II. ESTIMATION OF BASE POPULATION BY SEX AND AGE

44. The methods described here have been developed to project populations classified by five-year groups of ages over successive five-year intervals of time. Such projections must proceed from an enumeration or estimate of the population, divided according to sex and five-year age groups, for the date which has been selected as the starting point.

45. In some countries, estimates of the population by sex and age are made currently with the help of vital statistics showing the changes since the latest census date. In that case, no problem exists, provided that the estimates are sufficiently reliable and are presented in the form, with regard to age grouping, that is desired for projection. Likewise, there is no problem if a census has been taken quite recently, providing reliable data by sex and age groups in the desired form, and if the census date is a suitable starting point for the projection. This chapter deals with situations in which these ideal conditions are not met.

46. The first case to be considered is that of a country for which population data by sex and age in the desired form, and meeting the required standard of accuracy,[6] are available for a prior date. In this case the problem consists in bringing the data forward to the starting date for the projection. This problem is taken up in part A of the chapter.

47. Part B deals with the methods of adjusting inaccurate data. Such adjustments are indispensable if distortion of the projection is to be avoided. The adjustment can be made either before or after the data have been brought forward to the starting date of the projection. A different kind of adjustment is required if the data are not tabulated according to the desired grouping of ages; procedures for this kind of adjustment are described in part C.

48. Finally, it is necessary to consider the case of a country which lacks suitable data on population by sex and age to form the basis for a projection. In this case, it may nevertheless be possible to make an estimate that will satisfy the purpose. Some methods of dealing with this problem are taken up in part D.

A. Bringing statistics forward to the starting date for the projection

1. *"Pro-rating"*

49. The simplest method to carry forward census data on sex and age to a more recent date is to distribute the population total estimated for the latter date[7] by sex and age in the same proportions as those observed at the census. This is done by determining first the ratio between the total populations at the two dates and then multiplying all the observed numbers in each sex-age category by this ratio.

50. The method is illustrated in table 1 by statistics for Barbados (although, as will be shown later, the application of the method in this case is questionable). A sex-age distribution is available from the census of 9 April 1946, for an enumerated population of 192,800. For seven individuals no age was returned, so that the sex-age data refer to a population of 192,793. If the population projection is to start from mid-year 1950—for which date the total population of Barbados was estimated at 209,239—the required base population can be estimated by multiplying all numbers of the 1946 census by the ratio of 192,793 to 209,239, i.e., by 1.0853.

Table 1. Sex-age distribution of population of Barbados at 1946 census "pro-rated" to mid-year 1950

Ages	Census, 1946		Estimate, 1950	
	Males	Females	Males	Females
0–4	11,297	11,580	12,261	12,568
5–9	10,695	10,943	11,607	11,876
10–14	9,578	9,926	10,395	10,773
15–19	8,751	9,551	9,497	10,366
20–24	7,079	8,227	7,683	8,929
25–29	6,494	7,863	7,048	8,534
30–34	6,416	7,902	6,963	8,576
35–39	6,562	7,954	7,122	8,632
40–44	5,438	6,716	5,902	7,289
45–49	3,959	5,994	4,297	6,505
50–54	2,723	4,639	2,955	5,035
55–59	1,689	3,866	1,833	4,196
60–64	1,493	3,369	1,620	3,656
65–69	1,301	3,046	1,412	3,306
70–74	1,173	2,564	1,273	2,783
75–79	665	1,638	722	1,778
80–84	272	809	295	878
85 and over	136	485	148	526
Total	85,721	107,072	93,033	116,206
Age unknown	6	1	—	—

51. In this example, the number of persons of unknown ages was so small that it could be safely neglected. Methods of adjustment where the numbers of persons of unknown age are important are taken up in part C. If separate estimates of total numbers of males and females are available for the latter date, "pro-rating" should be done for each sex separately.

52. This simple method is quite adequate in almost any situation for a short time interval such as a year or less, unless a high degree of accuracy can be achieved in other steps of the estimating procedure. If the interval between the census and the date of the estimate is longer,

[6] Methods of testing the accuracy of population statistics by sex and age groups are described in: United Nations, *Methods of Appraisal of Quality of Basic Data for Population Estimates* (Manuals on methods of estimating population, No. II), ST/SOA/Series A, Population Studies, No. 23.

[7] Methods of estimating population totals for dates since the latest census are described in: United Nations, *Methods of Estimating Total Population for Current Dates* (Manuals on methods of estimating population, No. I), ST/SOA/Series A, No. 10.

"pro-rating" may still be adequate, provided (*a*) that there were no serious irregularities in the age structure at the time of the census, and (*b*) that no important changes occurred in the birth rate, mortality conditions affecting particular age groups, or the flow of migration during the post-censal intervals.

53. The figures for Barbados provide an example of an irregularity of age structure which is sufficient to distort considerably the "pro-rated" estimate for an interval as long as four years. According to the 1946 census, the numbers of both males and females 35 to 39 years old exceeded the numbers 25 to 29 and 30 to 34 years, possibly because of emigration of young adults during the years prior to the census. On the assumption that the census figures are correct, the "pro-rated" figures for the group 35 to 39 years old in 1950 are presumably excessive, since this group would be made up primarily of the survivors from the relatively small group 30 to 34 years old in 1946. The 1950 figures for persons 30 to 34 years old are also likely to be excessive, and those for persons 40 to 44 years old underestimated. Such errors can sometimes be avoided by making separate estimates for the age groups affected, and "pro-rating" the remainder.

54. Changes in mortality during a fairly short interval between the census date and the date of the estimate will ordinarily have little effect on the age distribution of the population, except that a sharp reduction of infant mortality may substantially increase the proportion of young children. The effect of changes in the birth rate will be limited to the youngest age groups. It is often desirable to make separate estimates, as explained further on, for the numbers of small children, and confine the "pro-rating" to the higher age groups. The effects of a moderate amount of migration during a relatively short period can often be safely disregarded where the scope and quality of the available data do not justify an attempt at a high standard of accuracy. However, in a situation such as the figures for Barbados imply, migration may be sufficient to invalidate any "pro-rated" estimates for young adults. Unless the previous heavy emigration suggested by the 1946 census results had ceased, not only the residual effect of that migration upon the 1950 age distribution but also the further effects of migration between 1946 and 1950 would have to be taken into account for a valid estimate.

2. *Utilization of data on births and deaths by age*

55. If adequate statistics on births and on deaths by age are available, the sex-age data of a census can be brought forward to a more recent date by adding the births and subtracting the deaths during each year from the numbers of persons of corresponding age in that year. This procedure ignores migration, which, if important, would have to be dealt with separately. The calculation can be carried to a great degree of refinement if data by single year of age are used, but it is laborious, and few countries have accurate enough statistics to warrant such meticulous calculation. A procedure utilizing data in five-year groups of ages for five-year intervals of time will ordinarily suffice, only the age group under one year being treated separately.

56. Deaths are tabulated according to the ages of the decedents at death, whereas the classification desired for the present purpose refers to their ages at the beginning of the period for which the calculations are made. Only a part of the deaths occurring in a given five-year age group during a period of five years belongs to the cohort[8] of that age group at the beginning of the period; the remainder is attributable to the next younger cohort. However, with the exception of infants and small children, annual numbers of deaths do not ordinarily change greatly from one five-year age group to the next. Hence, no serious error is incurred in calculations for the groups over five years of age by assigning one-half of the tabulated deaths of the same age group and one-half of those of the next age group to a given cohort. The procedure will presently be illustrated with data for Portugal (see tables 2 and 3).

57. Since infant deaths are always far more numerous than deaths at ages past infancy, the cohort aged initially less than five years, and the five-year cohort born during the period for which the calculation is made, must be dealt with in a different fashion. If high accuracy were desired, the separation of numbers of infant deaths and deaths of small children would have to be calculated with a somewhat complicated formula in which the distribution of infant deaths by month of age is also taken into account.[9] Moderately accurate results, however, can be derived by a much simpler formula which, for the present purpose, is quite adequate.

58. Since more infant deaths occur shortly after birth than during the remainder of the infants' first year of life, a great portion of infant deaths in the first year after the census will occur to infants born during that same year than to infants already born before the date of the census. A rough rule which has been widely employed and is here recommended is to attribute two-thirds of infant deaths occurring in a year to infants born during that year; and one-third to infants born during the preceding year.[10] This rule will have to be applied only with respect to infant deaths occurring during the first year after the census. None of the infants dying in subsequent years could have been born prior to the census.

