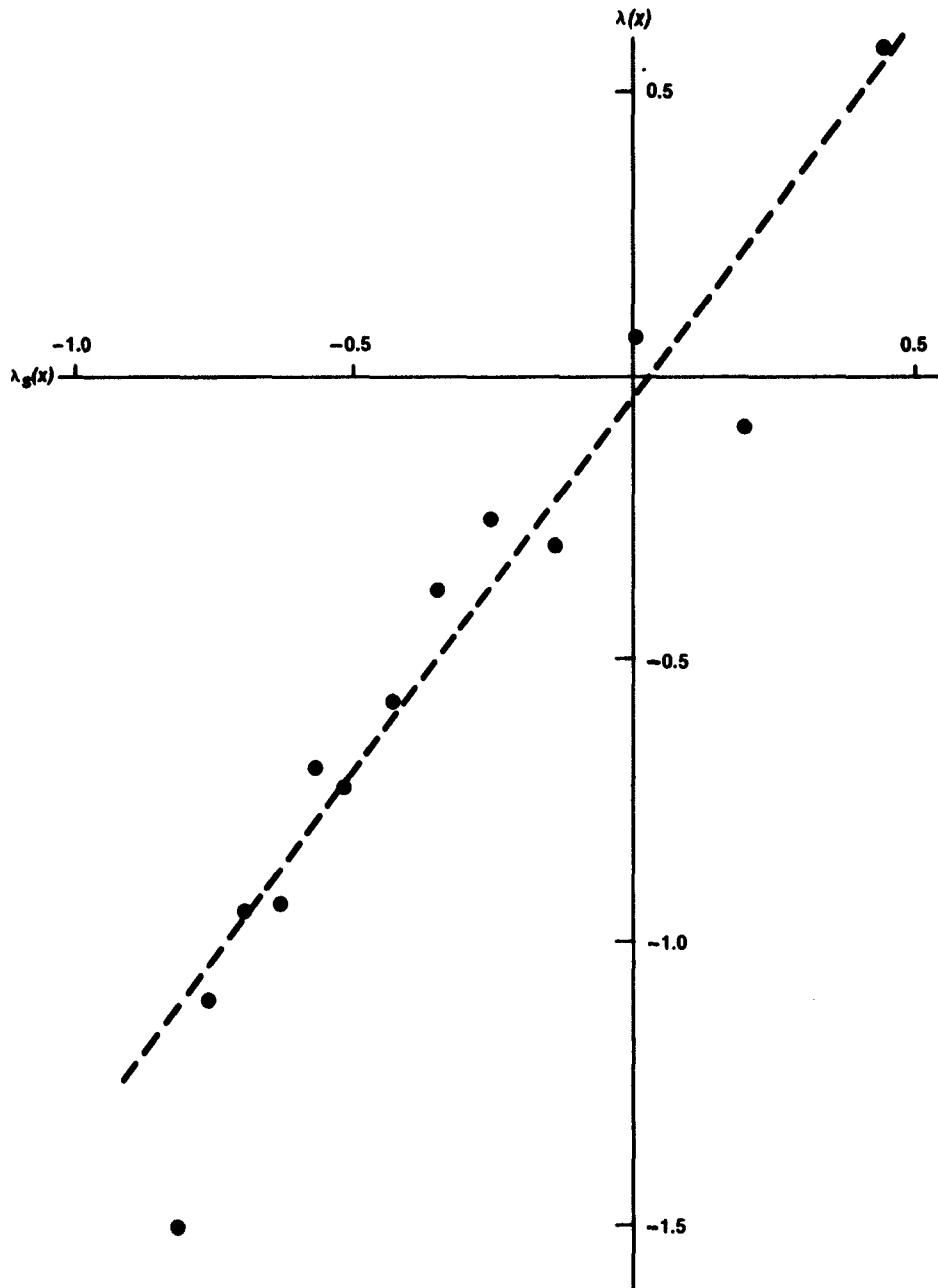
fitted by group means (see chapter V, subsection C.4) calculated on the basis of all the points except the first; it has an intercept ( $\alpha$ ) of -0.03 and a slope ( $\beta$ ) of 1.32, indicating that the level of mortality in the population in question is similar to that of the standard, but that observed mortality increases much more rapidly with age than it does in the standard. However, little confidence can be placed in the final results, given that the slope of the line is heavily affected by points referring to young women for whom age-reporting and coverage errors are substantial. The chaining process used by this smoothing procedure reduces the independence of the estimates derived from different cohorts, thus increasing the possibility of serious biases in the slope of the fitted line. Because of these problems, the calculation of a final life table is omitted. If desired, it would be calculated just as explained in step 5 of subsection B.3(c).

#### 4. Intercensal mortality estimated by using projection and cumulation

##### (a) General characteristics of method

Some of the effects of age-misreporting on intercensal mortality estimates can be eliminated by the use of cumulation. Instead of calculating survivorship ratios for cohorts, the initial population is projected forward to the date of the second census using a range of mortality levels. For each mortality level, the projected population over ages  $x$  from 10 or 15 to 50 or 55 is obtained by cumulation, and the observed population at the second census,  $N_2(x^+)$ , is used to interpolate within the projected values in order to determine the mortality level consistent with it. In this way, the effects of age-misreporting at the second census are limited to the effects of transfers across each age boundary  $x$ . Unfortunately, the effects of age-misreporting in the initial age distribution are not reduced.

Figure 22. Plot of the logit transformation of the estimated survivorship function,  $l(x)$ , against that of the standard, West model for females, level 15; Colombia



It should be noted in passing that there is no theoretical reason to prefer the forward projection of the first age distribution to make it comparable to the second age distribution over the reverse projection of the second age distribution to make it comparable to the first. The two procedures may, however, be expected to give somewhat different results because of different errors in the two distributions. The analyst might wish to carry out the calculations both ways, compare them and possibly obtain a final estimate by averaging the two results; but this discussion is confined to the description of the usual forward-projection procedure. Backward projection is carried out in an analogous way.

(b) *Data required*

The only data required are two census enumerations separated by  $t$  years with populations classified by age and sex.

(c) *Computational procedure*

The steps of the computational procedure are described below.

*Steps 1-3.* These steps, by which two comparable age distributions for time-points separated by an exact number of years are obtained, are identical to those described in subsection B.2(b) and are not repeated here.

**Step 4: cumulation of the second age distribution.** The age distribution from the first census is not cumulated, but the second age distribution is cumulated by summation from the uppermost age group downward. The ages above which populations should be cumulated will depend upon the intercensal interval  $t$ ; normally, the oldest initial population that should be used is the population aged 45 and over at the first census, and this population will be aged  $45+t$  and over at the second census; the next age group will be aged  $40+t$  and over, the next  $35+t$  and over, and so on.

**Step 5: projection of initial population with different mortality levels.** A suitable family of Coale-Demeny model life tables is selected, and life tables from different levels are used to project the initial population, five-year age group by five-year age group. If the intercensal interval is  $t$  years, and the initial population aged from  $x$  to  $x+4$  is  ${}_5N1_x$ , the population projected by using level  $\nu$  of the life tables,  ${}_5NP'_{x+t}$ , is given by:

$${}_5NP'_{x+t} = ({}_5N1_x)({}_5L'_{x+t} / ({}_5L'_x))$$

where  ${}_5L'_x$  and  ${}_5L'_{x+t}$  come from a model life table of level  $\nu$ . If  $t$  is not a multiple of five, values of  ${}_5L'_{x+t}$  can be estimated by using the technique described in step 5 of subsection B.2(b). Each age group has to be projected using several levels  $\nu$ , and the populations over each age  $x+t$  are then obtained by cumulation. The observed population aged  $x+t$  and over at the time of the second census,  $N2((x+t)+)$ , is then used to determine the level consistent with it by interpolating linearly between the projected estimates  $NP'((x+t)+)$ . Once mortality levels have been determined in this way for each initial age  $x = 5, 10, 15, \dots, 45$ , the median of these levels can be used as an estimate of adult mortality during the intercensal period.

**(d) A detailed example: Panama, 1960-1970**

The steps of the computational procedure are described below.

**Steps 1-3.** These steps have already been performed in subsection B.3(c) and need not be repeated here. Therefore, the starting-point of this example is the 1960 and 1970 female age distributions for Panama given in columns (3) and (4), respectively, of table 173.

