CHAPTER III. THE ACCURACY OF AGE AND SEX STATISTICS

This chapter is concerned primarily with methods of testing the accuracy of population statistics classified by age groups. Tests of the accuracy of age statistics are necessary not only because such data are of major importance for population estimates and demographic analysis, but also because errors in these statistics are often indicative of deficiencies in the head count of the population, or in the records of the numbers of vital events. Statistics classified by age groups may be affected both by errors in the reporting of ages and by variations in the completeness of enumeration, or of recording of vital events, for the different age groups. Both types of error will be considered here.

The classification by sex is also of fundamental importance in demographic statistics. Sex, unlike age, is seldom reported incorrectly in census enumerations or records of vital events. It is therefore not necessary here to develop tests for the accuracy of sex reporting. However, statistics classified by sex may be in error because the enumeration, or the recording of events, is more nearly complete for one sex than for the other. Analysis of statistics by sex, like the tests of age statistics, may therefore be useful in evaluating the reliability of head counts and the completeness of vital statistics registration. In this chapter, certain tests applicable to population data classified by sex are presented together with the tests of age statistics. In addition, the methods explained in the preceding chapters for testing census and vital statistics for both sexes combined, can be applied to data for each sex separately.[1]

A. Evaluation of census results classified by sex only, or by sex and broad age groups

In some under-developed countries where a great majority of the population is illiterate, it is very difficult to obtain even approximately correct statements, in a census, of the ages of the people. In such situations, the census returns are sometimes classified by sex only, or for each sex by such indefinite age groupings as "infants", "children", and "adults". Although it is impossible to apply to data of this kind such rigorous tests as are possible where the exact ages in years are recorded, it is useful to examine the consistency of returns for the two sexes and for whatever broad age categories are presented. Such an examination may provide useful indications of apparent deficiencies in the head count.

Testing procedures of the types described below are especially appropriate where there is any *a priori* reason to suspect—in view of the prevailing customs and attitude of the people or the methods of enumeration employed—that there may have been a tendency toward an error in the count affecting especially persons of one sex or age category. If this is the case, the results of the tests should be interpreted with full regard to such information concerning factors tending to produce an error. However, the application of these tests is advisable even where no *a priori* suspicion of error exists, since they may bring to light errors that have not been suspected.

A first test consists in the computation of sex ratios on the basis of the returns for the whole country and each of its parts. The sex ratio can be expressed as the number of males per 100 females, or females per 100 males. If broad age categories are also given, the sex ratios can usefully be computed separately for each of these categories. Computation of the ratio of children to women is another useful test.

Such ratios, computed for each area, should be compared and evaluated in the light of what is known about conditions in various parts of the country. In general, a population not much affected by migration has approximately (although hardly ever exactly) equal numbers of males and females. This is to be expected because, as a rule, slightly more males than females are born,[2] but female mortality is less than that of males; consequently, a slight excess of males among children is usually more or less counter-balanced by an excess of females among adults. Depending on the precise sex ratio of births and mortality conditions, the total number of females may or may not slightly exceed the total number of males.

This situation, however, is often modified in various ways. If the population is very small, the ratio of males to females is subject to chance variations. Following a war, with many military casualties, the ratio of males to females in the adult population may be abnormally low. Large-scale migration can have a great effect on sex ratios. Very often a majority of the migrants are young men in search of gainful employment, so that the relative numbers of males are reduced in areas of emigration, and increased in areas of immigration. Predominantly female migratory movements also occur, as in the case where many young rural women move to towns where they seek employment as domestic servants. Some knowledge of the types of movements and employment conditions in the country will enable the statistician to appraise the sex-ratios computed from the census data.

The following example, based on Canadian statistics, is given as an illustration of the variations in sex ratios which may in fact occur in different parts of a country. In this example, the ratios computed from the census figures give no basis for suspecting important errors in enumeration.

At the 1951 census of Canada, 7,089,000 males and 6,921,000 females were enumerated. For every 100 females, there were 102.4 males. Since Canada is an immigration country, the excess of males is not surprising. Among the rural population, the ratio was 112.3 males per 100 females; among the urban population, 95.8 males per 100 females. This difference

[1] See also the discussion of tests of vital statistics registration by reference to sex ratios and age patterns in chapter II.

[2] See chapter II, section B.

is related to the differences in employment conditions for the two sexes in the countryside and in towns.

For the rural and urban populations of the various provinces, the ratios are as shown in table 5 below.

Table 5

MALES PER 100 FEMALES IN RURAL AND URBAN AREAS OF THE CANADIAN PROVINCES IN 1951

Province	Sex ratio (males per 100 females) Rural	Urban
Newfoundland	111	98
Prince Edward Island	111	86
Nova Scotia	110	96
New Brunswick	109	91
Quebec	109	95
Ontario	115	96
Manitoba	117	94
Saskatchewan	117	94
Alberta	121	99
British Columbia	121	98
Yukon Territory	164	120
North West Territories	122	184

Conditions in the Yukon Territory and the North West Territories are special; these are located in high northern latitudes and are very sparsely inhabited, by a largely migrant population. In the remaining provinces the sex ratios show a male excess in rural areas and a male deficit in the cities. In Alberta and British Columbia, where the rural excess of males is particularly great, the urban deficit of males is almost negligible. The provinces are listed in their geographic order, from East to West, showing that the rural excess of males diminishes slightly in the progression from Newfoundland to Quebec, and then increases continuously towards the West. A similar, but less regular pattern of variation appears also in the urban sex ratios. This pattern is undoubtedly related to the economic features and demographic history of the country.

A comparison of sex ratios can also be made for communities of various size classes. For the entire country, the following sex ratios are obtained (males per 100 females):

94 in urban places of 100,000 inhabitants or more;
96 in urban places of 30,000 to 99,999 inhabitants;
96 in urban places of 10,000 to 29,999 inhabitants;
99 in urban places of less than 10,000 inhabitants;
110 in rural non-farm areas;
117 in rural farm areas.

The sex ratio rises regularly in progression from the most urban to the most intensely rural categories. Analogous ratios for each of the provinces (not presented here) show the same systematic sequence, with only a few minor deviations, as, for example, in the case of Nova Scotia, where the sex ratio for urban places with 10,000 to 29,999 inhabitants was found to be 97, but that for urban places with less than 10,000 inhabitants was only 94. Perhaps some special reason exists to account for this deviant observation.

In other countries, conditions may differ greatly from those noted in the case of Canada. Where there is a predominantly male migration from rural to urban areas, the pattern observed in Canada may be altogether reversed. The presence of mining enterprises, or of plantations employing seasonal labor, for example, may result in selective migrations reflected in the sex ratios. Whatever the special conditions of the case may be, if sex ratios disclose a systematic picture, then there is reason to suppose that the two sexes have been enumerated with comparable accuracy. Deviations from the pattern, which cannot readily be explained, should be singled out for further investigation of the accuracy of the returns. If the entire pattern of sex ratios is erratic there are strong grounds for distrusting the completeness of the enumeration.