59. There is always a considerable decrease in age-specific death rates from age 1 to age 4; more deaths occur to children aged 1 and 2 than to children aged 3 and 4. If deaths were uniformly distributed among ages 1-4 (which is never the case), then five-eighths of these deaths would be ascribed to the cohort aged 0-4 at the census and three-eighths to the cohort born in the subsequent 5 years.[11] Experimental calculations have shown, however, that it is usually fairly accurate to assume that

[8] The term "cohort" is used throughout this manual in the technical sense of a group of persons born during the same period (in this case a five-year period), who consequently belong to the same age group at any given time.

[9] See: W. P. D. Logan, "The measurement of infant mortality", *Population Bulletin of the United Nations*, No. 3 (October, 1953); also United Nations, *Foetal, Infant and Early Childhood Mortality, Vol. 1. The Statistics.* (ST/SOA/Series A., No. 13.)

[10] This rule is more accurate where mortality is high than where it is low. Where infant mortality is extremely low, up to 80 per cent of infant deaths occur to infants born during the same year. For the present purpose, however, great accuracy is not needed where mortality is low since the number of deaths involved is then comparatively small.

[11] This would be so because no children born after the census would reach age 1 until the second year following the census.

one-half of deaths occurring at ages 1-4 belong to the cohort initially aged 0-4 and one-half to the cohort born in the course of the five-year period.[12]

60. In the following example, the data of the Portuguese census of 12 December 1940 and of Portuguese vital statistics for the period 1941-1945 are employed. Both the census data and the death statistics include returns for which ages were not reported. Owing to their small numbers (0.3 per cent of the census population and 0.2 per cent of recorded deaths), these are neglected at the present stage of the computation.[13] The

[12] The calculations have shown this to be equally accurate where mortality is high and where it is low. This accuracy, though not perfect, is quite sufficient for the present purposes.

[13] Otherwise, "pro-rating" of the groups of unknown age would be recommended.

TABLE 2. DEATHS IN PORTUGAL, 1941–1945, ATTRIBUTABLE TO COHORTS OF GIVEN AGES ON 31 DECEMBER 1940

Sex and age group (years)	Deaths reported in this age group					Totals, 1941–1945	Deaths attributable to cohort initially in this age group[a]
	1941	1942	1943	1944	1945		
Males							
0	14,934	13,482	14,339	13,535	13,191	69,481	27,999[b]
1– 4	9,443	8,212	7,386	7,068	6,319	38,428	
5– 9	1,631	1,749	1,572	1,405	1,257	7,614	6,155
10–14	956	1,061	974	860	845	4,696	5,864
15–19	1,381	1,516	1,496	1,388	1,251	7,032	7,809
20–24	1,658	1,706	1,695	1,743	1,784	8,586	8,180
25–29	1,648	1,556	1,618	1,448	1,503	7,773	8,004
30–34	1,729	1,652	1,601	1,643	1,609	8,234	8,556
35–39	1,780	1,757	1,769	1,753	1,818	8,877	9,138
40–44	1,849	1,808	1,853	1,915	1,974	9.399	9,788
45–49	2,068	1,957	1,929	2,093	2,129	10,176	10,966
50–54	2,419	2,274	2,261	2,423	2,378	11,755	12,764
55–59	2,778	2,720	2,601	2,870	2,804	13,773	15,686
60–64	3,855	3,641	3,288	3,456	3,358	17,598	19,186
65–69	4,416	4,245	4,010	4,173	3,931	20,775	21,668
70–74	4,903	4,586	4,275	4,529	4,268	22,561	21,879
75–79	4,622	4,334	4,056	4,195	3,990	21,197	18,344
80–84	3,385	3,125	3,019	3,026	2,936	15,491	18,569
85 and over	2,433	2,136	2,004	2,088	2,162	10,823	
Females							
0	12,860	11,147	11,939	11,068	10,843	57,857	25,467[c]
1– 4	8,745	7,698	6,892	6,351	5,860	35,546	
5– 9	1,445	1,595	1,397	1,353	1,024	6,814	5,686
10–14	939	1,039	929	839	812	4,558	5,612
15–19	1,296	1,480	1,422	1,262	1,206	6,666	7,018
20–24	1,438	1,479	1,483	1,475	1,495	7,370	7,080
25–29	1,432	1,473	1,355	1,312	1,219	6,791	6,863
30–34	1,431	1,426	1,366	1,341	1,371	6,935	6,972
35–39	1,497	1,457	1,481	1,301	1,272	7,008	6,960
40–44	1,445	1,385	1,316	1,415	1,351	6,912	6,927
45–49	1,492	1,417	1,336	1,318	1,379	6,942	7,854
50–54	1,860	1,878	1,774	1,572	1,681	8,765	9,644
55–59	2,251	2,170	2,008	2,049	2,045	10,523	12,635
60–64	3,205	3,180	2,834	2,801	2,727	14,747	16,839
65–69	4,041	3,960	3,771	3,631	3,528	18,931	21,383
70–74	5,235	5,032	4,711	4,500	4,357	23,835	24,770
75–79	5,878	5,285	4,925	4,815	4,802	25,705	24,683
80–84	5,217	4,937	4,626	4,485	4,396	23,661	34,927
85 and over	5,090	4,769	4,363	4,486	4,389	23,097	

[a] For initial ages 5 and over, half the deaths at the same ages and half those at the next higher ages were attributed to the given cohort. For initial ages 0–4, the deaths attributed to the cohort comprise: one-third of infant deaths (age 0) in the year 1941 only, one-half of all deaths at ages 1–4 and one-half of all deaths at ages 5–9. For those born after 1940, the following deaths have been attributed: all infant deaths except one-third of those occurring in 1941, and one-half of all deaths at ages 1–4.

[b] The balance of 83,717 is attributable to the male cohort born 1941–1945.

[c] The balance of 71,343 is attributable to the female cohort born 1941–1945.

TABLE 3. UTILIZATION OF STATISTICS ON BIRTHS AND DEATHS IN BRINGING PORTUGUESE SEX-AGE DISTRIBUTION FORWARD FROM 1940 TO 1945

Age of cohort as of 31 December 1940 (years)	*Size of cohort enumerated as of 12 December 1940* [a]	*Births, 1941–1945* [b]	*Deaths attributed to cohort, 1941–1945*	*Size of cohort estimated as of 31 December 1945*	*Age of cohort as of 31 December 1945 (years)*
Males					
Unborn.....	—	463,256	83,717	379,539	0– 4
0– 4.....	425,058	—	27,999	397,059	5– 9
5– 9.....	426,760	—	6,155	420,605	10–14
10–14.....	408 344	—	5 864	402,480	15–19
15–19.....	373,088	—	7,809	365,279	20–24
20–24.....	314,829	—	8,180	306,649	25–29
25–29.....	297,457	—	8,004	289,453	30–34
30–34.....	267,017	—	8,556	258,461	35–39
35–39.....	230,164	—	9,138	221,026	40–44
40–44.....	194,495	—	9,788	184,707	45–49
45–49.....	170,842	—	10,966	159,876	50–54
50–54.....	154,011	—	12,764	141,247	55–59
55–59.....	125,538	—	15,686	109,852	60–64
60–64.....	114,063	—	19,186	94,877	65–69
65–69.....	82,026	—	21,668	60,358	70–74
70–74.....	57,148	—	21,879	35,269	75–79
75–79.....	33,848	—	18,344	15,504	80–84
80–84.....	16,860	—	18,569	7,836	85 and over
85 and over.	9,545				
Females					
Unborn.....	—	441,188	71,343	369,845	0– 4
0– 4.....	404,083	—	25,467	378,616	5– 9
5– 9.....	408,961	—	5,686	403,275	10–14
10–14.....	395,012	—	5,612	389,400	15–19
15–19.....	375,394	—	7,018	368,376	20–24
20–24.....	315,853	—	7,080	308,773	25–29
25–29.....	311,429	—	6,863	304,566	30–34
30–34.....	289,619	—	6,972	282,647	35–39
35–39.....	265,763	—	6,960	258,803	40–44
40–44.....	227,893	—	6,927	220,966	45–49
45–49.....	204,339	—	7,854	196,485	50–54
50–54.....	192,155	—	9,644	182,511	55–59
55–59.....	159,884	—	12,635	147,249	60–64
60–64.....	148,265	—	16,839	131,426	65–69
65–69.....	112,244	—	21,383	90,861	70–74
70–74.....	82,996	—	24,770	58,226	75–79
75–79.....	53,126	—	24,683	28,443	80–84
80–84.....	30,575	—	34,927	15,572	85 and over
85 and over..	19,924				

[a] It is assumed for the purpose of the calculation, that the enumerated population as of 12 December 1940 equals the actual population as of 31 December 1940.