**Step 4: cumulation of the second age distribution.** The Coale-Demeny model life tables tabulate the stationary-population function,  ${}_5L_x$ , only up to age 80, the final category being the stationary population aged 80 and over. For a 10-year survival period, therefore, the highest age group for which a model survivorship ratio can be calculated is the initial open-ended age group 70 and over. The initial age distribution therefore needs to be tabulated by five-year age group up to age group 65-69, with the last age group being 70 and over. For the final age distribution, however, less detail is required, since the highest initial age group to be used is that aged 45 and over. Thus, the final population in 1970 is required in age groups 55+ (survivors of the initial population aged 45 and over, 50+, 45+ and so on

down to 15+ (survivors of the initial population aged 5 and over). The easiest way to cumulate the age distribution is to begin with the number observed in the oldest age group and add in successively the number corresponding to the age group immediately below it. Thus, given the 1970 age distribution in column (4) of table 173, the population over age  $x$ ,  $N2(x+)$ , is calculated as

$$N2(x+) = N2((x+5)+) + {}_5N2_x$$

where  ${}_5N2_x$  is the population aged from  $x$  to  $x+4$  in 1970. For example, for  $x = 75$ ,

$$N2(75+) = 9,873;$$

and for  $x = 70$ ,

$$N2(70+) = N2(75+) + {}_5N2_{70} = 9,873 + 6,690 = 16,563.$$

Then for  $x = 65$ ,

$$N2(65+) = N2(70+) + {}_5N2_{65} = 16,563 +$$

$$10,061 = 26,624.$$

Full results are shown in column (15) of table 180.

**Step 5: projection of initial population with different mortality levels.** For a 10-year intercensal interval, the probability of surviving from the age group from  $x$  to  $x+4$  to the age group from  $x+10$  to  $x+14$  is approximated by  ${}_5L_{x+10} / {}_5L_x$ , and such model survivorship ratios can be calculated for each level of any family of Coale-Demeny model life tables. The last survivorship ratio, that for the population aged 70 and over to 80 and over 10 years later, is approximated as  $T_{80} / T_{70}$ , that is,  $T_{80} / ({}_5L_{70} + {}_5L_{75} + T_{80})$ . In a growing population with an age distribution 70 and over that is younger than the equivalent stationary population, this model survivorship ratio is likely to be somewhat lower than the true value, but an adjustment would not generally be worthwhile. Table 174 shows 10-year model survivorship probabilities for initial age groups up to 60-64 and for a range of mortality levels in the West family of model life tables; in order to project the entire population over any age, 10-year model survivorship probabilities for the initial age groups 65-69 and 70+ need to be added. For level 16, for example,

$${}_{10}S_{65,69}^{16} = {}_5L_{75}^{16} / {}_5L_{65}^{16} = 134,741 / 249,225 = 0.5406$$

and

$$\begin{aligned} {}_{10}S_{70+}^{16} &= T_{80}^{16} / T_{70}^{16} = T_{80}^{16} / [{}_5L_{70}^{16} + {}_5L_{75}^{16} + T_{80}^{16}] \\ &= 102,911 / (196,255 + 134,741 + 102,911) \\ &= 0.2372. \end{aligned}$$

In the application of forward projection with cumulation, there is no alternative to carrying out the calcula-

TABLE 180. PROJECTION OF THE INITIAL FEMALE POPULATION OVER 10 YEARS USING DIFFERENT MORTALITY LEVELS, PANAMA, 1960-1970

Age $x$ (1)	Female population aged from $x$ to $x+4$ in 1960 (2)	Projection at West level 16		Projection at West level 18		Projection at West level 20		Projection at West level 22		Actual 1970 population aged $x+10$ and over (15)	Interpolated mortality level (16)				
		Stationary- population age group ratio $5^Lx + 10^L / 5^Lx$ (3)	Projected 1970 population aged		Stationary- population age group ratio $5^Lx + 10^L / 5^Lx$ (6)	Projected 1970 population aged		Stationary- population age group ratio $5^Lx + 10^L / 5^Lx$ (9)	Projected 1970 population aged						
			From $x+10$ to $x+14$ (4)	$x+10$ and over (5)		From $x+10$ to $x+14$ (7)	$x+10$ and over (8)		From $x+10$ to $x+14$ (10)			$x+10$ and over (11)	From $x+10$ to $x+14$ (13)	$x+10$ and over (14)	
5.....	75 242	0.9758	73 421	393 970	0.9836	74 008	339 759	0.9905	74 527	405 151	0.9956	74 911	410 558	398 119	17.43
10.....	62 509	0.9705	60 665	320 549	0.9796	61 234	325 721	0.9875	61 728	330 624	0.9941	62 140	335 647	324 738	17.62
15.....	53 468	0.9623	51 452	259 884	0.9734	52 046	264 518	0.9831	52 564	268 896	0.9919	53 035	273 506	261 728	16.80
20.....	44 826	0.9556	42 836	208 432	0.9683	43 405	212 472	0.9794	43 902	216 332	0.9897	44 364	220 471	210 804	17.81
25.....	37 149	0.9493	35 266	165 596	0.9631	35 778	169 067	0.9754	36 235	172 429	0.9868	36 659	176 107	169 919	18.51
30.....	31 609	0.9417	29 766	130 331	0.9563	30 228	133 289	0.9697	30 651	136 194	0.9822	31 046	139 449	133 804	18.35
35.....	28 216	0.9316	26 286	100 565	0.9465	26 706	103 061	0.9604	27 099	105 543	0.9743	27 491	108 402	104 395	19.07
40.....	23 550	0.9147	21 541	74 278	0.9303	21 913	76 354	0.9454	22 264	78 444	0.9613	22 639	80 911	79 035	20.48
45.....	20 253	0.8866	17 956	52 737	0.9045	18 319	54 441	0.9220	18 673	56 180	0.9414	19 066	58 273	57 260	21.03
50.....	14 801	0.8422	12 465	34 781	0.8639	12 787	36 122	0.8855	13 106	37 507	0.9101	13 470	39 207	39 628	>22
55.....	11 787	0.7741	9 124	22 316	0.8002	9 432	23 336	0.8266	9 743	24 400	0.8580	10 113	25 736	26 624	>22
60.....	10 101	0.6750	6 818	13 191	0.7042	7 113	13 904	0.7342	7 416	14 657	0.7719	7 797	15 623	16 563	>22
65.....	6 618	0.5406	3 578	6 373	0.5704	3 775	6 791	0.6016	3 981	7 241	0.6426	4 253	7 826	-	-
70.....	11 785	0.2372	2 795	2 795	0.2559	3 016	3 016	0.2766	3 260	3 260	0.3032	3 573	3 573	-	-



tions for a range of mortality levels. For each mortality level  $\nu$ , the projected population of each age group from  $x$  to  $x + 5$  in 1960,  ${}_5NP_{x+10}^\nu$ , is obtained by applying the survivorship ratio,  ${}_{10}S_{x,x+4}^\nu$ , to the initial population,  ${}_5N1_x$ . Thus, for age group 15-19 and mortality level 16, the initial population is 53,468, and the survivorship probability is 0.9623, so

$${}_5NP_{25}^{16} = ({}_5N1_{15})({}_{10}S_{15,19}^{16}) = (53,468)(0.9623) = 51,452,$$

whereas for the same age group but at mortality level 22,

$${}_5NP_{25}^{22} = (53,468)(0.9919) = 53,035.$$

The 1960 population of each five-year age group, from 5-9 upward, has been projected forward using survivorship probabilities from each mortality level (note that in practice it is sufficient to work with steps of two levels, that is, to use levels 16, 18, 20 and 22, and not to repeat the calculations for the intermediate levels), the results being shown in columns (4), (7), (10) and (13) of table 180. Each projected population can then be cumulated from the oldest age group towards the youngest to find the population over ages 55, 50, 45 and thus down to the population over 15. The results for the mortality levels being used are shown in columns (5), (8), (11) and (14) of table 180. The reported populations over each of these ages can then be compared with the projected populations, and linear interpolation can be used to estimate the mortality level implied by each. Thus, for instance, the reported population over age 20 in 1970 is 324,738; the projected population over age 20 for mortality level 16 (column (5) of table 180) is 320,549, whereas for mortality level 18 it is 325,721 (column (8) of table 180). Therefore, the mortality level implied by the observed population is obtained as

$$\nu(20+) = 16.0 + (2.0)(324,738 - 320,549) / (325,721 - 320,549) = 17.62$$

where 16 is the level associated with the smaller projected population; and the interpolation factor has to be multiplied by two because the levels used, 16 and 18, are two units apart. Column (15) of table 180 shows the cumulated populations observed in 1970, and column (16) shows the mortality levels yielding projected populations over each age consistent with the observed.