If the census also provides figures on "adults" and "children", various other ratios can be computed, such as ratios of children to adults, children to women, women to men, and girls to boys. An example of an analysis of this kind, based on the 1931 census in the Northern Provinces of Nigeria, is reproduced below:[3]

According to the 1931 "census" returns there were 3,497,807 men, 3,898,114 women, and 4,037,178 children. The proportion of children was 37.1 per cent. It oscillated between 29.2 per cent in Niger Province and 41.5 per cent in Zaria Province. The ratio of women to 100 men was 111.4, varying between 92.7 in Plateau Province and 129.5 in Bornu Province. The ratio of children to 100 women was 103.6, varying between 78.4 in Adamawa and 130.2 in Zaria. Large as these differences between Provinces may seem, the differences between the returns for the various Divisions of one and the same Province are in many cases not less startling. I shall quote a few striking examples:[a]

Province	Division	Men	Women	Children	Children per cent	Women to 100 men	Children to 100 women
Kabba	Igala	73,096	80,184	72,232	32.0	109.6	90.8
	Igbira	30,217	42,263	76,196	51.2	139.8	180.3
	Kabba	15,219	22,440	19,557	34.2	147.4	87.2
	Koton Karifi	12,339	13,664	5,319	17.0	110.7	37.5
Niger	Abuja	23,634	26,912	23,870	32.1	113.8	88.7
	Agaie-Lapai	20,148	23,915	12,077	21.5	118.6	50.5
	Bida	57,759	66,546	47,653	27.7	115.2	71.6
	Kontagora	22,163	23,046	23,970	34.7	103.9	104.0
	Kuta	23,548	21,300	19,836	30.7	90.4	93.1
	Zungeru	12,958	13,176	10,556	28.8	101.6	80.1
Plateau	Jemaa	22,436	24,103	29,780	39.0	107.4	123.6
	Jos	57,282	41,330	33,685	25.5	72.1	81.5
	Pankshin	58,149	54,375	48,228	30.0	93.5	88.7
	Shendam	35,859	39,632	42,811	36.2	110.5	108.0
	Southern	28,969	28,459	23,640	29.2	98.2	83.1
Zaria	Katsina	219,037	284,457	422,354	45.6	129.8	148.5
	Zaria	142,279	148,266	141,016	32.7	104.2	95.1

[a] See *Census of Nigeria, 1931*, vol. ii, pp. 33, 52-3. Figures comprise total population.

[3] From R. R. Kuczynski, *Demographic Survey of the British Colonial Empire*, Vol. I, West Africa (1948), p. 608.

The Census Officer made the following comment:

One knows of no such conditions in Kabba Province as would account for the enormous disproportions of non-adults shown in the Division of Koton Karifi . . . and Igbira . . . or for the low figure . . . in Agaie-Lapai Division of Niger Province. In Jos Division, however, where there is a large immigrant adult male population to the mining areas, one would expect to find a lower proportion of adult females and non-adults.[a]

As a matter of fact all ratios of sex and age derived from the General Censuses are open to grave doubts, since these "censuses", also in the Northern Provinces, were based largely on the tag registers and, as regards women and children, on rough estimates or guesses. In some Divisions, such as Katsina, where a large excess of females (30 per cent) was recorded for adults and a large deficiency of females among children (18 per cent), this anomaly may be due to the custom of counting even the youngest wives as adults.[b]

[a] See *Census of Nigeria*, 1931, vol. ii, p. 33.

[b] It should be noted, however, that the ratio of children to women was extraordinarily high in Katsina Division, and that if the number of women was swelled by counting girls as adults, the ratio of children to women would have been higher still than shown in the census returns.

It should be noted that in those censuses in which "adults" and "children" are distinguished without any precise definition of the age limits of these categories (as has to be done in some areas of Africa where people are often quite unaware of their ages, in years), it very frequently happens that females are reported as "adults" at an earlier age than males. Consequently, among the children, there would often be a heavy preponderance of males because adolescent boys aged, say, 15 or 16 would be regarded as "children", while adolescent girls of, say, 13 or 14 years would be regarded as "adults". This results also in an undue preponderance of females among the population considered as "adults".

B. Examination of detailed age classifications of the population at a single census date

For a proper interpretation of the results of tests applied to more detailed census statistics on ages of the population, it is important to bear in mind that errors in these statistics may be produced both by differences in the relative completeness of enumeration of persons in different age groups, and by mis-statement of the ages of those who are enumerated. The factors which may be responsible for both these types of error, under the conditions which exist in the given country, should also be taken into account.

The causes of mis-statement of ages include:

Ignorance of age, negligence in reckoning the precise age, deliberate mis-statement, and misunderstanding of the question.[4] If ages have been mis-stated mainly because of ignorance, the returns may nevertheless represent a fairly close approximation to the true distribution of ages. If ages have been deliberately mis-stated, there is more likelihood of a bias in the distribution towards either overstatement or understatement of age, affecting certain broad age ranges.

[4] The standard definition of age, used in most censuses and vital records, is the number of completed years, that is to say, the exact age, in years, at the last birthday. Local customs in reckoning age sometimes diverge from this definition. In some cases, the question refers to the date of birth instead of the age attained. It is believed by the authorities in many countries that more precise answers are obtained if the question is put in this form, provided, of course, that date of birth is well known to a majority of the population; nevertheless, mis-statements of date of birth analogous to mis-statement of age also occur. The censuses of Brazil require statement of date of birth if known and statement of age if birth date is not known.

Balancing equations, such as those already described with reference to the total population, can be applied separately to each sex-age cohort.[5] Failure of these equations to balance may be accounted for by mis-statement of ages, by varying completeness of enumeration or by both. It is often impossible to distinguish clearly the effects of these two kinds of error if one test only has been used. Hence the need to apply several different tests. Examination of the data from different points of view increases the likelihood of a correct interpretation.

The subject of the present chapter cannot be entirely detached from the topics dealt with in chapters I and II. Some tests of completeness of enumeration or vital registration presuppose that age statements are fairly accurate. Some tests of age-accuracy, on the other hand, presuppose reasonably complete enumeration or registration. Direct checks of completeness of a census or vital registers can be conveniently combined with direct checks on the accuracy of age statements, as well as statements relating to various other population attributes.

Some effects of age errors are most readily apparent if statistics by single years of age are examined; others can best be observed in grouped data, preferably the conventional groups of 5 years. The age distribution should therefore be tested in both forms. Essentially the same principles can be applied to single-year and grouped data. Complete examination, in either case, involves four steps, namely: (1) inspection of the data, (2) comparison with an expected configuration, (3) analysis of ratios computed from the data, and (4) measurement of age-accuracy by means of an index.