[b] There were 904,444 registered births, of which it was assumed that 51.22 per cent were male and 48.78 per cent female.

census having been taken so near the end of the year, the census population is assumed to equal the population at the end of the year as of 31 December 1940.

61. The results of these computations for Portugal would have to be further "pro-rated" to agree with an available estimate of total population for the end of 1945.

62. With modifications, this method is also applicable in different situations. If the date of the census was not so near the end of the year, numbers of births and deaths would have to be calculated for a following five-year period not coinciding with calendar years. Statistics on births and deaths for each month or quarter, if available, may be helpful in this connexion but, except where

high accuracy is attainable, it is sufficient to estimate numbers of births and deaths for fractions of years from the annual figures, by assuming that these numbers are distributed uniformly throughout the year. Though seasonal fluctuations introduce an error in the estimate for the fraction of the first year, this error is mostly compensated by a similar error for the remaining fraction of the last year.

63. If deaths of infants are not distinguished in the statistics from deaths of children aged 1-4, another, though less accurate, rough formula can be used to separate deaths attributable to the cohort initially aged 0-4 years from those to children born during the period. Experimental calculations have shown that, on an average, about four-fifths of all deaths at ages 0-4 occur to children born within the five-year period, and only one-fifth to the initial cohort of small children. The fraction tends to be lower where mortality is high and higher where mortality is low. It is preferable to use the formulas previously mentioned wherever the statistics permit a distinction of infant deaths from deaths at ages 1-4.

64. Where the statistics of the population and of deaths are not sufficiently accurate, or where they are not available in the required five-year age groups, modifications of the data are required before the method can be applied. Some of these modifications will be discussed further on in this chapter. Where an estimate of migration is also needed, methods such as those suggested in chapter VI also have to be used.

3. *Interpolation with respect to time*

65. The procedure which has been illustrated with statistics for Portugal serves to bring the census statistics forward through a time-interval of exactly five years. By repetition, the data can also be brought forward by 10 or 15 years. The time-interval between the census and the base date for the projection, however, is not necessarily an exact multiple of five years.

66. As already indicated, "pro-rating" can be used to make the estimates agree with an estimate of total population for some different date. However, if there is a considerable interval between that date and the one to which the detailed estimates refer, "pro-rating" results in some loss of accuracy, Since the procedure which has been described is designed to maintain a fairly high degree of accuracy, "pro-rating" of its results for a time-interval of one year or more is not desirable. It is greatly preferable to interpolate the figures estimated both for the beginning and the end of the five-year period in question, with respect to numbers in each sex-age group. "Pro-rating" implies an assumption that the proportionate numbers in the different sex-age groups remain unchanged; this assumption is not justified unless demographic conditions have been very stable. Interpolation, on the other hand, depends on the assumption that the relevant conditions change in a gradual, regular fashion; under ordinary conditions, this assumption is usually approximately correct and the errors to which it may give rise within a five-year period are usually very slight.

67. Using once more the data for Portugal, let us assume, by way of an example, that an estimate is required for the middle of 1944. This date is 3½ years after the date of the census (end of 1940) and 1½ years before the date of the detailed estimates (end of 1945). Interpolation, therefore, must be carried out in the ratio of 3½ to 1½, or 7 to 3. This is done, as in table 4, by multiplying all estimated figures for 1945 by 7, adding 3 times the census figures for the same age groups, and dividing the sums by 10. The results are then "pro-rated" to the mid-year 1944 official estimate of total population.

TABLE 4. INTERPOLATION OF PORTUGUESE SEX-AGE DISTRIBUTIONS OF END-1940 AND END-1945 TO OBTAIN ESTIMATE FOR MID-1944

Sex and age group (years)	*Enumerated, 12 December 1940*	*Estimated, 31 December 1945*	*Interpolated, mid-1944*	
			Not "pro-rated"	*"Pro-rated" to 1944 estimate of total population*
Males				
0–4.....	425,058	379,539	393,195	396,878
5–9.....	426,760	397,059	405,969	409,771
10–14.....	408,344	420,605	416,927	420,832
15–19.....	373,088	402,480	393,662	397,349
20–24.....	314,829	365,279	350,144	353,424
25–29.....	297,457	306,649	303,891	306,737
30–34.....	267,017	289,453	282,722	285,370
35–39.....	230,164	258,461	249,972	252,313
40–44.....	194,495	221,026	213,067	215,063
45–49.....	170,842	184,707	180,548	182,239
50–54.....	154,011	159,876	158,116	159,597
55–59.....	125,538	141,247	136,534	137,813
60–64.....	114,063	109,852	111,115	112,156
65–69.....	82,026	94,877	91,021	91,875
70–74.....	57,148	60,358	59,395	59,951
75–79.....	33,848	35,269	34,843	35,169
80–84.....	16,860	15,504	15,911	16,060
85 and over	9,545	7,836	8,349	8,427
TOTAL			3,805,382	3,841,024
Females				
0–4.....	404,083	369,845	380,116	383,676
5–9.....	408,961	378,616	387,720	391,351
10–14.....	395,012	403,275	400,796	404,550
15–19.....	375,394	389,400	385,198	388,806
20–24.....	315,853	368,376	352,619	355,922
25–29.....	311,429	308,773	309,570	312,469
30–34.....	289,619	304,566	300,082	302,893
35–39.....	265,763	282,647	277,582	280,182
40–44.....	227,893	258,803	249,530	251,867
45–49.....	204,339	220,966	215,978	218,001
50–54.....	192,155	196,485	195,186	197,014
55–59.....	159,884	182,511	175,723	177,369
60–64.....	148,265	147,249	147,554	148,936
65–69.....	112,244	131,426	125,671	126,848
70–74.....	82,996	90,861	88,502	89,331
75–79.....	53,126	58,226	56,696	57,227
80–84.....	30,575	28,443	29,083	29,355
85 and over	19,924	15,572	16,878	17,036
TOTAL			4,094,484	4,132,833

68. The difference between the total of the interpolated estimates and the independent estimate of total population is partly accounted for, in this instance, by the fact that persons whose ages were not reported in the census were excluded from the computations. Differences may also arise because of the fluctuation of the number of births during the five-year interval, or because of migratory movements. If an examination of the data shows that the latter factors are important, a modified estimate for the age group 0-4 may be substituted and estimates of the effects of migration added or subtracted, before "pro-rating" is done.

69. It is also possible to make a short-range projection in order to move the estimates beyond the date of the most recent vital statistics to a date considered more appropriate as the starting point for long-range projections. For example, if estimates have been made for 1952 by bringing forward the data of a 1947 census, and if the desired starting point is 1955, the 1952 estimates may be projected to 1957 with suitable assumptions—preferably simple extrapolation of trends in the vital rates—and the results interpolated with the 1952 figures to obtain estimates for 1955.

70. The validity of interpolated estimates by five-year age groups depends on the assumption that the population in each group is distributed fairly evenly by single-year ages within the group. (The same assumption is involved where statistics of deaths of five-year age groups are used to bring forward population figures for five-year age groups over an interval of less than five years.) Where this assumption does not hold, for example because of large annual variations in the number of births during the period when a given five-year cohort was born, it may be necessary to make the calculations separately for single-year age groups. Census data classified by single-year age groups are often too unreliable for this purpose, because of errors in age reporting. The distribution of single-year groups within a five-year age group can often be estimated, however, by reference to the numbers of births in the corresponding years.

B. Adjustment for errors due to inaccurate age reports or faulty enumeration

71. Both census data on the age distribution of population and statistics of deaths by age are subject to errors due to inaccurate reporting of ages and to under-enumeration or under-registration. Those errors which have very little effect on the numbers in the five-year age groups can generally be ignored for the present purpose, even though they may affect substantially the distribution by single-year age groups. In the case of errors which substantially affect one or more of the five-year groups, at least a rough adjustment is necessary if a useful population projection is to be obtained.