The mortality levels shown in column (16) are certainly less variable than those obtained for individual age groups in table 174. However, after four rather consistent estimates of levels in the range from 16.8 to 17.8 associated with the populations over 15, 20, 25 and 30 in 1970, the estimates show a steady tendency to rise as the lower age boundary increases. One possible cause of this outcome is that the West mortality pattern is not a good representation of adult mortality in Panama, but it seems more likely that systematic age-reporting errors may be distorting the second age distribution (and probably the first as well, though the method provides information only about relative differences). In the circumstances, the best estimate of mortality level that one

can obtain from these data is the average of the first four values, 17.4, rather than the median of all the values, which is more likely to be affected by the apparent tendency towards age exaggeration. Note that this value of 17.4 is substantially higher than the final estimate based on individual age groups, 16.1. The level based on cumulated data is probably the better of the two estimates, although it should be remembered that the cumulation procedure is really only applied to the second age distribution, not to the first, so the results are still dependent upon the age detail of the initial age distribution. The greater consistency of the results obtained from cumulated data should not be interpreted as necessarily indicating greater accuracy.

### C. INTERCENSAL SURVIVAL WITH ADDITIONAL INFORMATION ON THE AGE PATTERN OF MORTALITY

#### 1. Basis of method and its rationale

The two most serious problems affecting intercensal-survival techniques are age-misreporting and different levels of census coverage. A procedure often used to reduce the effects of age-misreporting is cumulation; the population over age  $x$  is affected only by erroneous transfers of people across the boundary  $x$  and not by errors over or under  $x$ . The second problem, that of coverage changes between the first and second censuses, can play havoc with mortality estimates derived from intercensal survival, since the coverage change will appear either as excess deaths (when the second census is the less complete) or as a deficiency of deaths (when the second census is the more complete). However, a change in coverage that is more or less constant by age will inflate or deflate intercensal deaths by amounts proportional to the population at each age, rather than by amounts proportional to the number of deaths at each age. A change in coverage will therefore have much more effect on deaths at younger ages, where there are in reality few deaths but large numbers of people, than at older ages, where there are many more deaths but small numbers of people.

A technique that is simple to understand but rather laborious to apply makes it possible to use what is essentially an intercensal-survival procedure while employing cumulated data and also making allowance for differential census coverage.<sup>3</sup> The method is based on the simple idea that, in a closed population, the number of people in a particular age group at a first census should be equal to the number of survivors of the same cohort at the second census plus the deaths of cohort members during the intercensal period. It can be simply shown that if the coverages of the first and second censuses and of intercensal deaths are invariant with age and are denoted by  $C_1$ ,  $C_2$  and  $k$ , respectively, then

$${}_nN1_x / {}_nN2_{x+t} = C_1 / C_2 + (C_1 / k) {}_nD_x / {}_nN2_{x+t} \quad (C.1)$$

<sup>3</sup> S. Preston and K. Hill, "Estimating the completeness of death registration". *Population Studies*, vol. XXXIV, No. 2 (July 1980), pp. 349-366.

where  $n$  is the cohort width in years;  $t$  is the length of the intercensal period;  ${}_nN1_x$  and  ${}_nN2_{x+t}$  are the enumerated cohort populations at the first and second censuses, respectively; and  ${}_nD_x$  is the registered number of intercensal deaths to the cohort aged from  $x$  to  $x+n$  at the first census. Equation (C.1) defines a straight line with slope  $C_1/k$  and intercept  $C_1/C_2$ ; that is, its slope is the completeness of the first census in relation to the completeness of death registration and the intercept is the completeness of the first census in relation to that of the second. The fitting of a straight line to points  $[{}_nN1_x/{}_nN2_{x+t}, {}_nD_x/{}_nN2_{x+t}]$  for different cohorts should therefore provide estimates of the relative completeness of the two censuses and of the completeness of death registration in relation to that of the first census.

Equation (C.1) is valid for any cohort, be it an initial five-year age group, the initial population over some age  $x$ , or even the population between two ages,  $x$  and  $y$ , at the first census. All that is required is that the range of the ratios  ${}_nN1_x/{}_nN2_{x+t}$  and  ${}_nD_x/{}_nN2_{x+t}$  be wide enough for the robust estimation of the parameters (mainly the slope) of the straight line to be possible. Cumulation can therefore be used to reduce the effects of some age errors, though the procedure remains sensitive to systematic age exaggeration on the part of the elderly.

## 2. Data required

The data listed below are required for this method:

(a) Two census enumerations with populations classified by age (and sex) for two points in time not more than 15 years apart. (It may be necessary that the age classification be by single year for at least one census if the intercensal interval is not a multiple of five years);

(b) Information on deaths by age (and sex) for the intercensal period; registered deaths for each intercensal year can be used but the calculations are lengthy, and deaths for every fifth year are adequate. If no information on deaths is available, a model life table can be used to supply the deficiency.

## 3. Computational procedure

The steps of the computational procedure are described below.

*Step 1: adjustment for net intercensal migration and territorial coverage.* See step 1 in subsection B.2(b). Note, however, that before applying this method, intercensal deaths should also be adjusted for migration and coverage changes, though if the age pattern of deaths is not much affected, the adjustment is not crucial to the final mortality estimates.

*Step 2: grouping of data from the two censuses by cohort.* See step 2 in subsection B.2(b).

*Step 3: adjustment of intercensal interval that is not an exact number of years.* See step 3 in subsection B.2(b).

*Step 4(a): cumulation of cohort deaths from registration data.* Registered deaths are normally tabulated by calendar year, five-year age group and sex. Given that the two censuses being used probably do not have reference dates at the beginning of a year and that a cohort

will be continually moving across standard five-year age groups as it moves through the intercensal period, the task of cumulating intercensal cohort deaths is tedious and imprecise. Since the value of the information on intercensal deaths lies in their age pattern, not in their precise overall level, a degree of simplification is in order.

If the first census was held in year  $a$  and initial cohorts are defined by standard five-year age group, deaths to the cohort aged from  $x$  to  $x+4$  in year  $a$  over the first five years of the intercensal period (that is, between  $a$  and  $a+5$ ) can be approximated by summing the deaths in year  $a$  to persons aged from  $x$  to  $x+4$  and the deaths in year  $a+5$  to persons aged from  $x+5$  to  $x+9$ , and multiplying the sum by 2.5. Thus, if  $D_{x,x+4}^j$  denotes, in general, the number of deaths to the cohort aged from  $x$  to  $x+4$  at the beginning of the period (that is, in year  $a$ ) and recorded during year  $j$ , and  ${}_5D_x^j$  denotes the number of deaths to persons aged from  $x$  to  $x+4$  in year  $j$ , then

$$\sum_{j=a}^{a+5} D_{x,x+4}^j = 2.5 [{}_5D_x^a + {}_5D_x^{a+5}]. \quad (C.2)$$

Similar approximations can be applied for a second five-year period between  $a+5$  and  $a+10$ :

$$\sum_{j=a+5}^{a+10} D_{x,x+4}^j = 2.5 [{}_5D_x^{a+5} + {}_5D_x^{a+10}]. \quad (C.3)$$

Cohort deaths for intercensal periods that are multiples of five can therefore be approximated rather simply from registered deaths for calendar years five years apart.

The case of an intercensal interval that is not a multiple of five years is slightly more complicated, but adequate approximations can be arrived at by suitable weighting of registered deaths. If the interval is between five and 10 years, cohort deaths for the first five years can be approximated as described above using the deaths registered in years  $a$  and  $a+5$ . Cohort deaths over the period from  $a+5$  to  $a+t$ , where  $t$  is the length of the intercensal interval, are approximated by averaging the number of deaths recorded in years  $a+5$  and  $a+t$  belonging to the appropriate age groups and then weighting the averages according to the number of years between  $a+5$  and  $a+t$ . Thus, letting

$${}_5\bar{D}_{x+5} = 0.5 [{}_5D_{x+5}^a + {}_5D_{x+5}^{a+t}], \quad (C.4)$$

$$\sum_{j=a+5}^{a+t} D_{x,x+4}^j = w(t-5) {}_5\bar{D}_{x+5} + (t-5-w(t-5)) {}_5\bar{D}_{x+10}. \quad (C.5)$$

Values of  $w(t-5)$  are shown in table 181.

If the intercensal interval is between 10 and 15 years, cohort deaths for the first 10 years can be obtained from equations (C.2) and (C.3). Cohort deaths for the extra period can then be obtained by using equations (C.4)

and (C.5) with  $a + 5$  substituted by  $a + 10$ ,  $t - 5$  by  $t - 10$ ; and  $x + 5$  and  $x + 10$  replaced by  $x + 10$  and  $x + 15$ , respectively. The necessary weights can still be calculated from table 181 using as point of entry  $(t - 10)$  years.

TABLE 181. WEIGHTING FACTORS FOR APPROXIMATION OF COHORT DEATHS FOR INTERVALS THAT ARE NOT MULTIPLES OF FIVE

Interval (years) $t-5$ (1)	Weighting factors	
	$w(t-5)$ (2)	$(t-5-w(t-5))$ (3)
1.....	0.9	0.1
2.....	1.6	0.4
3.....	2.1	0.9
4.....	2.4	1.6

*Step 4(b): approximation of cohort deaths from a model life table.* If no information is available about the age distribution of deaths over the intercensal period, model life tables can be used to fill the gap. The exact level of mortality is not critical, since the method will estimate the "completeness" of death registration; but because the age pattern of mortality is important, a model that adequately represents the pattern of mortality experienced by the population being studied and that roughly approximates its level should be chosen.