The use of these methods will be illustrated by a detailed analysis of the age distribution of the 1945 census of Turkey. These data have been selected not only because errors in the age statements were evidently considerable, but also because the frequency and magnitude of errors differ greatly for the two sexes. The types of errors present in the Turkish statistics are common to most statistics relating to age, though they are often of much smaller magnitude.

Inspection of the data

Table 6 reproduces the results of the 1945 census of Turkey, showing the population by sex and single years of age. The same data are presented in figure 1. It can be seen at a glance that unusually large numbers of individuals were reported at ages which are multiples of 5 and 10, and relatively small numbers at other ages.

Closer inspection of the figures reveals that deficiencies in reported numbers are greatest at ages with final digits 1 or 9; this is not surprising in view of the strong attraction exerted by the figures terminating in 0.

[5] An age-cohort is the number of persons surviving from the same birth-period. Thus, persons aged 35-39 years in 1940 and 45-49 years in 1950 belong to the same "cohort", born in the 1900-1905 period.

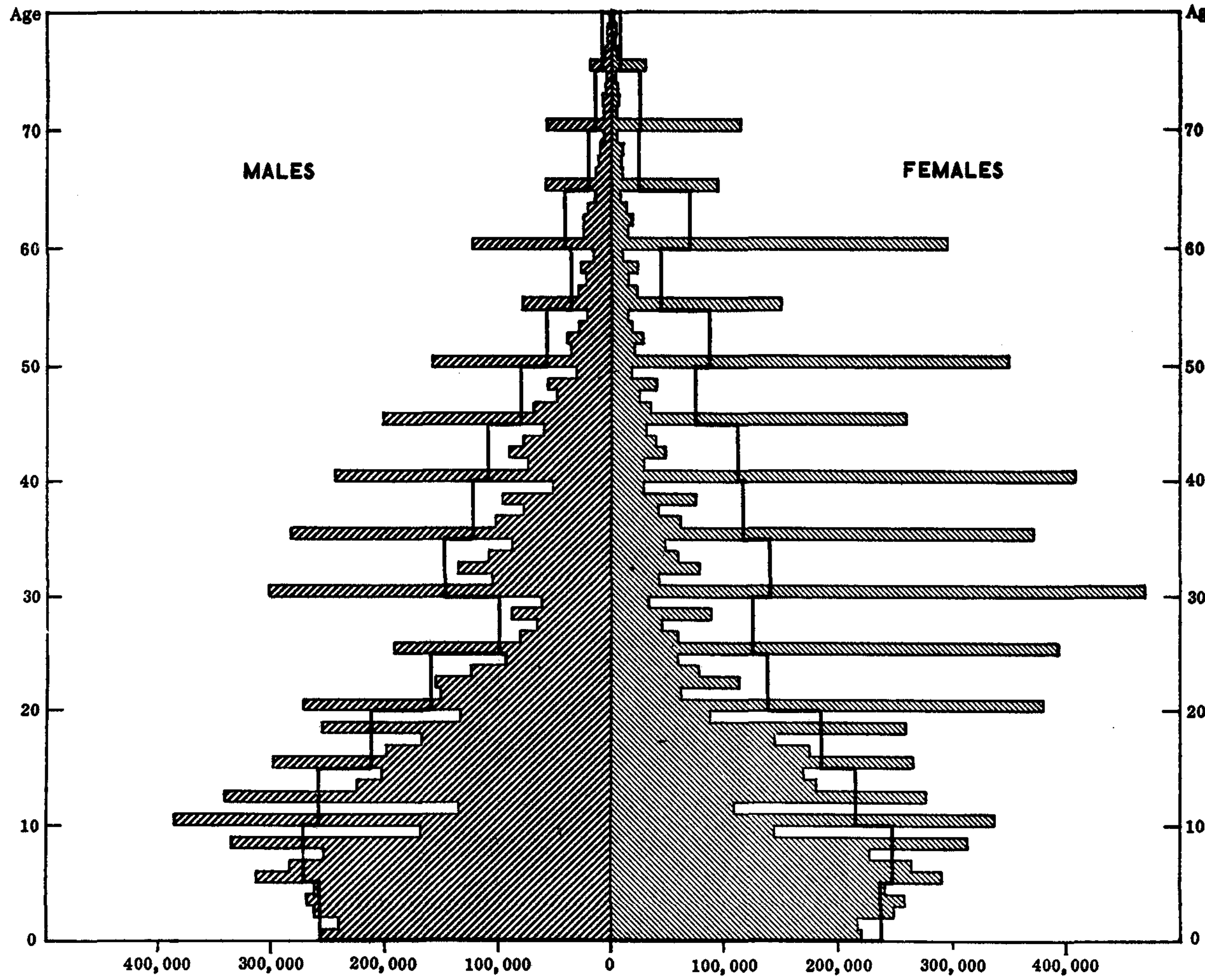

Figure 1. POPULATION OF TURKEY, 1945, BY SEX, BY SINGLE YEARS OF AGE AND 5-YEAR AGE GROUPS, ACCORDING TO CENSUS

Marked deficiencies at ages ending in 4 or 6 are likewise explained by the considerable attraction of the figures with terminal digit 5. As for ages ending in 2, 3, 7, and 8, it can be noted that the "even" figures are somewhat more attractive (or less repugnant) than the "odd" ones; numbers at ages ending in 2 are somewhat larger than those at ages ending in 3; numbers at ages ending in 8 are somewhat larger than those at ages ending in 7.

Finally, the forces of attraction of preferred final digits (or repulsion of disliked digits) are more marked in the case of females than of males, as can be seen from the wider fluctuations in the figures for females. Evidently, the age statements of females are less accurate.

These findings are common to inaccurately reported ages in almost any country. The intensity of attraction and repulsion of certain figures is not always as pronounced as in the present example, nor is the difference between male and female age-accuracy always as great. The details of the pattern of age mis-statements vary somewhat from country to country, but the characteristic features are usually the same.[6]

[6] For a study of the relative attractions and repulsions exerted by various figures of age in different countries, the reader is referred to Bachi's article in the *Bulletin of the International Statistical Institute,* Vol. XXXIII, Part IV.

Whereas these effects of age errors are readily apparent, it is usually difficult to interpret accurately their causes. Probably most age mis-statements result in statements of round figures (for example, persons aged either 39 or 41 may report their ages as 40), though this is not necessarily always the case (a person aged exactly 40 years may report his age as 41 or 39, or some other figure). Where ignorance is the main cause, the great majority of mis-stated ages are likely to be reported in round figures. If ages are deliberately mis-stated, they may often be given at other than round figures in order to create an impression of accurate reporting.