72. The problem of errors due to incomplete census enumeration arises most commonly in the case of the age group 0-4. Children of this age are often very incompletely enumerated in the census. In the registration of deaths also, omissions are generally most likely to occur in this same age group. Incomplete enumeration and registration may also affect other age groups, and in some censuses the military population, which consists predominantly of young men, is either excluded or treated separately.

73. If incomplete enumeration of a particular age group must be assumed, unless there are compensating errors in the enumeration of other age groups, it follows that the total population as shown by the census is understated. Nevertheless, it may be desirable to keep the projections in line with the published census total. Otherwise, unwarranted comparisons might be made, leading to confusion. Therefore, if an adjustment is made with respect to some age group, it is usually recommendable to "pro-rate" the resulting figures once more so that they fall in line again with the totals established by the census or current official estimates.

74. In the case of data substantially affected by errors in age reporting, smoothing is indicated. Such errors, both in census enumerations and in the registration of deaths, commonly result in alternating understatement and overstatement of the figures for successive five-year groups, caused by the different attraction of the various terminal digits for statements of age. This has been called the "saw-tooth" effect of age errors.[14] A simple moving average would not suffice to smooth out such irregularities. Nevertheless, for the present purposes, the simplest possible formula is recommended. It cannot be claimed for any graduation formula, no matter how elaborate, that it precisely corrects the errors in the data. Though a more refined formula[15] may produce smoother and apparently more consistent results, it remains doubtful whether a closer approximation to reality is thereby attained.

75. Procedures for smoothing age distributions affected by errors in age statements, and for correcting the effects of under-enumeration or under-registration in certain age groups, are described below, separately with reference to age groups in the range from 10 to 74 years, ages 0-4, 5-9, and 75 and over.

1. *Ages from 10 to 74*

76. Theoretically, different formulae should be used to smooth data exhibiting different degrees of inaccuracy, so that fairly accurate statistics would be modified only slightly while less accurate data would be more radically transformed. Further discussion of this question would lead far afield. In the interest of simplicity, only one smoothing formula for age groups in the range from 10 to 74 years is described here. This formula is appropriate for use where the data are markedly inaccurate. It is derived from a simple parabola and has been widely used for relatively simple work.

77. The formula employs five terms; that is, in order to adjust the figure for one five-year age group, data for the two preceding and the two following age groups must also be inserted in the formula. If the age statistics

[14] The characteristic effects of age errors are discussed in further detail in chapter III of *Methods of Appraisal of Quality of Basic Data for Population Estimates*, ST/SOA/Series A, Population Studies, No. 23. See also "Acurracy tests for census age distributions tabulated in five-year and ten-year groups", *Population Bulletin* No. 2, October 1952.

[15] A suitable procedure for more refined graduation of age data, by means of "pivotal values" and "osculatory interpolation" has been very well described by T. N. E. Greville in *United States Life Tables and Actuarial Tables, 1939-1941*, United States Bureau of the Census, Washington, D.C., 1946.

are tabulated by five-year groups up to age 85, smoothing can be effected by such a formula for all groups between the ages of 10 and 75. The numbers at the youngest and the oldest ages have to be dealt with separately.

The formula may be stated as follows:
$\Sigma = 1/16 (- S_{-2} + 4 S_{-1} + 10 S + 4 S_1 - S_2)$ where Σ is the adjusted number of persons in one five-year group, to be computed, S is the reported number of persons in the same five-year group, S_{-2} and S_{-1} are reported numbers in the two preceding five-year groups, and S_1 and S_2 are reported numbers in the two subsequent five-year groups.

78. An example of the use of this formula is given in table 5 by its application to the age statistics of the 1950 census of the Dominican Republic. Inspection of the census figures reveals marked irregularities which increase as age advances. In particular, it seems unreasonable that there should have been more females aged 60-64 than 55-59, and more aged 70-74 than 65-69. These and other irregularities are largely—if not entirely—overcome by use of the formula.

TABLE 5. "SMOOTHING" OF SEX-AGE DATA OF THE 1950 CENSUS OF THE DOMINICAN REPUBLIC

Age (years)	Males		Females	
	Population according to census	Figures adjusted by formula	Population according to census	Figures adjusted by formula
All ages...	1,070,065	—	1,065,088	—
0-4.....	189,383	—	186,458	—
5-9.....	150,704	—	147,061	—
10-14.....	141,661	133,194	135,179	133,819
15-19.....	101,552	110,903	124,194	124,585
20-24.....	105,152	97,933	109,241	106,924
25-29.....	77,620	79,599	79,196	80,580
30-34.....	59,618	62,180	60,011	61,666
35-39.....	60,137	57,150	55,494	53,567
40-44.....	47,183	48,016	42,978	42,833
45-49.....	36,551	37,244	30,621	32,710
50-54.....	30,712	29,414	28,908	25,588
55-59.....	21,049	22,884	14,894	18,769
60-64.....	19,716	17,530	18,760	15,212
65-69.....	9,502	11,350	8,688	11,342
70-74.....	8,176	7,007	9,628	7,740
75-79.....	3,941	—	4,073	—
80-84.....	3,697	—	4,725	—
85 and over	3,711	—	4,979	—

As applied to females aged 70-74, for instance, the formula works as follows:

$- S_{-2}$	− 18,760
$+ 4 S_{-1}$	+ 34,752
$+ 10 S$	+ 96,280
$+ 4 S_1$	+ 16,292
$- S_2$	− 4,725
TOTAL	123,839
SAME, divided by 16	7,740

79. Nevertheless, after the adjustment, some irregularity remains. It will be noted that the adjusted figures for females exceed considerably those for males at ages between 15 and 25 and in the group 70 to 74, whereas they are appreciably smaller than corresponding numbers of males between the ages of 35 and 65. A very plausible explanation is the tendency of many women of young and middle ages to understate their age; the resulting error cannot be removed by graduation. There are also probably some tendencies among men to misstate their ages, and there is a bias towards the exaggeration of age among the older persons of either sex. The correction of errors due to these biases would require the reconstruction of each sex-age group on the basis of statistics on births, deaths and migration over a period of almost a century. This is not possible in the case of the Dominican Republic. In fact, few countries possess reliable statistics covering such a long period of time.

80. If inspection of the data justifies a presumption that the age reports for one sex are more accurate than those for the other sex, an adjustment of the more faulty distribution can be made by reference to the more accurate one. The evidence is seldom clear, but occasionally the presumption as to relative accuracy is quite strong. The procedure is somewhat arbitrary and is recommended only if the data for one sex are obviously much more accurate than those for the other sex. This is not necessarily the case in the present example; nevertheless, the method is illustrated with reference to the Dominican statistics in table 6.

TABLE 6. ESTIMATE OF FEMALE POPULATION BY AGE GROUPS ON THE BASIS OF AGE DISTRIBUTION OF MALES, FOR THE DOMINICAN REPUBLIC, 1950

Age (years)	Adjusted number of males	Estimated ratio of females to males	Number of females: first estimate	Number of females: "pro-rated" estimate
10-14.....	133,194	1.005	133,860	130,468
15-19.....	110,903	1.010	112,012	109,173
20-24.....	97,933	1.015	99,402	96,883
25-29.....	79,599	1.020	81,191	79,133
30-34.....	62,180	1.025	63,734	62,119
35-39.....	57,150	1.035	59,150	57,651
40-44.....	48,016	1.045	50,177	48,905
45-49.....	37,244	1.055	39,292	38,296
50-54.....	29,414	1.065	31,326	30,532
55-59.....	22,884	1.075	24,600	23,977
60-64.....	17,530	1.085	19,020	18,538
65-69.....	11,350	1.095	12,428	12,113
70-74.....	7,007	1.105	7,743	7,547
TOTAL			733,935	715,335

The adjusted figures for ages 10-14 (see table 5) indicate a ratio of females to males of about 1.005; those for ages 70-74 give a ratio of 1.105. According to the experience of most life-tables, the ratio of females to males tends to rise very gradually with age, slowly at young ages and more rapidly at old ages. A smooth series

of hypothetical ratios can be inserted,[16] rising from 1.005 at ages 10-14 to 1.105 at ages 70-74. The numbers of males, as adjusted by the smoothing formula, are then multiplied by these ratios to obtain first estimates of corresponding numbers of females of the same ages. Initially, the adjusted figures for females between ages 10 and 75 totalled 715,335, whereas the newly computed numbers of females total 733,935. "Pro-rating" is used to bring the total back to 715,335.