What is required as input is an estimate of the number of deaths occurring to an initial cohort during the intercensal period. The life-table  ${}_5L_x$  function can conveniently be used to synthesize the required number. If the intercensal period is five years in length, the number of deaths occurring to the cohort aged from  $x$  to  $x + 4$  at the time of the first census is given by  ${}_5L_x - {}_5L_{x+5}$ ; the proportion of the initial cohort dying during the period is given by  $({}_5L_x - {}_5L_{x+5})/{}_5L_x$ . In general, for any intercensal interval  $t$ , the number of deaths occurring during the intercensal interval to the cohort aged from  $x$  to  $x + 4$  at the first census can be estimated as

$$\sum_{j=a}^{a+t} D_{x,x+4}^j = ({}_5N1_x)({}_5L_x - {}_5L_{x+t})/{}_5L_x \quad (C.6)$$

where  ${}_5N1_x$  is the population aged from  $x$  to  $x + 4$  at the time of the first census (year  $a$ ). If  $t$  is not a multiple of five,  ${}_5L_{x+t}$  values cannot be read directly from the Coale-Demeny model life tables, though they can be estimated from the tabulated values of  $l(x)$  for adjacent ages that are multiples of five, as described in step 5 of subsection B.2(b), using the coefficients and equation given in table 172.

*Step 5(a): calculation and plotting of population and death ratios for five-year cohorts.* Working with five-year age groups, one can calculate the ratios of the cohort aged from  $x$  to  $x + 4$  at the first census to its survivors  $t$  years later,  ${}_5N1_x / {}_5N2_{x+t}$ , and of the intercensal deaths of its members to its survivors  $t$  years later,

$$\sum_{j=a}^{a+t} D_{x,x+4}^j / {}_5N2_{x+t}$$

These points can then be plotted. Typically, they will be very erratic, often to the extent that no linear trend can be plausibly associated with them. If an examination of their plot suggests that they do represent a line, its parameters may be estimated by using group means (see chapter V, subsection C.4). The slope can then serve as an adjustment factor for the recorded number of deaths, adjustment that will make them consistent with the coverage of the first census, while the intercept is an estimate of the coverage of the first census in relation to that of the second.

*Step 5(b): calculation and plotting of population and death ratios for open-ended cohorts.* Cohorts can also be defined in terms of open-ended age intervals, that is, as all those of aged  $x$  and over at the time of the first census; and the ratios  $N1(x+)/N2((x+t)+)$  and  $\sum_{j=a}^{a+t} D_{x,\omega}^j / N2((x+t)+)$  can be calculated for values of  $x$  of 5, 10, 15 and so on. Many of the irregularities observed in the five-year ratios will be smoothed out by this cumulation, and a group mean procedure (see chapter V, subsection C.4) can be used to fit a straight line to these points.

*Step 5(c): calculation and plotting of population and death ratios for truncated cohorts.* It is often useful to exploit the advantages of cumulation without using information for the elderly, whose age-misreporting may be substantial. Therefore, initial cohorts aged between  $x$  and 60, or  $x$  and 65, can be used for values of  $x$  that are multiples of five. The use of truncated cumulation will have a substantial smoothing effect, but the slope of the resulting line may be highly sensitive to the upper age-limit chosen.

*Step 6: interpretation of results.* Often, selective migration, age exaggeration and other reporting problems may distort the slope of the fitted line. In such cases, this method may not be very useful for mortality estimation purposes. The estimate of the intercept, on the other hand, appears to be more robust, so that the main value of this method lies in the assessment it provides of the relative coverage of successive censuses.

#### 4. A detailed example

The method is applied to the case of Panama, 1960-1970, since some of the necessary calculations have already been made. The steps of the procedure are given below.

*Step 1: adjustment for net intercensal migration and territorial coverage.* As no basis exists for making the necessary adjustments, the basic population data used are those presented in table 173. Deaths do not require any adjustment either.

*Step 2: grouping of data from the two censuses by cohort.* This step has been fully described in subsection B.2(c); standard five-year age groups in both 1960 and 1970 are used to define cohorts, since the intercensal interval is approximately 10 years.

*Step 3: adjustment for length of the intercensal interval.* As described in subsection B.2(c), the 1960 female popu-

TABLE 182. FEMALE POPULATION, 1960 AND 1970; AND REGISTERED DEATHS OF FEMALES, 1960, 1965 AND 1970; BY AGE GROUP, PANAMA

Age group (1)	Female population		Registered deaths		
	1960 <sup>a</sup> (2)	1970 (3)	1960 (4)	1965 (5)	1970 (6)
0-4	88 477	114 017	1 670	1 497	1 608
5-9	75 242	106 944	138	122	167
10-14	62 509	85 253	63	64	71
15-19	53 468	73 381	76	84	95
20-24	44 826	63 010	106	102	99
25-29	37 149	50 924	103	79	104
30-34	31 609	40 885	87	91	94
35-39	28 216	36 115	98	112	116
40-44	23 550	29 409	99	111	110
45-49	20 253	25 360	123	119	137
50-54	14 801	21 775	111	141	161
55-59	11 787	17 632	132	115	186
60-64	10 101	13 004	147	188	233
65-69	} 18 403	10 061	152	174	268
70-74		{ 6 690	184	239	270
75+		{ 9 873	513	742	913

<sup>a</sup> Adjusted to approximate an intercensal interval of 10 years.

lation was moved back from the actual census date, 11 December 1960, to a date exactly 10 years before the 1970 census, 10 May 1960. Results are shown in column (2) of table 182.

*Step 4(a): cumulation of cohort deaths from registration data.* Registered deaths by age and sex are available for Panama for 1960, 1965 and 1970; and these data can be used to estimate intercensal deaths for each cohort. The numbers of female deaths for each year and age group are shown in table 182.

The approximate procedure for estimating cohort deaths is too crude to give reasonable results for the cohort aged 0-4 in 1960, so the cohort aged 5-9 is the starting-point. Deaths occurring over the first five years, from mid-1960 to mid-1965, to this cohort are estimated as

$$\sum_{j=1960}^{1965} D_{5,9}^j = 2.5 [{}_5D_5^{1960} + {}_5D_{10}^{1965}] = 2.5 (138 + 64) = 505.$$

Deaths for the cohort aged 5-9 in 1960 over the second five years, from mid-1965 to mid-1970, are estimated as

$$\sum_{j=1965}^{1970} D_{5,9}^j = 2.5 [{}_5D_{10}^{1965} + {}_5D_{15}^{1970}] =$$

$$2.5(64 + 95) = 397.5.$$

Cohort deaths for the 10-year period are then obtained by summing the deaths during the two five-year periods:

$$\sum_{j=1960}^{1970} D_{5,9}^j = \sum_{j=1960}^{1965} D_{5,9}^j + \sum_{j=1965}^{1970} D_{5,9}^j =$$

$$505 + 397.5 = 902.5.$$

The results for each cohort are shown in table 183. The only age group that requires special treatment is the open-ended interval 65 and over. For the period 1960-1965, all the annual deaths of persons over 70 belong to

this cohort, as do some of the deaths of persons aged 65-69. The total number of deaths of persons over 70 during the period 1960-1965 can be estimated by summing the deaths over age 70 in 1960 and 1965 and multiplying the sum by 2.5:

$$\begin{aligned} \sum_{j=1960}^{1965} \omega_{-70} D_{70}^j &= 2.5(\omega_{-70} D_{70}^{1960} + \omega_{-70} D_{70}^{1965}) \\ &= 2.5(184 + 513 + 239 + 742) \\ &= 2.5(1,678) = 4,195 \end{aligned}$$

where, as usual,  $\omega_{-70} D_{70}^j$  denotes the number of deaths occurring during year  $j$  to persons aged 70 and over.