The extent of mis-statements cannot be directly inferred from the data. While mis-statements by one year only are probably most frequent, it is possible that in some cases the error is considerably larger. At advanced ages, mis-statements by several years are probably very frequent.

The data show no immediate evidence as to whether there are any tendencies, in certain age ranges, for overstatements of age to be more frequent than understatements, or *vice versa.* This question is of considerable importance. If errors in both directions are of nearly the same frequency and magnitude, combination of the returns into five-year groups, or the application

Table 6

POPULATION OF TURKEY, BY SEX AND SINGLE YEARS OF AGE, ACCORDING TO THE CENSUS OF 1945 (In thousands)[a]

Age	*Males*	*Females*	*Age*	*Males*	*Females*	*Age*	*Males*	*Females*
TOTAL	9,447	9,344						
0–4	1,287	1,185	35–39	607	578	70–74	80	133
0	256	219	35	283	373	70	58	116
1	240	217	36	101	61	71	6	5
2	262	249	37	77	42	72	7	6
3	268	258	38	96	74	73	5	4
4	261	242	39	50	28	74	3	2
5–9	1,348	1,242	40–44	542	558	75–79	30	39
5	312	292	40	244	409	75	19	30
6	282	265	41	73	30	76	4	3
7	253	228	42	90	48	77	3	2
8	334	313	43	76	39	78	3	3
9	167	144	44	59	31	79	1	1
10–14	1,285	1,074	45–49	401	378	80–84	29	46
10	386	337	45	201	260	80	24	42
11	133	109	46	68	36	81	2	2
12	341	277	47	47	25	82	2	1
13	223	181	48	55	40	83	1	1
14	201	170	49	30	17	84	1	1
15–19	1,049	931	50–54	283	434	85–89	7.5	9.4
15	298	267	50	158	350	85	5.4	7.2
16	198	174	51	36	21	86	0.8	0.9
17	166	144	52	38	28	87	0.6	0.6
18	255	259	53	28	19	88	0.5	0.5
19	132	88	54	22	15	89	0.3	0.3
20–24	789	692	55–59	171	219	90–94	7.2	11.5
20	271	381	55	78	150	90	6.2	10.4
21	150	62	56	29	23	91	0.3	0.3
22	154	113	57	22	15	92	0.3	0.3
23	123	77	58	27	23	93	0.2	0.2
24	92	59	59	16	10	94	0.2	0.1
25–29	484	619	60–64	200	349	95–99	2.5	3.6
25	191	393	60	121	297	95	1.1	1.8
26	79	59	61	23	14	96	0.3	0.3
27	65	46	62	23	17	97	0.2	0.2
28	89	88	63	20	13	98	0.3	0.3
29	61	33	64	13	8	99	0.7	1.0
30–34	731	700	65–69	99	125	100 and over	1.5	2.8
30	301	471	65	58	94	Unknown	11.3	13.1
31	103	42	66	13	9			
32	133	78	67	11	8			
33	107	60	68	10	9			
34	87	48	69	6	4			

[a] Source: Turkey, Central Office of Statistics, *Recensement général de la population du 21 octobre 1945,* Publication no. 286, Vol. no. 65, Ankara, 1950.

of a rather simple smoothing formula, will suffice to remedy most of their defects. If the mis-statements are highly tendentious, it is doubtful whether even a complicated graduation formula can eliminate the resulting distortions in the age distribution.

Figure 1 also shows the 1945 population of Turkey in five-year groups of ages. The corresponding numbers appear in table 7. Since Turkey has a high birth rate, we should expect its age structure to resemble a pyramid, with each successive layer somewhat smaller than the preceding one. Deviations from this pattern in a high birth-rate country are possible if there have been marked previous changes in the birth rate, large immigration or emigration, or special mortality losses affecting particular age groups (for example, military casualties). Turkey has indeed experienced large population transfers affecting certain ethnic groups but the transferred populations probably represented a fair cross-section of age groups; hence, no great irregularity in the age distribution is to be expected on this ground. Military casualties during the First World War and subsequent years must have resulted in some reduction in the number of male survivors aged between 40 and 60 years in 1945. During both the First World War and the Second World War, the male population of military ages was mobilized; the absence of many young men from their homes could have resulted in some reduction of the birth rate during both periods and, hence, in reduced numbers of survivors of either sex aged 0-4 and 25-29 by 1945.

Figure 1 and table 7 show the following departures from a regular age pyramid.

1. Numbers aged 0-4 are indeed smaller than those aged 5-9. As it is common, in the censuses of many countries, that small children are enumerated incompletely, one may at first suspect this to be the result of a great deficiency in child enumeration. However, as already pointed out, the differences may have been due partly to mobilization during the Second World War. Possibly both incomplete enumeration and a birth deficit combined to produce this effect.

2. Fewer persons are reported at ages 25-29 than either at ages 20-24 or 30-34. A possible explanation is a marked birth deficit (or increased infant and child mortality) during 1915-1920. If this was the case, it is probable that the relatively small number of children aged 0-4 should also be largely accounted for in the same terms. Closer inspection of the deficit at ages 25-29, however, reveals additional peculiarities. Females aged 30-34 outnumber not only those aged 25-29, but also those aged 20-24; in the case of males, this is not true. This observation cannot be explained in terms of a birth deficit, which would affect both sexes alike; it indicates either that the ages of females were mis-stated differently from those of males, or that many women aged 20-29 were not enumerated. It has been observed in many censuses that young women tend to understate their ages. In the present instance, it is possible that many women in their early 20's reported their ages in the 'teens, and also that some women in the early 30's reported themselves in the upper 20's. This conclusion is not certain since males may also have mis-stated their ages in some unknown fashion. Moreover, there might have been a tendency to omit young women in the census enumerations. The First World War birth deficits, if any, could also have continued for one or two years after 1920, owing to internal warfare in Turkey affecting the birth cohort 20-24 years in 1945. A similar analysis of the age returns from an earlier or a later census of Turkey, using some of the methods to be discussed later in this chapter, would be very helpful in evaluating the possibility of explaining the irregularities in terms of a birth deficit. Even without reference to another census, a statistician with intimate knowledge of the country and its people could find the most plausible explanation for the apparent inconsistencies in the returns.

3. Females aged 50-54, 60-64, 70-74, 80-84, 90-94, and males aged 60-64, appear in each case more numerous than those in the preceding 5-year group. It is altogether improbable that birth rates could have fluctuated during the entire half-century from 1850 to 1900 so as to produce these effects. The alternating excesses and deficiencies in these five-year groups must be attributed to the very powerful attraction exerted by figures which are multiples of 10; they imply that many of the people over 50 were quite ignorant of their real ages, that the reporting was very negligent, or that there was a definite reluctance to report the ages exactly.