81. When deciding whether to smooth census age data, one should always recall that no formula has the power of correcting errors precisely. All that any formula can do is to eliminate irregularities and put the statistics in a plausible form. Irregularities may, however, be the result of actual past fluctuations or discontinuities in demographic trends as well as of errors in the data. Therefore, smoothing should not be done if the statistics are fairly accurate: in such a case, a smoothing formula, by eliminating irregularities in population composition which the data truly reflect, would introduce greater errors than those which it corrects. If the data are inaccurate but past demographic trends are known to have been irregular, it may be appropriate to smooth the figures for some, but not all, age groups.

82. A situation of the latter type was encountered in preparing a population projection for Mexico.[17] The census age data for 1940 were found to be affected by irregularities, evidently due to inaccurate age reporting, similar to those observed in the Dominican Republic. In addition, however, the numbers of both sexes reported at ages 20-24 and 25-29 were considerably smaller than might have been expected. A plausible explanation for this circumstance was found in the Mexican Civil War, which lasted from 1910 to 1921 and severely disrupted life in various areas of the country at various times; as a result, birth rates during that period may have been abnormally low.[18]

83. Before applying the smoothing formula to the Mexican census data, therefore, it was necessary to insert, somewhat arbitrarily, such numbers for age groups 20-24 and 25-29 as might have survived to 1940 if birth rates during the Civil War period had remained fairly constant. The numbers inserted were obtained by linear interpolation between reported numbers at ages 15-19 and 30-34. The smoothing formula was then applied to obtain adjusted estimates for ages between 10 and 20, and between 30 and 75. For ages 20-24 and 25-29, the numbers originally reported were retained in the adjusted age distribution, on the assumption that they probably did not stray very much from the true numbers.

84. The smoothing formula is less well suited to statistics of deaths by age, in view of the large numbers of deaths in infancy and early childhood, followed by comparatively few deaths during later childhood and adolescence. With respect to deaths at ages 20 and over, however, the smoothing formula gives fairly satisfactory results.

2. *Ages 0-4*

85. As already stated, children under five years of age are incompletely enumerated in the censuses of many countries. Where the census figure for this age group is found to be seriously deficient, it should be replaced by an independent estimate. The deficiency is sometimes found to be important only in the case of infants under one year of age; if so, the independent estimate may be limited to these infants, and combined with the census figure for ages 1-4.

86. The method of deriving an estimate of the number of children under five years of age from the number of births during the preceding five-year time period, less the deaths among these children, has been described in part A of this chapter. This method depends upon the availability of accurate vital statistics. A modification of the method is necessary where nearly accurate statistics of births are available but the statistics of infant deaths are deficient. It must be noted in this connexion that infant deaths are commonly less completely registered than deaths in other age groups. The modification consists in substitution for the registered numbers of infant deaths, of an estimate based on assumed mortality conditions. Needless to say, the assumption as to mortality used for this purpose should correspond to the assumptions concerning this factor of population change to be made in later steps of the projection.

87. Where sufficiently reliable statistics of births also are not available, the number of births, too, may be estimated by means of appropriate assumptions. The relevant methods are discussed in subsequent chapters of this manual. Even though the estimate of the population 0-4 years of age obtained by such methods is very uncertain, it may still be much to be preferred over the census figure.

3. *Ages 5-9*

88. The numbers of children aged 5-9 as reported in the census can ordinarily be accepted, unless there are special reasons to doubt their accuracy. Examination of census data for many countries yields little evidence that numbers aged 5-9 years are subject to any special inaccuracy of reporting; in fact, even where the age data are generally very defective, the figures for this age group commonly appear to be more accurate than others. Experience has demonstrated that children of this age are not nearly as likely to be overlooked in a census enumeration as younger children. It is easier to obtain from their parents at least a nearly accurate estimate of age for children in this group than to get accurate reports on older groups. Any gross mis-statement of the age of a child in this group would ordinarily be apparent if the child were in sight, and there is usually no such obvious motive for mis-stating the age as may apply, for example, to a youth just under the legal age for employment or for marriage.

---

[16] This rough series suffices at the present stage; eventual "pro-rating" of the resulting estimates will result in a somewhat different series of ratios.

[17] See *The Population of Central America (including Mexico), 1950-1980*, ST/SOA/Ser. A, Population Studies No. 16. At the time when the calculations were made, the age statistics of the Mexican census of 1950 had not yet become available; hence, data of the 1940 census had to be used. Later, when the detailed statistics for 1950 were received, they were compared with the estimates derived from the data for 1940 and found to be quite consistent. The discrepancies between 1950 census data and estimates based on 1940 census data could be quite satisfactorily explained in terms of inaccurate age reporting at the 1950 census.

[18] Vital statistics of this disturbed period have, unfortunately, not become available.

89. For the development of certain estimates, especially those pertaining to fertility, it is often indispensable to assume that at least the data for one age group have relatively high accuracy, so that other quantities can be estimated by reference to these data. In such cases, it may be well to choose the figures for the 5-9 group.

90. In cases of extreme inaccuracy of age statistics, however, the accuracy of data for the 5-9 group is also subject to considerable doubt. In such cases, the reports at all age levels are usually strongly affected by attraction of even numbers (e.g., ages ending in 6, 8 and 10), and by the fact that the question relating to age in "completed years" is often misunderstood as referring to age "at the nearest (possibly the next-following) birthday". Allowance can be made for these influences by using as an estimate for the 5-9 group, the sum of one-half the numbers reported at ages 5 and 10, plus the reported numbers at ages 6, 7, 8 and 9.

4. *Ages 75 and over*

91. Owing to the tendency of older persons to exaggerate their ages, the numbers of persons reported as aged 75 and over are excessive in most censuses and greatly in excess in some. To estimate the correct numbers for the advanced ages would in any case be difficult, whether by extrapolation or by reconstruction from birth and death statistics pertaining to a distant period of the past.

92. For the purpose of a population projection, errors in numbers at the most advanced ages are only of small importance, because the numbers in these cohorts are rapidly depleted by deaths. In the population projections so far undertaken by the United Nations staff, no adjustments have been attempted with respect to persons aged 75 years or more though their numbers were admittedly quite inaccurate in most instances. However, in cases where the data are patently inaccurate, more reasonable figures may be substituted with the aid of observations derived from stationary populations.[19]

93. Table 7 shows, for a model stationary population, the percentages in each of the advanced age groups corresponding to a given percentage of total population at ages 70 and over. Given a certain percentage aged 70-74, corresponding percentages aged 75-79, 80-84 and 85 and over can be derived by interpolation.

94. The Dominican Republic census showed 1,070,742 males and 1,065,130 females, and our adjusted estimates for the age group 70-74 were 7,007 males, i.e., 0.654 per cent of all males, and 7,743 females, i.e., 0.73 per cent of all females. In accordance with the table, only about 0.3 per cent of males should be aged 75-79, 0.1 per cent 80-84, and 0.01 per cent aged 85 and over; of females, about 0.34 per cent should be aged 75-79, 0.11 per cent 80-84, and 0.02 per cent 85 and over. Accordingly, the following figures for these age groups may be substituted for the reported numbers:

Age (years)	*Males*		*Females*	
	Reported	*Estimated*	*Reported*	*Estimated*
70–74.....	8,176	7,007	9,628	7,743
75–79.....	3,941	3,210	4,073	3,621
80–84.....	3,697	1,070	4,725	1,172
85 and over...	3,711	107	4,979	213

TABLE 7. PERCENTAGES OF PERSONS AT ADVANCED AGES IN STATIONARY POPULATIONS[a]

Percentage of total population aged 70 and over	*Estimated percentage of total population in specified age group*				
	70–74	*75–79*	*80–84*	*85 and over*	*75 and over*
1.0.....	0.62	0.28	0.09	0.01	0.38
1.5.....	0.90	0.43	0.14	0.03	0.60
2.0.....	1.16	0.58	0.21	0.05	0.84
2.5.....	1.41	0.73	0.29	0.07	1.09
3.0.....	1.64	0.89	0.37	0.10	1.36
3.5.....	1.86	1.05	0.45	0.14	1.64
4.0.....	2.08	1.20	0.54	0.18	1.92
4.5.....	2.28	1.36	0.63	0.23	2.22
5.0.....	2.48	1.51	0.73	0.28	2.52

[a] These stationary populations are interpolated from the stationary populations corresponding to the model life tables used for the projections (see following chapter and appendix).