For age group 65-69, the average number of deaths per annum between 1960 and 1965 is estimated as  $0.5(152 + 174)$ , or 163; and since the cohort aged 65-69 in 1960 averaged 2.5 years of exposure to the risk of dying during the period 1960-1965, the deaths of persons aged 65-69 belonging to that cohort are estimated as  $2.5(163) = 407.5$  during that period. Hence, the total

TABLE 183. ESTIMATED COHORT DEATHS, FEMALE POPULATION, PANAMA, 1960-1970

Cohort age group 1960 (1)	Number of cohort deaths		
	1960-1965 (2)	1965-1970 (3)	1960-1970 (4)
5-9	505.0	397.5	902.5
10-14	367.5	457.5	825.0
15-19	445.0	515.0	960.0
20-24	462.5	432.5	895.0
25-29	485.0	517.5	999.5
30-34	497.5	555.0	1 052.5
35-39	522.5	620.0	1 142.5
40-44	545.0	700.0	1 245.0
45-49	660.0	817.5	1 477.5
50-54	565.0	870.0	1 435.0
55-59	800.0	1 140.0	1 940.0
60-64	802.5	1 110.0	1 912.5
65+	4 602.5	4 773.8	9 376.3

number of deaths during the period 1960-1965 for the cohort aged 65 and over in 1960 is

$$\sum_{j=1960}^{1965} {}_5D_{65}^j + \sum_{j=1960}^{1965} {}_{\omega-70}D_{70}^j = 4,195 + 407.5 = 4,602.5.$$

Between 1965 and 1970, all the deaths at age 75 and over belong to the initial 65+ cohort, as do a proportion of the deaths of persons aged 70-74. Deaths at age 75 and over are estimated as

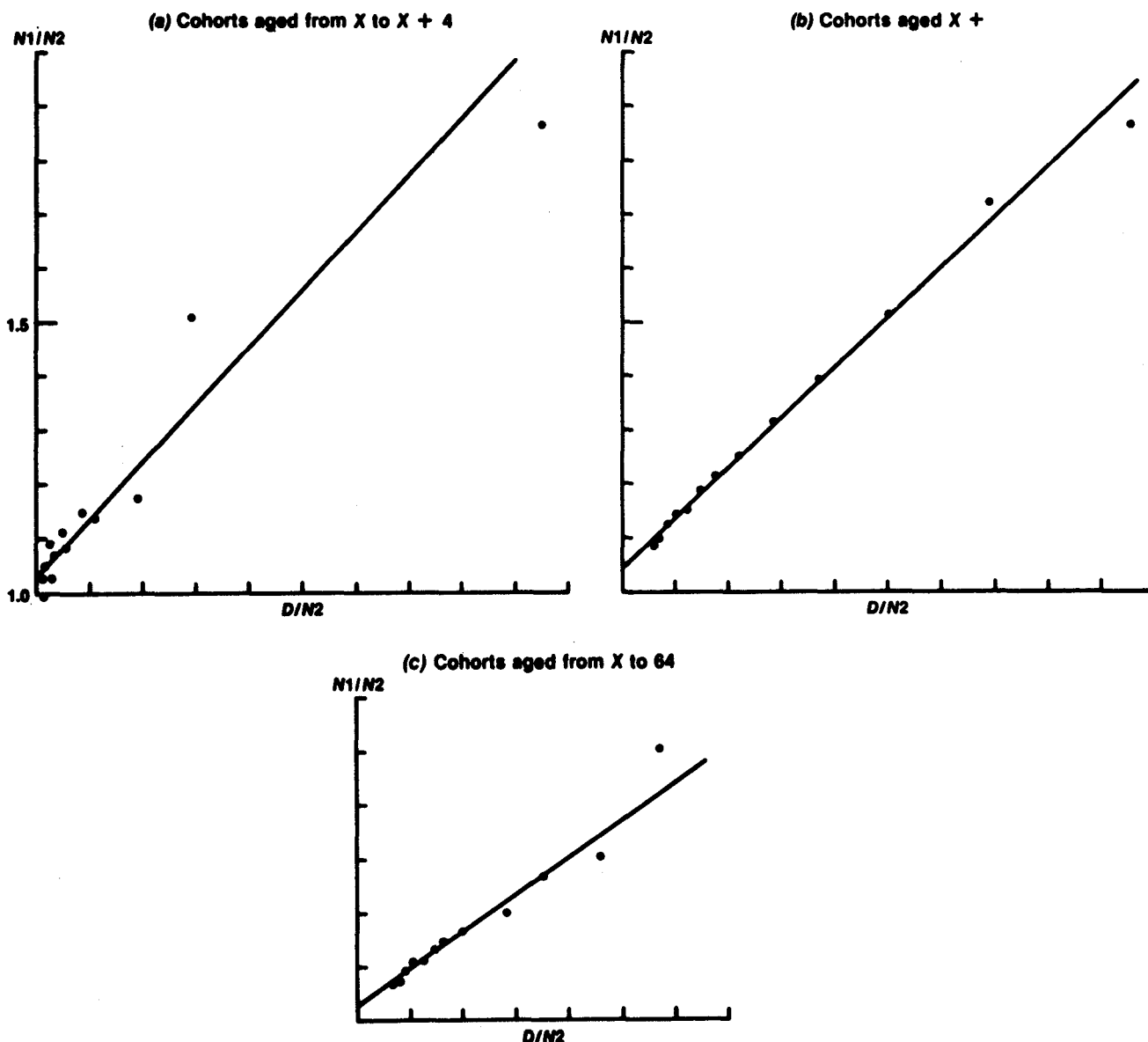
$$\begin{aligned} \sum_{j=1965}^{1970} {}_{\omega-75}D_{75}^j &= 2.5({}_{\omega-75}D_{75}^{1965} + {}_{\omega-75}D_{75}^{1970}) \\ &= 2.5(742 + 913) = 4,137.5. \end{aligned}$$

Deaths occurring to the cohort during the period 1965-1970 at ages 70-74 can be estimated from the average annual number of deaths at these ages,  $0.5(239+270)$ , and the average exposure to risk, 2.5 years, giving  $(2.5)(0.5)(239+270) = 636.25$ . Hence, the deaths occurring during the intercensal period to the cohort aged 65 and over at the beginning of it (1960) are

$$\begin{aligned} \sum_{j=1960}^{1970} D_{65,\omega}^j &= 4,137.5 + 636.25 + 4,602.5 \\ &= 9,376.25 \end{aligned}$$

*Step 5(a): calculation and plotting of population and death ratios for five-year cohorts.* Ratios of the initial to the final cohort size,  $N1/N2$ , and of cohort deaths to final cohort size,  $D/N2$ , are calculated for each cohort.

Figure 23. Plots of cohort population ratios,  $N1/N2$ , against ratios of cohort deaths over population,  $D/N2$ , for various types of cohorts



Thus, for the cohort aged 5-9 in 1960,

$${}_5N_{15}/{}_5N_{215} = 75,242/73,381 = 1.0254$$

$$\sum_{j=1960}^{1970} D_{5,9}^j / {}_5N_{215} = 902.5/73,381 = 0.0123.$$

Values for all cohorts are shown in columns (2) and (3) of table 184 and are plotted in panel (a) of figure 23. The points show a fair amount of variability; and the straight line, fitted by group means, is heavily influenced by the last three points and does not approximate the others very well. The intercept, which estimates the coverage of the first census in relation to that of the second, is 1.0278, suggesting that the 1960 census was some 3 per

cent more complete than the 1970 census. The slope, which estimates the coverage of the first census in relation to the completeness of death registration, is 1.0623, suggesting that death registration was some 6 per cent less complete than the enumeration of the 1960 population. These estimates are by no means unreasonable, but the estimate of the slope is sensitive to the fitting procedure used; for instance, if the line is fitted by using least squares (see chapter V, subsection C.4) the value of the intercept becomes 1.046, not very different from that obtained by group means; but the slope is 0.914, a value totally different from that associated with the line fitted by using group means. Furthermore, the former value is rather implausible, because it implies that the coverage of deaths is more complete than that of the population.

TABLE 184. COHORT POPULATION RATIOS,  $N1/N2$ , AND RATIOS OF COHORT DEATHS OVER POPULATION,  $D/N2$ , FOR DIFFERENT TYPES OF COHORTS, PANAMA, 1960-1970

Age $x$ in 1960 (1)	Ratios for cohorts aged from $x$ to $x+4$		Ratios for cohorts aged $x$ and over		Ratios for cohorts aged from $x$ to 64	
	${}_5N_x/{}_5N_{x+10}$ (2)	$\sum D_{x,x+4}^j / {}_5N_{x+10}$ (3)	$N1(x+)/N2((x+10)+)$ (4)	$\sum D_{x,\omega}^j / N2((x+10)+)$ (5)	$N1(x,64)/N2(x+10,74)$ (6)	$\sum D_{x,64}^j / N2(x+10,74)$ (7)
5	1.0254	0.01230	1.0849	0.06069	1.0651	0.03525
10	0.9920	0.01309	1.0983	0.07163	1.0743	0.04060
15	1.0500	0.01885	1.1239	0.08572	1.0949	0.04749
20	1.0964	0.02189	1.1418	0.10188	1.1063	0.05474
25	1.0286	0.02768	1.1527	0.12112	1.1088	0.06313
30	1.0748	0.03579	1.1862	0.14634	1.1322	0.07347
35	1.1126	0.04505	1.2176	0.17749	1.1501	0.08519
40	1.0815	0.05718	1.2513	0.21998	1.1638	0.09991
45	1.1487	0.08380	1.3158	0.28189	1.2016	0.14276
50	1.1382	0.11035	1.3902	0.37004	1.2330	0.17770
55	1.1716	0.19282	1.5133	0.49687	1.3067	0.22999
60	1.5099	0.28587	1.7209	0.68156	1.5099	0.28587
65	1.8640	0.94969	1.8640	0.94969	-	-
<i>Parameters of straight lines fitted by group means to all points:</i>						
Intercept.....	1.0278			1.0409		1.0239
Slope.....	1.0623			0.9399		1.3916