Table 7

Population of Turkey, 1945, by sex and 5-year groups of age, according to the census (In thousands)[a]

Age-group	*Males*	*Females*
Total	9,447	9,344
0–4	1,287	1,185
5–9	1,348	1,242
10–14	1,285	1,074
15–19	1,049	931
20–24	789	692
25–29	484	619
30–34	731	700
35–39	607	578
40–44	542	558
45–49	401	378
50–54	283	434
55–59	171	219
60–64	200	349
65–69	99	125
70–74	80	133
75–79	30	39
80–84	29	46
85–89	8	9
90–94	7	11
95 and over	4	6
Unknown	11	13

[a] Source: Turkey, Central Office of Statistics, *Recensement général de la population du 21 octobre 1945*, Publication no. 286, Vol. no. 65, Ankara, 1950.

Comparison of data with an expected configuration

Nothing is to be gained by a further examination of the Turkish data by single years of age because the effects of attractions of particular figures of age are immediately evident from inspection of the data. However, the statistician examining census age statistics which, though possibly inaccurate, are not so erratic as those of Turkey in 1945, may wish to use additional tests. For some populations, ages are reported fairly well with the exception of the advanced ages. In such instances, it is advantageous to compare the reported numbers at individual years of age with those obtained by some method of graduation or smoothing. For this purpose, comparison with a 5-year moving average may be entirely sufficient. A brief example is given in table 8.

In the case of the Turkish census, more can be learned from a comparison of the returns with an expected pattern of age composition in five-year groups. As has already been mentioned, the Turkish age structure should resemble a pyramid, with the exception of any effects of war-time birth deficits or military casualties. Such a pyramid can be constructed by computing a stable population[7] from the L_x values[8] of a suitable life table by applying an appropriate geometric rate of increase. Since great precision is not required, any stable population which fits the data tolerably well

[7] A stable population is a population of such age structure as would result if the same rates of mortality and fertility prevailed over a period of several generations. It differs from a stationary population in that the birth rate and the death rate may differ, whereas in a stationary population the birth rate equals the death rate. The actual population of a country with high birth rates often resembles a stable population fairly closely even though mortality has changed in the past.

[8] The L_x values in a life table indicate the age composition of a stationary population that would result from the mortality rates of the life table and a constant number of births (usually 100,000) every year.

Table 8

POPULATION OF THE UNITED STATES, 1940, REPORTED IN THE CENSUS AT AGES 80 TO 90, COMPARED WITH 5-YEAR MOVING AVERAGES

Age	Population reported at census	5-year moving average	Excess (+) or deficit (—) of reported numbers	Excess or deficit as percentage of moving average
78..........	65,045			
79..........	58,066			
80..........	53,512	50,189	+3,323	+6.6
81..........	38,307	43,245	—4,938	—8.9
82..........	36,016	36,831	— 715	—2.2
83..........	30,324	30,318	+ 6	+0.0
84..........	25,997	25,844	+ 153	+0.6
85..........	20,946	21,232	— 286	—1.3
86..........	15,938	17,134	—1,196	—7.0
87..........	12,957	13,481	— 524	—3.9
88..........	9,834	10,610	— 776	—7.3
89..........	7,728	8,166	— 438	—5.4
90..........	6,596	6,162	+ 434	+7.0
91..........	3,716			
92..........	2,932			

can be used. In the present instance, a good fit was obtained by applying a 2 per cent annual rate of increase to the stationary population of the 1936-1938 life table for Egypt.[9] This was done by applying successive powers of 1.02^5 to the ${}_5L_x$ values (that is to say, the stationary population in five-year age groups).

Ordinarily, a comparison of a stable with an actual population is made by multiplying the figures of the stable population by a constant, namely, the ratio between the totals of the two populations. In the present instance, in order to allow for possible birth deficits resulting in reduced numbers of persons aged 0-4 and 25-29, the stable population was multiplied, separately for each sex, by the ratio between the total stable and actual populations for all ages other than 0-4 and 25-29. The results are shown in table 9 and figure 2.

Since the comparison is rough, no importance should be attached to the exact size of the differences between the census data and the stable population figures for any particular age-sex category. Nevertheless, significant observations can be made. There is a considerable deficit at ages 0-4, which may represent in part the effect of a decline in the birth rate during 1940-1945. This deficit, however, appears to be larger for males than for females. On the other hand, the reported numbers of males appear markedly in excess at ages 10-14. A possible, but not necessarily correct, explanation would be a tendency to overstate the ages of boys under 10 and to understate the ages of boys just past 15 years of age. At ages 20-24 and 25-29, both sexes show deficient numbers, but this deficit is differently

[9] These assumptions do not necessarily imply that Turkish mortality has at any time been identical with that of Egypt in 1936-1938, nor that the Turkish population has always been increasing at 2 per cent per annum. Provided that the birth rate has been fairly constant, the stable population selected for this illustration may still provide a good fit if death rates in the past have been higher (and the rate of population increase consequently lower) or *vice versa*.

Table 9

POPULATION OF TURKEY, 1945, BY SEX AND 5-YEAR AGE GROUPS. COMPARISON OF CENSUS FIGURES WITH STABLE POPULATION

(Numbers in thousands)

Age-group	Males			Females		
	Census (1)	Stable population (2)	Ratio (1):(2)	Census (1)	Stable population (2)	Ratio (1):(2)
TOTAL	9,435[a]	10,148	0.93	9,330[a]	9.772	0.95
0-4..............	1,287	1,716	0.75	1,185	1,516	0.78
5-9..............	1,348	1,334	1.01	1,242	1,180	1.05
10-14..............	1,285	1,160	1.11	1,074	1,052	1.02
15-19..............	1,049	1,013	1.03	931	937	0.99
20-24..............	789	886	0.89	692	831	0.83
25-29..............	484	768	0.63	619	734	0.84
30-34..............	731	662	1.10	700	644	1.09
35-39..............	607	569	1.25	578	561	1.03
40-44..............	542	485	1.12	558	487	1.15
45-49..............	401	408	0.98	378	421	0.90
50-54..............	283	337	0.84	434	360	1.21
55-59..............	171	271	0.63	219	303	0.72
60-64..............	200	209	0.96	349	250	1.40
65-69..............	99	152	0.65	125	199	0.63
70-74..............	80	98	0.81	133	144	0.92
75-79..............	30	53	0.58	39	90	0.43
80-84..............	29	21	1.36	46	43	1.07
85 and over........	19	7	2.60	27	20	1.34

[a] Excluding persons of unknown age.