Numbers at advanced ages, though estimated only very roughly, are evidently much smaller than reported numbers.

5. *Final adjustment of estimates*

95. The various adjustments by procedures described in this section result in figures by age groups which, if added up, do not agree with the initial figure of total population. Since it is usually expedient to maintain consistency with published figures on total population, "pro-rating" is indicated.

96. It is unnecessary to "pro-rate" after each stage of the calculations since "pro-rating" at the final stage usually has practically the same effect. It is usually sufficient to "pro-rate" only once, after the last adjustment has been carried out.

[19] Though actual populations are not identical with stationary ones, the percentages of total population at the most advanced ages in fact often conform reasonably well with those of stationary populations. The assumption is admittedly imperfect but suffices to effect a considerable improvement on inaccurate data. Better approximations may sometimes be obtained by means of stable population models.

C. Adjustment of data not presented by conventional five-year age groups

97. For the purpose of the methods of projection discussed here, it is necessary that the population be classified by five-year age groups up to the age of 85. If the statistics are not presented in this form, adjustments are required.

98. In many cases, although a suitable classification by age for most of the population is given, the ages of some individuals remain unreported. Statistics are sometimes presented only by ten-year age groups for part or all of the age scale. Furthermore, the age groups may be irregular and may not fit into the required five-year, or even ten-year groups. Finally, in some cases only a few broad age groups are distinguished and these are sometimes identified only by vague criteria such as "infants", "children", "adults", etc. In the first three instances, adjustments are possible as described in the following pages. The last case can be dealt with by special methods which will be explained in another publication.

1. *Persons whose ages were not reported*

99. The persons whose ages are not reported in a census are usually only a small minority. This does not imply that the ages of other persons, which were reported, were always well ascertained. "Pro-rating" of the population of reported ages to total population is usually quite sufficient. The "pro-rating" should be done for each sex separately.

100. In some censuses ages are not ascertained or not tabulated in detail in respect of a certain segment of the population; for example, aboriginal tribes, other minorities, or even majority groups, the enumeration of which involves special difficulties. If the groups not classified by age are very small, no problem is presented; "pro-rating" or any reasonable estimate of the age distribution of these groups will be adequate. Where they represent a substantial fraction of the population, care must be taken to make the best possible estimates for them. It cannot ordinarily be assumed that such a population segment resembles the remainder of the population in its age structure. Estimates may be made by such methods as are described in part D of this chapter, for use where adequate statistics on age compositions are not available.

2. *The halving of ten-year age groups*

101. To reduce data presented by ten-year age groups to the required five-year groups, Newton's formula can be used:

$$f_{na} = 1/2\ [f_n + 1/8\ (f_{n-1} - f_{n+1})]$$

where f_{na} is the number in the first half of the given 10-year group, to be computed, f_n is the given number in the entire 10-year group, and f_{n-1} and f_{n+1} are numbers in the preceding and following 10-year groups.

102. As an example in the application of this formula, we may consider the official mid-year 1953 estimate of the Pacific Islands Trust Territory, which was given by sex and age, as follows:

Age (years)	Population	
	Male	Female
Total	25,956	24,929
0–4	4,640	4,310
5–9	2,245	2,148
10–14	2,093	1,978
15–19	2,015	2,037
20–24	2,025	2,026
25–34	3,233	3,124
35–44	2,651	2,637
45–54	2,209	2,155
55–64	1,758	1,605
65–74	1,126	1,024
75 and over	606	521
Age unknown	1,355	1,364

103. For the number of males aged 25-29, Newton's formula gives:

$$f_{na} = 1/2\ [3{,}233 + 1/8\ (4{,}040 - 2{,}651] = 1{,}703.$$

The number of males aged 30-34 is obtained by subtracting 1,703 from 3,233, i.e., 1,530. This method makes it possible to construct the estimates for five-year age groups up to age 64. To halve the 65-74 group, estimates for age groups over 75 years of age are required. These can be obtained by reference to table 7 above. In this example, 2.46 per cent of males and 2.21 per cent of females of known age are aged 75 years and over. By interpolation, about 1.48 per cent of males should be aged 75-79, 0.71 per cent 80-84, and 0.27 per cent 85 and over; 1.36 per cent of females should be aged 75-79, 0.63 per cent 80-84, and 0.23 per cent 85 and over. The corresponding numbers are:

Age (years)	Males	Females
75–79	364	320
80–84	175	148
85 and over	66	54

104. It is now possible to enter an estimate for the 75-84 group in Newton's formula and thus to halve the 65-74 group. The following results are obtained for the five-year age groups from 25 to 74:[20]

Age (years)	Males	Females
25–29	1,703	1,651
30–34	1,530	1,473
35–39	1,390	1,379
40–44	1,261	1,258
45–49	1,160	1,142
50–54	1,049	1,013
55–59	947	873
60–64	811	732
65–69	639	583
70–74	487	441

[20] In the application of formulas using several successive age groups, one must bear in mind that such formulas cannot be used where there have been past irregularities in the annual numbers of births. The number of the people affected by these irregularities have to be replaced by proper numbers before applying the formulas. See for example what is said about the population of Mexico in part B of this chapter.

105. In view of the large number of persons whose ages were not reported, it is important in this case to take account of any clues which may be available concerning possible peculiarities of the age distribution of this group. In case it can be assumed that their age distribution was similar to that of the remainder of the population, the estimates for all age groups should be "pro-rated" to the total numbers of each sex.

3. *Interpolation for ages tabulated by irregular groupings*

106. A very rough method for dealing with statistics by age in irregular groups which require transformation into conventional five-year groups comprises the following steps. First, the number in each group is multiplied by a factor to make it equal roughly the number that would be found in a five-year group having the same central age. The numbers resulting from this calculation are interpolated, with reference to the central ages of adjacent groups, to obtain the numbers which may be expected in five-year groups centring on exact ages 2½, 7½, 12½, 17½, etc., which are central in the conventional grouping.

107. This method is rough because it must be assumed that ages are distributed, by single years, in a linear form. Such perfect distribution of ages cannot be assumed with regard to the entire age scale but may, in fact, be approximated within relatively short ranges of ages if the birth rates have not fluctuated greatly in the past. This procedure also tends to reduce somewhat the errors in the original statistics, though it does not eliminate them.

108. The following statistics of the 1930 census of El Salvador may be taken as an illustration:

Age (years)	Enumerated population		Exact central age (years)	Width of age group (years)
	Male	Female		
0–1	39,510	38,649	1	2
2–4	64,533	64,049	3	3
5–7	62,125	60,361	6.5	3
8–14	129,666	118,109	11.5	7
15–17	40,057	44,959	16.5	3
18–22	78,347	85,107	20.5	5
23–29	80,994	85,933	26.5	7
30–39	96,327	93,515	35	10

109. The age group 0-4 can be obtained directly by addition. It is not necessary to interpolate to obtain age group 35-39, since this will be obtained by subtraction from the total for 30-39, once the 30-34 group has been calculated.

110. Estimates for five-year age groups, having the same central ages as the given groups, are obtained by multiplying the figures of each given group by five and dividing by the number of years comprised in the original group. Five times the numbers in age groups 2-4, 5-7, and 15-17 are divided by three, and five times the numbers in age groups 8-14 and 23-29 are divided by seven. No multiplication is needed for the 18-22 group, which is already a five-year group. The numbers in the 30-39 group are divided by two to obtain an estimate of a five-year group centring around the exact age of 35 years. The results of the multiplications are as follows:

Central age (years)	Estimated number in corresponding five-year group	
	Males	Females
3.5	107,555	106,748
6.5	103,542	100,602
11.5	92,619	84,364
16.5	66,762	74,932
20.5	78,347	85,107
26.5	57,853	61,381
35	48,164	46,758

111. To interpolate these figures with respect to the central ages of conventional five-year groups, note must be taken of the distances, in years, of the latter central ages from the central ages in the five-year groups listed above. The interpolation is then carried out by multiplying each of the above figures in inverse proportion to its distance from the required central age. The procedure can be illustrated, as follows, with respect to males aged 5-9. The exact central age of this group is 7.5 years; that is, one year more than 6.5 and 4 years less than 11.5. The figure relating to central age 6.5 is multiplied by four-fifths, and that relating to central age 11.5 by one-fifth, and the two results are added. The result is 101,358, that is, the required estimate for the age group 5-9. When the other interpolations have been made by the same method, the distribution is obtained:

Age group (years)	Central age (years)	Estimated population	
		Male	Female
0–4	2.5	104,043	102,698
5–9	7.5	101,358	97,355
10–14	12.5	87,447	82,477
15–19	17.5	69,658	77,476
20–24	22.5	71,516	77,198
25–29	27.5	56,713	59,661
30–34	32.5	51,014	51,059
35–39	37.5	45,313	42,456

112. The results are not entirely satisfactory. In particular, there appear to be relatively too few females aged 10-14 and too few males aged 15-19. This is probably a reflection of inaccuracies in the original data. Before using these results, it might be desirable to apply some of the corrections and adjustments discussed in part B of this chapter.