*Step 5(b): calculation and plotting of population and death ratios for open-ended cohorts.* The most convenient way to cumulate is to begin with the higher ages and work towards younger ones, beginning with the cohort aged 65 and over in 1960. In this case, the values of  $N1$ ,  $N2$  and  $D$  are exactly the same as those obtained in step 4(a):

$$N1(65+)/N2(75+) = 18,403/9,873 = 1.8640$$

$$\sum_{j=1960}^{1970} D_{65,\omega}^j / N2(75+) = 9,376.25/9,873 = 0.94969.$$

For the initial cohort aged 60 and over,

$$N1(60+)/N2(70+) = (18,403 + 10,101)/(9,873 + 6,690) = 1.7209$$

$$\sum_{j=1960}^{1970} D_{60,\omega}^j / N2(70+) = (9,376.25 + 1,912.5) / (9,873 + 6,690) = 0.68156.$$

The calculations continue in the same way, adding a

new age group to the previous sums at each step, until the youngest cohort, aged 5-9 in 1960, has been added in. Columns (4) and (5) of table 184 show the ratios obtained, which are plotted in panel (b) of figure 23.

The linearity of the points has been greatly improved by the cumulation, and the straight line fits the observations quite closely. However, the parameters of the line—an intercept of 1.0409 and a slope of 0.9399—imply a similar enumeration coverage differential between the 1960 and 1970 censuses as that obtained in step 5(a), but a quite different coverage of death registration, which is now estimated to overrecord deaths, in relation to the 1960 census, by some 6 per cent.

*Step 5(c): calculation and plotting of population and death ratios for truncated cohorts.* Cumulations similar to those carried out in step 5(b) can also be made so that the age groups considered are not entirely open-ended, but rather exclude the last one or two age groups. Such truncated cumulation can be useful if there is reason to suppose that special types of errors affect the oldest age groups. Estimates of completeness of death registration will then refer to registered adult deaths excluding both those at young ages and those at old ages.

The cumulations are carried out as before. However, they begin not with the last, open-ended cohort, but with the oldest cohort that one intends to include in the calculations. In this case, the cohort aged 60-64 in 1960 was chosen as the upper limit. For this cohort, the ratios required are the same as those calculated in step 5(a), in which ratios were calculated for individual five-year age cohorts. Thus,

$$N1(60, 64)/N2(70, 74) = 10,101/6,690 = 1.5099$$

$$\sum_{j=1960}^{1970} D_{60,64}^j / N2(70, 74) = 1,912.5/6,690 = 0.28587.$$

For the cohort aged 55-64 in 1960,

$$N1(55, 64)/N2(65, 74) = (10,101 + 11,787)/$$

$$(6,690 + 10,061) = 1.3067$$

$$\sum_{j=1960}^{1970} D_{55,64}^j / N2(65, 74) = (1,912.5 +$$

$$1,940.0)/(6,690 + 10,061) = 0.22999.$$

The cumulations continue downward with age, adding the next younger age group each time, until age group 5-9, the youngest to be used, has been included. The resulting ratios are shown in columns (6) and (7) of table 184 and are plotted in panel (c) of figure 23.

The plot shows that the degree of linearity of the points is somewhere between that in the fully cumulated case and that in which no cumulation was used. The intercept of the fitted line is similar to those obtained earlier; but the slope is once more quite different, indicating that registered deaths in the approximate age range from 10 to 70 are only about 72 per cent complete with respect to the 1960 census coverage.

#### D. ESTIMATION OF A POST-CHILDHOOD LIFE TABLE FROM AN AGE DISTRIBUTION AND INTERCENSAL GROWTH RATES

##### 1. Basis of method and its rationale

Traditional intercensal survival techniques are greatly complicated by intercensal intervals that are not exact numbers of years in length or are not multiples of five. Furthermore, the application of the method described in subsection B.4, which uses cumulation to reduce the impact of age-reporting errors, is very time-consuming since it involves the projection of an initial population using different mortality levels.

Preston and Bennett<sup>4</sup> propose a different method to estimate adult mortality during the intercensal period from the age distributions produced by two consecutive censuses. The application of the method proposed is simple whatever the length of the intercensal period, and

<sup>4</sup> Samuel H. Preston and Neil G. Bennett, "A census-based method for estimating adult mortality", *Population Studies*, vol. 37, No. 1 (March 1983), pp. 91-104.

it is not very sensitive to certain types of age-misreporting, particularly heaping. It is also innovative, because it uses the two census age distributions to estimate age-specific growth rates (rather than cohort survivorship probabilities) and then uses these growth rates to transform the observed population age structure into the equivalent of a stationary-population (life-table)  ${}_5L_x$  function.

Bennett and Horiuchi<sup>5</sup> show that in any closed population, at a particular time  $t$ , the number of persons aged  $y$ ,  $N(y)$ , is equal to the number of persons aged  $x$ ,  $N(x)$ , multiplied by the probability of surviving from age  $x$  to age  $y$ ,  $l(y)/l(x)$  measured at time  $t$ , and by an exponential factor involving the integral of the population growth rates also at time  $t$  between ages  $x$  and  $y$ . Thus,

$$N(y) = N(x) \{l(y)/l(x)\} \exp\left(-\int_x^y r(u) du\right), \quad (D.1)$$

which can be regarded as being equivalent to a stable population relationship, except for the replacement of the exponential of the stable growth rate times the number of years between  $x$  and  $y$ ,  $\exp[-(y-x)r]$ , by the exponential of the integral of the variable growth

rates between  $x$  and  $y$ ,  $\exp\left[-\int_x^y r(u) du\right]$ .

If  $N(x)$ ,  $N(y)$  and the set of  $r(u)$  values for  $u$  between  $x$  and  $y$  are all known, then the period survivorship probability,  $l(y)/l(x)$ , can be estimated from equation (D.1). However, in order to introduce a certain amount of smoothing, Preston and Bennett propose the estimation of the expectation of life at each age  $x$ , using extensive cumulation both of the reported population and of the observed age-specific growth rates.

In discrete terms, using five-year age groups, the basic equation proposed by Preston and Bennett is

$$e_x = \frac{\sum_{y=x}^{\omega} {}_5N_y \exp\left[5.0 \sum_{u=x}^{y-5} {}_5r_u + 2.5{}_5r_y\right]}{N(x)} \quad (D.2)$$

where  $N(x)$ , the number of people aged  $x$ , is estimated as

$$N(x) = \frac{{}_5N_{x-5} \exp[-2.5{}_5r_{x-5}] + {}_5N_x \exp[2.5{}_5r_x]}{10} \quad (D.3)$$

The advantages of this method are: (a) its application is relatively simple even in cases where the intercensal period does not have a convenient length; (b) it makes no assumptions concerning stability; (c) it introduces an element of cumulation, thus limiting the effects of age errors, and (d) the use of growth rates eliminates the effects of age errors for which the pattern is the same at

<sup>5</sup> Neil G. Bennett and Shiro Horiuchi, "Estimating the completeness of death registration in a closed population", *Population Index*, vol. 47, No. 2 (Summer 1981), pp. 207-221.

both censuses. The estimates yielded by this procedure are probably as reliable as those obtained from any of the intercensal techniques available, and its simplicity of application makes it extremely attractive.

### 2. Data required

The data required for this method are listed below:

(a) Two census enumerations with populations classified by the same age groups (and sex), separated by an intercensal interval which should not exceed 20 years;

(b) Sufficient information to adjust one census or the other for net intercensal migration and territorial coverage, if necessary.

### 3. Computational procedure

The steps of the computational procedure are described below.

*Step 1: adjustment for net intercensal migration and territorial coverage.* See step 1 in subsection B.2(b).

*Step 2: calculation of age-specific intercensal growth rates.* The rate of growth of the population in each five-year age group from the first to the second census is calculated as

$${}_5r_x = [\ln({}_5N2_x) - \ln({}_5N1_x)]/t \quad (D.4)$$

where  ${}_5r_x$  denotes the intercensal growth rate of the population of the age group from  $x$  to  $x + 4$ ;  ${}_5N2_x$  is the population aged from  $x$  to  $x + 4$  at the second census;  ${}_5N1_x$  is the population aged from  $x$  to  $x + 4$  at the first census; and  $t$  is the length of the intercensal interval in years (with a decimal portion if necessary). Both age distributions must share the same open interval,  $A +$ . The value of  $A$  should be set as high as the two age distributions permit, since age exaggeration is a less severe problem with this method than with the death distribution methods described in chapter V.