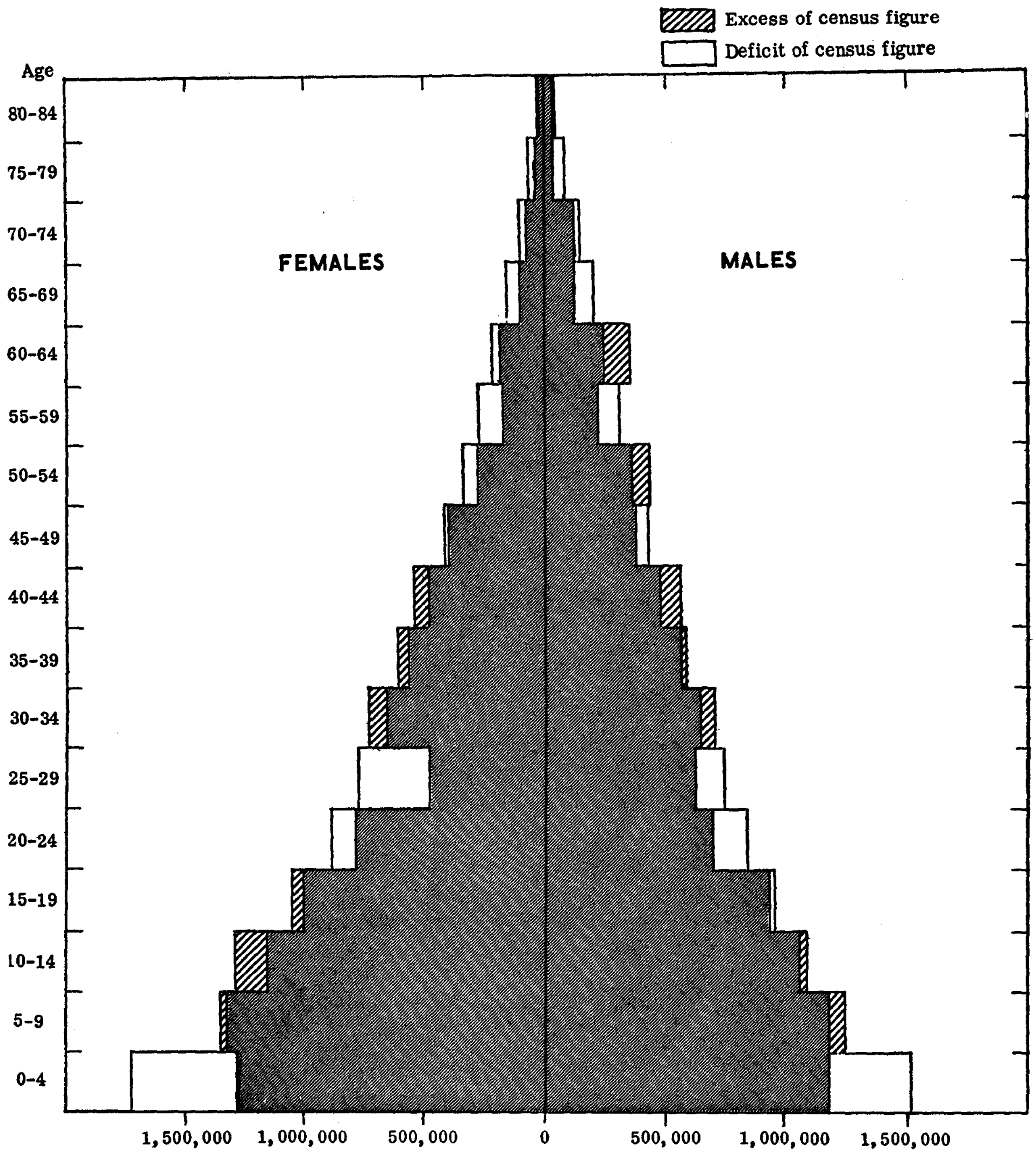

Figure 2. POPULATION OF TURKEY, 1945, BY SEX AND 5-YEAR AGE GROUPS. COMPARISON OF CENSUS FIGURES WITH A STABLE POPULATION

distributed in the two cases. The deficiency in the case of females is about the same at ages 20-24 as at 25-29. It could be explained as the result of a birth deficit only if the birth rate were as much depressed in 1920-1925 as in 1915-1920, which seems very unlikely. The relative deficits of males aged 20-24 and 25-29 could more plausibly be accounted for by the possible birth deficit. The observations tend to confirm that either the ages of many young women were mis-stated tendentiously, or many women in the age-range 20-29 were omitted from the census count.

The comparison of the census data with the stable population becomes less and less accurate as we proceed to more advanced ages. It is unlikely that the hypotheses of 1936-1938 Egyptian mortality and an annual rate of natural increase of 2 per cent are realistic for the elder cohorts, who lived under the conditions of a more remote past. Nevertheless, the comparison does not lose all meaning. The consistent deficit of males above age 45 could be attributable to military casualties during the First World War and the years immediately following. The alternation of excesses and deficits among females aged over 40, on the other hand, shows once more the attraction of ages ending in zero.

Analysis of ratios computed from the data

For the examination of the accuracy of age statistics, it is useful to compute sex ratios for each age category, and ratios for each sex separately, between the numbers reported in adjacent age categories. These ratios can be defined in several ways. In the present context, we shall define the "sex-ratio" as the number of males per 100 females in the same age class; and the "age-ratio" as 100 times the number of persons in a given age class divided by the arithmetic average of numbers in the two adjoining age categories. Sex-ratios should ordinarily change only very gradually from one age to another, as they are determined mainly by the sex ratio of births and sex differences in mortality at various ages. Age-ratios should ordinarily deviate very little from 100, except at advanced ages or as a result of major fluctuations in past birth rates.[10] The variations in sex-ratios and age-ratios computed from grouped age data, which may occur as a result of age mis-statement or for other reasons, have been studied by the Population Branch of the United Nations.[11]

[10] Migration and military casualties may disturb the normal patterns of sex-ratios and age-ratios.

Sex-ratios and age-ratios, computed by single years of age from the data of the Turkish census of 1945, are shown in table 10. It will be noted that the sex-ratios are invariably low at those round ages upon which the reports tend to be concentrated, and high at other ages, the reason for this being more frequent mis-statement of ages of females than of males. It would also appear that the enumeration of female infants was less complete than that of male infants, as witnessed by high sex-ratios at ages 0 and 1. Age-ratios reflect in a certain way the relative powers of attraction of certain figures of age; up to age 18, these ratios are very similar for males and females. At ages above 20, the attractive force of round figures is usually several times as great in the case of females as in that of males. At advanced ages, the reports for males are almost as erratic as for females.