D. Estimation of age distribution of a population for which statistics are not available

113. A different order of estimation is required to obtain the basis for population projections where no census has been taken, or the latest census figures are too old to be brought forward to a suitable starting point, or where the census provides no age classification, or a classification that is too inaccurate or too general in its groupings to serve as a basis for the estimates. In such situations any estimate of the age distribution will necessarily be questionable, but estimates may nevertheless be made, in many such cases, with sufficient accuracy to provide the basis for usable projections.

1. *Analogy with another similar population*

114. If the statistics of country A are insufficient for an estimate of the population by sex and age, while those for country B are adequate, and if it can be assumed that population trends in both countries have been

similar, the age-distribution of country B can be "pro-rated" to the population total of country A. Such a procedure is not justified, of course, unless there is firm ground for believing that the two age structures in fact resemble each other closely. An analogous procedure can be used with more assurance if sex-age data are available for a part of the country which can be presumed to be fairly representative of the whole; in that case, the statistics for the given part are "pro-rated" to an estimate of the country's total population.

115. An estimate of sex-age structure can also be arrived at if the population trends of country B are similar in some but not all respects to those of country A. An example is afforded by the population projection for Uruguay made by the United Nations staff.[21]

116. The most recent official sex-age statistics for Uruguay which could be found were those of the census of 1908, but a base estimate was required for the year 1950. Although fairly reliable vital statistics were available for most of the intervening years, it was not considered possible to construct a realistic sex-age estimate for 1950 by carrying the 1908 data forward over a period of 42 years, especially since migration—for which statistics were not adequate—had been of some importance.

117. Comparison of Uruguayan and Argentine statistics for the half-century since 1900 brought out the following facts:

(*a*) Birth rates were always appreciably lower in Uruguay than in Argentina; in fact, in most years the birth rate of Uruguay was very nearly five-sixths of that of Argentina, despite considerable changes in the rates of both countries.

(*b*) Death rates in Uruguay were at nearly all times very slightly lower than those in Argentina. The rates in both countries declined fairly regularly and about equally rapidly, those of Argentina apparently following the Uruguayan rates with a time-lag of a few years.

(*c*) International migration appeared to have affected the population of Uruguay in approximately the same proportions as that of Argentina, except that in the last decade before 1950 the net effect of migration in Uruguay seemed to have been negligible. The migration statistics were not very accurate but an error in the assumption of equal proportionate effects on the two populations would presumably introduce only a comparatively small error in the resulting population estimates.

118. In Argentina, censuses were taken in 1914 and in 1947. In view of the slight time-lag in the respective mortality rates, and the similarity of the effects of migration, the combined effects of deaths and migration in Argentina, from 1914 to 1947, were regarded as comparable to those on the population of Uruguay in a period of equal length (33 years) beginning with the census of 1908 and ending in the year 1941. Since the 33-year period is not a multiple of five years, adjustments, by interpolation between adjacent age groups, were carried out on the five-year age data of the Argentine (1914) and Uruguayan (1908) censuses so that the age groups would correspond to those of cohorts whose ages would fall within the conventional five-year groups by the end of the 33-year period.

119. Comparison of identical cohorts in the Argentine censuses of 1914 and 1947 permitted the computation of "survival-*cum*-migration ratios" (i.e., the ratio of survival, as modified by the effects of immigration and emigration) over the 33-year period for persons aged 35 years and over in 1947. "Survival-*cum*-migration ratios" were also computed, for periods of varying length, for five-year cohorts of births which occurred from 1917 to 1947, aged under 30 years in 1947. One cohort was constituted partly from small children at the 1914 census and partly from births during 1914-1917 (aged 30-34 years in 1947), and the same type of ratio was computed for this cohort. The ratios were then applied to the corresponding Uruguayan census data of 1908 and birth statistics of 1908-1941 to obtain a first estimate of the Uruguayan population, by sex and age, in 1941, which was further "pro-rated" to accord with the official estimate of total population for the same year.

120. This estimate for Uruguay in 1941 was then brought forward, first to 1946, and then to 1951, by means of Uruguayan statistics on births and on deaths by sex and age, with suitable interpolations and extrapolations for those years for which the requisite vital statistics had not yet become available. Interpolation between the resulting estimates for 1946 and 1951 yielded the required base estimate for the projection of the population from 1950 onward.

121. It is recognized that this estimating procedure may have involved errors and that high accuracy could not be claimed for the results. Nevertheless, the approximation is believed to be reasonable and not without value, no other estimate being available for any date since 1908.

122. The same degree of similarity in population trends might perhaps not be found between any other two countries, but the example cited here may be helpful in suggesting devices by which estimates can be derived in different situations.

2. *The method of stable populations*

123. An approximate sex-age distribution can also be constructed for a population for which a few demographic indices are known, by applying the theory of stable populations. As Lotka has demonstrated,[22] a population with constant birth rates and constant mortality rates will exhibit constant proportions of numbers in the different sex-age groups, which are determined by the birth and mortality rates. It has been further demonstrated[23] that changes in mortality, unless they are very radical, have only slight effects on this stable age structure. Hence, an estimate of the birth rate, combined with a very rough estimate of the death rate, suffices to construct an approximate age distribution if it can be assumed that the birth rate has been nearly constant for a long period. In this connexion, the model life tables described in chapter III are very helpful. The use of stable populations will be described more fully in a subsequent publication.

[21] United Nations, *The Population of South America, 1950-1980.*

[22] Alfred T. Lotka, *Théorie analytique des associations biologiques,* Paris, Hermann, 1934-1939.

[23] *Population Bulletin of the United Nations,* No. 4, December 1954. "The cause of ageing of populations: declining mortality or declining fertility?"

E. The example of Costa Rica

124. A population projection for Costa Rica was made taking the year 1950, when the last census was taken, as a starting point.[24] Since the time when that projection was made, more recent information has become available. The latest official estimate of total population now at hand relates to mid-year 1955 and amounts to 951,093. This figure can now be taken as the basis for a new projection, with 1955 as the starting date. This recent official estimate is well in line with the results of earlier projection, according to which, assuming constant high fertility, the population would have increased from 804,900 individuals in mid-year 1950 to 945,900 in mid-year 1955. A slight revision of the earlier projection, as illustrated in chapter VII, might suffice. Nevertheless, in order to illustrate the procedures discussed in this chapter, the derivation of base estimates for 1955, which would be the starting point for a new projection, is explained here.

125. In bringing the census statistics forward from 1950 to 1955, the procedure utilizing data on births and on deaths by age might be employed. The interval is five years and the required statistics are available. A study of Costa Rican statistics,[25] however, has shown that while death registration is presumably fairly accurate, the accuracy of birth statistics is affected by frequent delays in the registration of births. The latter finding is also reflected in the series of annual recorded birth rates. The actual birth rates were probably fairly constant up to 1950, as witnessed by the following five-year averages:

1925-29	46.2
1930-34	44.6
1935-39	44.5
1940-44	43.7
1945-49	43.7
1950-54	51.1

For the individual years from 1950 to 1954, the following series was obtained:

1950	46.5
1951	47.6
1952	54.8
1953	53.9
1954	52.6

The series is difficult to interpret unless it is assumed that, especially in 1952 and 1953, there was a catching-up in the registration of births that had occurred in earlier years. There is little plausible reason to suppose that the actual birth rates should have fluctuated in this manner.