*Step 3: calculation of average intercensal age distribution.* Equation (D.2) requires the use of an average intercensal age distribution,  ${}_5N_x$ . An adequate approximation to this age distribution can be obtained by simply averaging the initial and final populations of each age group. Thus,

$${}_5N_x = 0.5({}_5N1_x + {}_5N2_x). \quad (D.5)$$

*Step 4: cumulation of age-specific growth rates from age 5 upward.* The calculation of the growth rate "inflater" appearing in equation (D.2) requires the summation of the age-specific growth rates,  ${}_5r_x$ , calculated in step 2. It is normally convenient to begin the cumulation process with age 5 and continue upward.

The only difficulty involved in the cumulation is the treatment of the inflation factor associated with the open age interval,  $A +$ . Although the relative importance of the open interval is much less in this case than in the conceptually similar death distribution techniques discussed in chapter V, because in the calculation of expect-

tations of life the open interval is always present, it is sound practice to minimize the influence of biases due entirely to the weight it may be assigned in the inflation factor. For this reason, a special procedure is suggested to deal with the open interval.

If  $R(x)$  is used to denote the inflation factor for the age group from  $x$  to  $x + 4$ , then, according to equation (D.2),

$$R(x) = 2.5 {}_5r_x + 5.0 \sum_{y=5}^{x-5} {}_5r_y \quad (D.6)$$

for  $x = 10, 15, \dots, A - 5$ . In the case of  $x = 5$ , (D.6) becomes

$$R(5) = 2.5 {}_5r_5 \quad (D.7)$$

$$\text{and for } x = A, R(A) = \rho(A) + 5.0 \sum_{y=5}^{A-5} {}_5r_y \quad (D.8)$$

where  $\rho(A)$  is calculated by using an equation derived from simulated stable populations and whose form is

$$\rho(A) = a(A) + b(A)r(10+) + c(A) \ln(N(45+)/N(10+)) \quad (D.9)$$

where  $r(10+)$  is the growth rate of the population over age 10, that is,

$$r(10+) = \ln[N2(10+)/N1(10+)]/t; \quad (D.10)$$

$N(10+)$  and  $N(45+)$  are the mid-period populations aged 10 and over, and 45 and over, respectively, and estimated as

$$N(10+) = 0.5(N1(10+) + N2(10+))$$

and

$$N(45+) = 0.5(N1(45+) + N2(45+)); \quad (D.11)$$

and  $a(A)$ ,  $b(A)$  and  $c(A)$  are constant coefficients depending upon the actual value of  $A$ . Their values are shown in table 185.

TABLE 185. COEFFICIENTS FOR ESTIMATION OF THE EQUIVALENT GROWTH RATE OVER AGE  $A$ ,  $\rho(A)$ , FROM THE GROWTH RATE OVER AGE 10 AND THE RATIO OF THE POPULATION OVER AGE 45 TO THE POPULATION OVER AGE 10

Age $A$ (1)	Coefficients		
	$a(A)$ (2)	$b(A)$ (3)	$c(A)$ (4)
45.....	0.229	20.43	0.258
50.....	0.205	18.28	0.235
55.....	0.179	16.02	0.207
60.....	0.150	13.66	0.176
65.....	0.119	11.22	0.141
70.....	0.086	8.77	0.102
75.....	0.053	6.40	0.063
80.....	0.025	4.30	0.029
85.....	0.006	2.68	0.006

Estimation equation:

$$\rho(A) = a(A) + b(A)r(10+) + c(A) \ln(N(45+)/N(10+))$$



*Step 5: reduction of age distribution to a stationary form.* The average intercensal age distribution obtained in step 3 is converted into a stationary population by multiplying each value  ${}_5N_x$  by the exponential of  $R(x)$ . The results can be regarded as "pseudo" ${}_5L_x^*$  values, analogous to the values of  ${}_5L_x$  in the usual life table. However, the life-table radix corresponding to the pseudo  ${}_5L_x^*$  values is not known. Therefore, in general, the pseudo  ${}_5L_x^*$  values cannot be manipulated as can normal  ${}_5L_x$  values.

Recapitulating, the estimation of the  ${}_5L_x^*$  values is carried out according to the following equations:

$${}_5L_x^* = {}_5N_x \exp(R(x)) \quad (D.12)$$

and

$${}_{\omega-A}L_A^* = {}_{\omega-A}N_A \exp(R(A)). \quad (D.13)$$

It should be noted that the sequence of  ${}_5L_x^*$  values obtained in this way is likely to be more erratic than one derived from a set of  ${}_5q_x$  values calculated on the basis of observed central mortality rates. In some instances, the estimated  ${}_5L_x^*$  values may even increase with age. Errors in the age distributions used as input are usually the cause of this erratic behaviour. It is in order to minimize the effects of such errors and also to generate a measure that is comparable with those usually found in other sources that the pseudo  ${}_5L_x^*$  values are converted into expectation of life in the manner described below.

*Step 6: calculation of expectation of life.* The expectation of life at age  $x$ ,  $e_x$ , is calculated by cumulating the pseudo  ${}_5L_x^*$  values obtained in the previous step and dividing the sum by an estimate of  $l(x)$ , the number of survivors to exact age  $x$  in the life table. An adequate estimate of  $l(x)$  can be obtained as

$$l^*(x) = ({}_5L_x^* - {}_5L_x^*) / 10.0 \quad (D.14)$$

where the \* has been added to remind the reader that these are also pseudo  $l^*(x)$  values with unknown radix. Then, letting  $T_x^*$  be the number of person-years lived above age  $x$ , its value is calculated as

$$T_x^* = \sum_{y=x}^{A-5} {}_5L_y^* + {}_{\omega-A}L_A^*, \quad (D.15)$$

so by combining equations (D.14) and (D.15),  $e_x$  can be estimated as

$$e_x = 10.0(T_x^*) / ({}_5L_x - {}_5L_x^*). \quad (D.16)$$

Once the life expectancy figures have been calculated, usually for  $x$  ranging from 10 to 50, the levels they imply in a model life-table system can be found, and a final estimate of mortality can be obtained by averaging the most reliable estimates of mortality level (those left after discarding any clearly unsuitable values). In practice, the mortality estimates for values of  $x$  up to age 30 or so are reasonably consistent, but after age 30 or 35

they often show progressively lower mortality as age  $x$  increases. The best estimate of overall mortality may therefore be an average of the levels associated with  $e_x$  for  $x$  ranging from 10 to 30, though this conclusion implies that the results will not be a useful basis for the selection of an age pattern of mortality, nor will they be good indicators of the necessity of adjustment when errors in the growth rates arise because of changes in enumeration completeness.

#### 4. A detailed example: Panama, 1960-1970

The case of Panama between 1960 and 1970 is used to illustrate the application of this method, so it will be possible to compare its results with those obtained above through the application of the intercensal-survival techniques.

The basic data are shown in columns (2) and (4) of table 173, but for the sake of completeness, they are reproduced in columns (2) and (3) of table 186. Note that when using this method there is no need to adjust for intercensal intervals that are not round numbers of years; therefore, the populations just as enumerated in 1960 and 1970 can be used.

The computational procedure for this example is given below.

*Step 1: adjustment for net intercensal migration and territorial coverage.* As described in step 1 of subsection B.2(c), no adjustments are carried out in this case.

*Step 2: calculation of age-specific intercensal growth rates.* The interval between the 1960 and 1970 censuses was 9.41 years (see subsection B.2(c)). The growth rate for each age group is therefore calculated by dividing the difference between the natural logarithms of the final and initial populations of each age group by 9.41. Thus, for the population of age group 5-9,

$$\begin{aligned} {}_5r_5 &= [\ln({}_5N_{25}) - \ln({}_5N_{15})] / 9.41 \\ &= (11.5801 - 11.2463) / 9.41 = 0.03547. \end{aligned}$$

Results are shown in column (4) of table 186.

*Step 3: calculation of average intercensal age distribution.* An average age distribution for the intercensal period is obtained simply by calculating the arithmetic means of the initial and final populations of each age group. Thus for age group 5-9,

$$\begin{aligned} {}_5N_5 &= 0.5({}_5N_{15} + {}_5N_{25}) = 0.5(76,598 + \\ &106,944) = 91,771. \end{aligned}$$

Full results are shown in column (5) of table 186.