Ratios of numbers reported at certain digits of age may or may not reveal any tendencies which exist towards overstatement or understatement of age. For example, one may compare, for each age which is a multiple of 10, the numbers of persons reporting the

[11] See "Accuracy tests for census age distributions tabulated in five-year and ten-year groups", United Nations, *Population Bulletin*, No. 2, October 1952, pp. 59-79. For a study of the effect of inaccurate age statements on sex ratios, see also Louis Henry, "La masculinité par âge dans les recensements", *Population*, 3e année, no. 1, 1948.

Table 10

Sex-ratios and age-ratios computed from single-year age data of the Turkish census, 1945[a]

Age	Sex-ratio	Age-ratios: Males	Age-ratios: Females	Age	Sex-ratio	Age-ratios: Males	Age-ratios: Females	Age	Sex-ratio	Age-ratios: Males	Age-ratios: Females
0	117	—	—	33	178	98	96	66	149	39	18
1	111	93	93	34	182	45	22	67	137	90	86
2	105	103	105	35	76	301	688	68	111	127	154
3	104	102	105	36	166	56	29	69	130	16	7
4	108	90	88	37	182	73	63	70	50	985	2481
5	107	115	115	38	131	152	209	71	123	19	8
6	106	100	102	39	177	42	12	72	120	129	139
7	111	82	79	40	60	399	1399	73	139	101	90
8	107	159	168	41	239	43	13	74	134	25	14
9	116	46	44	42	187	122	140	75	66	563	1049
10	115	258	266	43	196	102	98	76	116	35	21
11	123	37	35	44	190	42	21	77	134	72	61
12	123	192	191	45	78	317	773	78	109	166	193
13	123	82	81	46	189	55	25	79	114	10	5
14	118	77	76	47	188	76	65	80	57	1616	3089
15	112	149	155	48	136	144	191	81	104	13	7
16	114	85	85	49	170	28	9	82	119	131	122
17	115	73	67	50	45	479	1806	83	123	86	87
18	99	171	223	51	170	37	11	84	120	25	17
19	150	50	27	52	138	118	137	85	75	671	937
20	71	193	508	53	147	94	89	86	93	27	22
21	241	70	25	54	142	41	18	87	97	88	86
22	137	113	162	55	52	310	781	88	102	114	109
23	160	100	90	56	125	57	28	89	96	9	6
24	156	59	25	57	148	79	65	90	59	2155	3269
25	49	224	667	58	118	141	186	91	84	9	6
26	133	62	27	59	169	22	6	92	99	134	128
27	140	77	63	60	41	613	2489	93	112	89	82
28	101	141	221	61	163	32	9	94	116	24	14
29	184	31	12	62	139	107	122	95	63	546	892
30	64	366	1248	63	153	109	104	96	95	37	27
31	244	48	15	64	159	33	15	97	113	86	65
32	170	126	152	65	62	442	1095	98	81	—	—

[a] The sex-ratio is here defined as the number of males of a given age per 100 females in the same age group; the age-ratio is defined as the number of persons of a given age group per 100 of the mean of numbers of the two adjoining age groups, of the same sex.

preceding age (terminal digit 9) with the number reporting the next subsequent age (terminating in 1), by taking the ratio of those reporting age 9 to those reporting age 11; of those reporting age 19 to those reporting 21 et cetera. If it can be presumed that a large part of age mis-statements are by one year only, then a tendency to overstate age might result in larger numbers reported with terminal digit 1 than with terminal digit 9; conversely, a tendency toward understatement might be reflected by a low ratio of numbers at ages ending in 1 to numbers at ages ending in 9. It must be understood, however, that these ratios are also affected by various other forms of age errors. From the Turkish data, the following ratios are obtained.

Ages which are multiples of 10 (x)	*Males*					*Females*
	Preceding age (x—1)		*Following age (x+1)*			
	Age	*Number reported*	*Age*	*Number reported*	*Ratio of number at age (x—1) to number at age (x+1)*	*Ratio of number at age (x—1) to number at age (x+1)*[a]
10	9	161,884	11	133,109	1.3	1.3
20	19	131,855	21	149,548	0.9	1.4
30	29	61,159	31	103,311	0.6	0.8
40	39	49,995	41	72,579	0.7	0.9
50	49	29,677	51	36,309	0.8	0.8
60	59	16,212	61	23,288	0.7	0.7

[a] Ratios computed in the same manner as in the case of males.

If high ratios in this case really indicate the prevalence of age understatement, and low ratios indicate overstatement, it appears that ages were understated around age 10 and overstated around ages 30, 40, 50 and 60. This interpretation is at best only partly correct. Relatively greater depletion of numbers at age 11 than at age 9, for instance, is probably in part a result of the strong attraction of age 12.

The differences between the ratios obtained for the two sexes are noteworthy. The ratio for females is much higher than that for males around age 20, an age at which women may readily be suspected of a desire to appear younger than they are. For ages 30 and 40, the female ratios are only slightly higher than the male, and for ages 50 and 60 the ratios for the two sexes are equal. Around ages 50 and 60, male and female sensitivities with respect to age may be nearly the same. The computation of sex-ratios and age-ratios for five-year age groups constitutes the basis of a method developed by the United Nations staff for examining the accuracy of age statistics.[12] It is useful to compute both ratios and examine them critically, particularly if there is some difficulty in constructing a stable population or some other expected age distribution with which the census data can be reasonably well compared. In the case of the 1945 Turkish census, analysis of the sex-ratios and age-ratios, for five-year groups, would yield essentially the same conclusions reached by the comparison with the stable population.

Measurement of age-accuracy by means of an index

It is of some interest to measure age-accuracy by an index for comparative purposes to establish, for example, whether the age statistics of one census are more accurate than those of another, whether there are great differences in age-accuracy between the urban and rural population et cetera. In the following, four methods of computing an index will be presented. Each of these methods has certain advantages and certain shortcomings. In particular, it is to be noted that three of these methods measure the extent of *digit-preference* rather than age-accuracy in a wider sense. Since not all age mis-statements result in statements at preferred digits, and since digit-preference can be present whether or not age mis-statements have any directional tendencies, a measure of digit-preference should be interpreted with some reservations. The method mentioned above employed by the United Nations staff is sensitive to both tendentious mis-statement and digit-preference, but has other defects which will be discussed.

Whipple's index

This index is obtained "by summing the age returns between 23 and 62 years inclusive and finding what percentage is borne by the sum of the returns of years ending with 5 and 0 to one-fifth of the total sum. The result would vary between a minimum of 100, representing no concentration at all, and a maximum of 500, if no returns were recorded with any digits other than the two mentioned".[13] The choice of 23 and 62 as the limits of the age band to be examined is arbitrary but has been found most suitable for practical purposes.