126. The course of the recorded crude death rates, by contrast, appears altogether plausible. The death rates declined, with increased momentum after 1940, as follows:

1925-29	23.2
1930-34	22.3
1935-39	20.0
1940-44	18.1
1945-49	14.0
1950	12.2
1951	11.7
1952	11.6
1953	11.7
1954	10.6

127. In the census of Costa Rica, as in the censuses of most countries, there is reason to doubt the completeness of enumeration of small children. In 1950, the Costa Rican census reported 30,261 children aged less than one year. The average annual number of births registered in 1949 and 1950[26] amounted to 35,399. If this had been the true number of births in the 12 months preceding the census, there would have had to be more than 5,000 infant deaths to result in 30,261 surviving infants, not counting the deaths of children who, by the time of the census, would have been more than 12 months old. If the latter were included, the total number of infant deaths would have had to be more than 6,000, possibly 7,000. However, on the average of 1949 and 1950, only 3,352 infant deaths were registered annually.

128. Similar rough computations can also be performed with respect to births during 1945-49, the deaths among them, and the number of children aged 1-4 in 1950. The 102,374 children of this age enumerated in the census would have been the survivors of about 129,000 births, if enumeration and registration had been exact. However, according to the statistics, only about 17,000 deaths followed those births. The conclusion is not definitive, since birth registration may have been inaccurate and since this calculation is not precise, but there is a strong indication that the enumeration of infants and small children in 1950 was quite deficient.

129. In bringing the statistics of the 1950 census forward to 1955, therefore, it is preferable not to place much reliance either on numbers of recorded births or on the number of children aged less than 5 enumerated in 1950. Instead, constant fertility in the recent period may be assumed and numbers of children aged under 5 in 1950 and those aged under 5 and 5-9 in 1955, may thus be worked out separately by the methods presented in chapter V.

130. The procedure, thus reduced to cohorts aged 5 and over in 1950, is illustrated in tables 8 and 9. The calculation is slightly inexact, since the Costa Rican census was taken on 22 May 1950 and the estimate desired is for mid-year 1955, whereas the statistics of deaths refer to the calendar years from the beginning of 1950 to the end of 1954. Most of the resulting error, however, will be eliminated when the results of the computation are "pro-rated" at a later stage. Since the data thus brought forward exclude children under 10 in 1955, "pro-rating" at this stage is not possible. Nor is it necessary, since no important errors are introduced by the possible inaccuracies in death registration or the exclusion from the calculation of almost negligible numbers of persons and deaths with undeclared ages. A definite estimate of the 1955 population by sex and age will be arrived at after use of the fertility assumption developed in chapter V.

[24] United Nations, *The Population of Central America (including Mexico), 1950-1980.*

[25] Ricardo Jiménez, "Causa de la diferencia entre el censo de población de 1950 y la estimación a esa fecha en Costa Rica", *Estadística*, vol. X, June 1952.

[26] The recorded birth rates of those two years averaged 45.0, which is plausible.

TABLE 8. DEATHS IN COSTA RICA, 1950–1954, ATTRIBUTABLE TO COHORTS OF GIVEN AGES ON 1 JANUARY 1950

Sex and age group (years)	Deaths reported in this age group					Total deaths, 1950–54	Deaths attributable to cohort initially in the given age group [c]
	1950	1951 [a]	1952	1953	1954 [b]		
Males							
5– 9	151	162	139	205	175	832	594
10–14	80	72	67	68	68	355	364
15–19	80	70	91	63	69	373	448
20–24	105	126	114	88	91	524	478
25–29	100	100	99	68	66	433	460
30–34	109	106	93	97	81	486	530
35–39	148	122	109	99	96	574	557
40–44	117	109	105	96	113	540	574
45–49	158	130	82	128	111	609	704
50–54	161	173	163	152	151	800	764
55–59	152	171	112	138	155	728	859
60–64	237	189	206	179	179	990	990
65–69	190	179	210	205	205	989	1,031
70–74	235	212	224	209	193	1,073	980
75–79	178	183	152	198	177	888	763
80–84	132	119	135	133	119	638	1,046
85 and over	155	136	135	159	142	727	
Females							
5– 9	148	158	133	137	156	732	502
10–14	63	56	51	49	52	271	310
15–19	83	73	68	62	64	350	429
20–24	107	129	88	88	96	508	492
25–29	107	107	107	76	79	476	470
30–34	104	102	97	91	70	464	541
35–39	160	131	99	126	102	618	598
40–44	123	115	118	106	115	577	536
45–49	114	93	94	103	91	495	564
50–54	127	137	125	122	122	633	645
55–59	126	142	122	137	130	657	770
60–64	204	163	182	171	163	883	880
65–69	174	163	185	194	163	879	960
70–74	220	198	208	228	187	1,041	904
75–79	154	158	143	167	145	767	737
80–84	149	134	147	148	129	707	1,280
85 and over	172	151	195	218	190	926	

[a] Except for infants, death statistics by age for 1951 were available for both sexes only; they were estimated for each sex assuming the same sex-ratios of deaths by age as in 1950.

[b] The available statistics for 1954 provided no break-down for ages 75 and over; it was assumed that deaths occurred at ages 75–79, 80–84, and 85 and over, in the proportions of 1953.

[c] Arithmetic mean of totals of each pair of successive age groups.

TABLE 9. UTILIZATION OF STATISTICS ON DEATHS IN BRINGING DATA ON COSTA RICAN POPULATION AGED 5 AND OVER IN 1950 FORWARD TO 1955

Sex and age of cohort in 1950 (years)	*Population enumerated, 1950* [a]	*Same, smoothed* [b]	*Deaths attributed to cohort*	*Estimated population, 1955*	*Age of cohort in 1955 (years)*
Males					
5– 9.....	56,789	—	594	56,195	10–14
10–14.....	49,734	48,814	364	48,450	15–19
15–19.....	40,418	41,773	448	41,325	20–24
20–24.....	37,671	36,212	478	35,734	25–29
25–29.....	28,647	29,327	460	28,867	30–34
30–34.....	23,851	24,297	530	23,767	35–39
35–39.....	22,908	22,184	557	21,627	40–44
40–44.....	18,310	18,446	574	17,872	45–49
45–49.....	14,140	14,568	704	13,864	50–54
50–54.....	12,313	11,579	764	10,815	55–59
55–59.....	7,889	8,747	859	7,888	60–64
60–64.....	7,667	6,965	990	5,975	65–69
65–69.....	4,716	5,088	1,031	4,057	70–74
70–74.....	3,331	3,180	980	2,200	75–79
75–79.....	1,860	—	763	1,097	80–84
80–84.....	1,073	—	1,046	746	85 and over
85 and over...	719	—			
Females					
5– 9.....	55,367	—	502	54,865	10–14
10–14.....	48,555	48,611	310	48,301	15–19
15–19.....	43,826	44,010	429	43,581	20–24
20–24.....	39,386	38,679	492	38,187	25–29
25–29.....	30,491	30,595	470	30,125	30–34
30–34.....	23,705	24,830	541	24,289	35–39
35–39.....	23,930	22,622	598	22,024	40–44
40–44.....	18,074	18,548	536	18,012	45–49
45–49.....	13,966	14,226	564	13,662	50–54
50–54.....	11,853	11,274	645	10,629	55–59
55–59.....	7,827	8,518	770	7,748	60–64
60–64.....	7,248	6,649	880	5,769	65–69
65–69.....	4,420	4,782	960	3,822	70–74
70–74.....	3,227	3,039	904	2,135	75–79
75–79.....	1,768	—	737	1,031	80–84
80–84.....	1,147	—	1,280	707	85 and over
85 and over...	840				

[a] Exclusive of individuals whose ages were not reported.

[b] By smoothing formula presented in p. 11. Smoothing appears desirable owing to irregularities probably due to inaccurate age reporting, such as the apparent excess of women aged 35–39 over women aged 30–34.

III. SOME CONCEPTS OF ABRIDGED LIFE TABLES AND LIFE-TABLE CONSTRUCTION

131. The core of the methods of population projections described here consists in the use of a system of model life tables to calculate the numbers of persons in each cohort of a population who will survive during successive time intervals in the future. The derivation of these model tables is briefly described in the appendix, where the survival ratios, needed for the population projections, as well as other functions of the life tables are tabulated.

132. Though the actual population projection may be carried out entirely by the tabulated survival ratios, it is first necessary to establish which of these ratios are appropriate to the given case, i.e., which ratios express most nearly the current and expected future mortality conditions. Survival ratios are not given directly in official statistics but must be calculated or otherwise derived by reference to the statistical information which exists. In almost every case, at least one step in the