*Step 4: cumulation of age-specific growth rates from age 5 upward.* Cumulated age-specific growth rates are required to estimate  ${}_5L_x$  for all values of  $x$  from 5 upward. The average population aged 5-9 years,  ${}_5N_5$ , needs to be inflated by 2.5 years of growth at the age-specific growth rate for the 5-9 age group, namely,  ${}_5r_5$ . The average population aged 10-14 years,  ${}_5N_{10}$ , needs to be inflated by five years of growth at the age-specific rate for the 5-9 age group,  ${}_5r_5$ , plus 2.5 years of growth at the

TABLE 186. ESTIMATION OF INTERCENSAL MORTALITY FOR FEMALES, USING INTERCENSAL GROWTH RATES, PANAMA, 1960-1970

Age group $x, x + 4$ (1)	Enumerated female population		Intercensal growth rate $s^r_x$ (4)	Average population 1960-1970 (5)	Cumulated growth rate $R(x)$ (6)	Stationary population			Expectation of life $e_x$ (10)	West mortality level (11)
	1960	1970				In five-year age groups ${}_5L_x^*$ (7)	At exact age $x$ $l^*(x)$ (8)	Over age $x$ $T_x^*$ (9)		
	(2)	(3)								
0.....	90 071	114 017	0.02505	102 044	-	-	-	-	-	-
5.....	76 598	106 944	0.03547	91 771	0.08868	100 281	-	-	-	-
10.....	63 635	85 253	0.03108	74 444	0.25505	96 072	19 635	1 123 232	57.21	16.9
15.....	54 431	73 381	0.03175	63 906	0.41213	96 500	19 257	1 027 160	53.34	17.4
20.....	45 634	63 010	0.03429	54 322	0.57723	96 753	19 325	930 660	48.16	16.8
25.....	37 818	50 924	0.03162	44 371	0.74200	93 185	18 994	833 907	43.90	16.7
30.....	32 179	40 885	0.02545	36 532	0.88468	88 488	18 167	740 722	40.77	17.8
35.....	28 724	36 115	0.02433	32 420	1.00913	88 935	17 742	652 234	36.76	18.1
40.....	23 974	29 409	0.02171	26 692	1.12423	82 154	17 109	563 299	32.92	18.6
45.....	20 618	25 360	0.02200	22 989	1.23350	78 926	16 108	481 145	29.87	20.3
50.....	15 068	21 775	0.03913	18 422	1.38633	73 691	15 262	402 219	26.35	21.3
55.....	11 999	17 632	0.04090	14 816	1.58640	72 393	14 608	328 528	-	-
60.....	10 283	13 004	0.02495	11 644	1.75103	67 076	13 947	256 135	-	-
65.....	6 737	10 061	0.04262	8 399	1.91995	57 286	12 436	189 059	-	-
70.....	5 242	6 690	0.02592	5 966	2.09130	48 297	10 558	131 773	-	-
75+.....	6 756	9 873	-	8 315	2.30650	83 476	-	83 476	-	-

rate for the 10-14 group,  $s^r_{10}$ . Values of  $R(x)$  are therefore found by successive cumulation of the age-specific rates following equations (D.6) and (D.7). For age group 5-9,

$$R(5) = 2.5(s^r_5) = 0.08868.$$

For age group 10-14,

$$R(10) = 2.5(s^r_{10}) + 5.0(s^r_5) = 0.25505.$$

For age group 20-24,

$$R(20) = 2.5(s^r_{20}) + 5.0(s^r_5 + s^r_{10} + s^r_{15}) = 0.57723.$$

To calculate  $R(A)$ , the inflation factor corresponding to the open-ended interval, equation (D.8) is used in conjunction with equation (D.9) to estimate  $\rho(A)$ . The latter uses as inputs the values of  $r(10+)$ ,  $N(10+)$  and  $N(45+)$ . The symbol  $r(10+)$  is the growth rate of the population 10 and over during the intercensal period, and it is calculated as any other growth rate, as is illustrated below:

$$\begin{aligned} r(10+) &= \ln(N_2(10+)/N_1(10+))/t \\ &= \ln(483,372/363,098)/9.41 \\ &= 0.2861/9.41 \\ &= 0.0304. \end{aligned}$$

The values of  $N(10+)$  and  $N(45+)$  are found by cumulating the necessary entries in column (5) of table 186. In the case of  $N(45+)$ , one begins with  ${}_5N_{45}$  and continues until  $N(75+)$ . For  $N(10+)$ , the starting-point is  ${}_5N_{10}$ . The resulting values are  $N(10+) = 423,238$  and  $N(45+) = 90,551$ . Hence,

$$\rho(75) = a(75) + b(75)r(10+) + c(75)\ln(N(45+)/N(10+))$$

$$= 0.053 + 6.40(0.0304) + 0.063 \ln(90,551/423,238)$$

$$= 0.1504$$

and

$$R(75) = \rho(75) + 5.0 \sum_{y=5}^{70} s^r_y$$

$$= 0.1504 + 2.1561$$

$$= 2.3065.$$

The complete set of  $R(x)$  values is shown in column (6) of table 186.

Step 5: reduction of age distribution to a stationary form. Values of  ${}_5L_x^*$  are now obtained for each  $x$  by multiplying the average population of the age group,  ${}_5N_x$ , by  $\exp[R(x)]$ . Thus, for age group 5-9,

$${}_5L_5^* = (91,771) \exp(0.08868) = 100,281$$

and for age group 45-49,

$${}_5L_{45}^* = (22,989) \exp(1.23350) = 78,926.$$

The open interval is treated in just the same way, remembering that the result is an estimate of  ${}_{\omega-A}L_A^*$ , or  $T_A^*$ . Full results are shown in column (7) of table 186.

Step 6: calculation of expectation of life. The expectation of life at age  $x$ ,  $e_x$ , is equal to the person-years lived from age  $x$  onward,  $T_x^*$ , divided by the number of survivors to age  $x$ ,  $l^*(x)$ . The value of  $l^*(x)$  can be estimated with sufficient accuracy by averaging adjacent values of person-years lived, that is,  ${}_5L_{x-5}^*$  and  ${}_5L_x^*$ . Thus, for age 10,

$$l^*(10) = ({}_5L_5^* + {}_5L_{10}^*)/10.0$$

$$= (100,281 + 96,072)/10.0 = 19,635.$$

Note that because 5 is the lowest value of  $x$  used in the calculations and no value of  ${}_5L_0^*$  is obtained,  $l^*(10)$  is the youngest population that can be calculated. It may be noted in passing that no value of  ${}_5L_0^*$  is obtained because age group 0-4 is generally severely distorted by omission and age-misreporting. Similarly, the highest age for which an  $l^*(x)$  value can be estimated is age  $A-5$ , since no value is available for  ${}_5L_A^*$ . Column (8) of table 186 shows the range of  $l^*(x)$  values.

The  $t_x^*$  column is calculated by cumulating from age  $A$  downward the successive  ${}_5L_x^*$  values. Thus, for 75,

$$T_{75}^* = {}_{A-75}L_{75}^* = 83,476$$

and for 50,

$$T_{50}^* = T_{55}^* + {}_5L_{50}^* = 328,528 + 73,691 = 402,219.$$

Full results are shown in column (9) of table 186.

The expectation of life at each age  $x$  from 10 to 50 is then calculated by dividing each  $T_x^*$  by the corresponding  $l^*(x)$ . Thus, for  $x = 10$ ,

$$e_{10} = T_{10}^* / l^*(10) = 1,123,232 / 19,635 = 57.21,$$

whereas for  $x = 50$ ,

$$e_{50} = T_{50}^* / l^*(50) = 402,219 / 15,262 = 26.35.$$

The figures for expectation of life shown in column

(10) of table 186 can be compared with those of a model life-table system. Given the assumption made earlier that the West family provides the best fit to the age pattern of mortality in Panama, the West mortality level implied by each female  $e_x$  in column (10) has been found by interpolation, the results being shown in column (11).

It will be noticed that the first five values, for the age range from 10 to 30, are more or less consistent, the average level being 17.1. Above age 25, however, the mortality levels rise steadily with age, no doubt as a result of age exaggeration. It is interesting to compare the results obtained by this method with those obtained by the projection technique using cumulated values and described in subsection B.4(d). The similarities between the mortality estimates given in column (11) of table 186 and those in column (16) of table 180 are striking. In the latter case, as in the former, a series of five approximately consistent estimates of mortality level, averaging a level of 17.4 instead of 17.1, is followed by a pronounced upward trend in level with age. The present technique may be preferred on two grounds, the first being its simplicity of application and the second that it does not have the drawback of giving different results according to the direction, forward or backward, in which the projection is performed. It must be pointed out, however, that the estimated adult mortality level of 17.1, which is the best available from table 186, is to some extent distorted by the age exaggeration detected, since it inflates the  $T_x$  values at all ages.