For the male and female population of Turkey as of the 1945 census, Whipple's index can be computed as follows (numbers in thousands):

Age	*Males*	*Females*
23	123	77
24	92	59
25–29	484	619
30–34	731	700
35–39	607	578
40–44	542	558
45–49	401	378
50–54	283	434
55–59	171	219
60	121	297
61	23	14
62	23	17
Sum, ages 23–62	3,601	3,950

Age	*Males*	*Females*
25	191	393
30	301	471
35	283	373
40	244	409
45	201	260
50	158	350
55	78	150
60	121	297
Sum, ages multiple of 5	1,577	2,703

Index for males

$$\frac{5 \times 1.577}{3{,}601} = 219.0$$

Index for females

$$\frac{5 \times 2.703}{3{,}950} = 342.2$$

[12] *Loc. cit.* See also pp. 42-43, *infra.*

[13] Census of India, 1921, Vol. I, Part I, Report by J. T. Marten, Calcutta, 1924, pp. 126-127.

Similar computations for the combined population of Turkish cities with more than 30,000 inhabitants yield an index of 146.3 for males and 223.3 for females. For the total of Turkish provincial and district capitals, the index is 165.8 for males and 292.8 for females. For the remainder of the country's population, which is mainly rural, the index amounts to 238.8 for males and 364.6 for females.

For purposes of comparison, the results of similar computations for other areas are shown below.

Bengal, 1901:		United States:	
Males	277.3	1880	144.8
Females	292.3	1890	131.3
Russia, 1897	175.2	1900	119.8
USSR, 1926	159.1	1910	120.9
Brazil, 1940:		1920	115.7
Males	143.3	1930	115.5
Females	153.1	1940	109.7
Poland, 1921	134.8	Sweden, 1920	100.4

Whipple's index is a very effective measure of age-accuracy so far as digit preference is concerned, and has the advantage that it can be computed very easily. Its main drawback, apart from measuring digit-preference only, is that it measures the preferences for only two digits, 0 and 5. As the preference for particular digits may be influenced by linguistic and other habits, it is probable that the average degree of preference for 0 and 5 varies among different peoples though, in general, these are always the most preferred digits.

Myers' index[14]

This index reflects preferences or dislikes for each of the ten digits, from 0 to 9. To determine such preferences, one might take successive sums of numbers recorded at ages ending in each of these digits; such a simple method, however, does not suffice since, with advancing terminal digits of age these sums will tend to decrease (for each successive digit, the population is one year older than for the preceding one). To avoid this inconvenience, the first step in Myers' method consists in the computation of a "blended" population in which ordinarily almost equal sums are to be expected for each digit. This being the case, the "blended" totals for each of the ten digits should be very nearly 10 per cent of their grand total. The deviations of each sum from 10 per cent of the grand total are added together, irrespective of whether they are positive or negative, and their sum is Myers' index.

The method of "blending" and the computation of the index are illustrated by means of the Turkish data in table 11. First, the sums of numbers at all ages terminating in each digit are computed for ages 10 and over, and for ages 20 and over. By multiplying the former with the successive coefficients 1, 2, 3 . . . 10 (for digits 0, 1, 2 . . . 9) and the latter with the successive coefficients 9, 8, 7 . . . 0 (for digits 0, 1, 2 . . . 9) and adding the results, the "blended" population

[14] This index is explained by R. J. Myers in "Errors and Bias in the Reporting of Ages in Census Data", *Actuarial Society of America. Transactions*, Vol. XLI, 1940, pp. 411-415.

Table 11

APPLICATION OF MYERS' METHOD TO AGE DATA FOR MALES, ACCORDING TO TURKISH CENSUS OF 1945

(Numbers in thousands)

Terminal digit	*Numbers at ages specified* 10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	*Sum for ages 10-99*	*Sum for ages 20-99*
0	386	271	301	244	158	121	58	24	6	1,569	1,183
1	133	150	103	73	36	23	6	2	0	526	393
2	341	154	133	90	38	23	7	2	0	788	447
3	223	123	107	76	28	20	5	1	0	583	360
4	201	92	87	59	22	13	3	1	0	478	277
5	298	191	283	201	78	58	19	5	1	1,134	836
6	198	79	101	68	29	13	4	1	0	493	295
7	166	65	77	47	22	11	3	1	0	392	226
8	255	89	96	55	27	10	3	0	0	535	280
9	132	61	50	30	16	6	1	0	1	297	165
Sum	2,333	1,275	1,338	943	454	298	109	37	8	6,795	4,462

Terminal digit	*Ages 10-99* Sum	Coefficient	Product	*Ages 20-99* Sum	Coefficient	Product	*"Blended" Sum*	*Percentage distribution*	*Deviation from 10 per cent*
0	1,569	1	1,569	1,183	9	10,647	12,216	21.7	+ 11.7
1	526	2	1,052	393	8	3,144	4,196	7.5	— 2.5
2	788	3	2,364	447	7	3,129	5,493	9.8	— 0.2
3	583	4	2,332	360	6	2,160	4,492	8.0	— 2.0
4	478	5	2,390	277	5	1,385	3,775	6.7	— 3.3
5	1,134	6	6,804	836	4	3,344	10,148	18.0	+ 8.0
6	493	7	3,451	295	3	885	4,336	7.7	— 2.3
7	392	8	3,136	226	2	452	3,588	6.4	— 3.6
8	535	9	4,815	280	1	280	5,095	9.0	— 1.0
9	297	10	2,970	165	0	0	2,970	5.3	— 4.7
Sum							56,309	100.1	39.3

for each digit is obtained. As was shown by Myers, all these sums should be very nearly equal if ages are reported accurately. In the present example, they are markedly unequal and, by adding together, irrespective of sign, the deviations from 10 per cent of the grand total, an index of 39.3 is obtained.

In the case of the Turkish data, positive deviations appear only at terminal digits 0 and 5; hence, in this instance, the advantage of Myers' over Whipple's method is not obvious. There are, however, populations for which positive deviations are also found at other age digits, such as 2 or 8, which would be overlooked if Whipple's index only were computed.

Theoretically, Myers' index can vary from 0 to 180. If ages are reported accurately, all "blended" sums are very nearly equal and deviations from 10 per cent are negligible, resulting in an index of almost zero. If all ages were reported with the same terminal digit (for example digit 0), then 100 per cent of the "blended" total would be at this digit; the absolute sum of deviations from 10 per cent would then amount to 180. The following indices, computed for male populations, provide a few examples of the results which may be obtained from actual observations:

Bengal, 1901	62.6
Russia, 1897	20.5
Brazil, 1940	16.3
Australia, 1933	4.0
Sweden, 1939	1.2

Bachi's index[15]

Despite their demonstrated practical usefulness, both Myers' and Whipple's indices have some minor theoretical defects: it is not possible to formulate the precise theoretical conditions under which Whipple's index would be exactly 100 and Myers' index exactly zero. An ingenious method, developed by Bachi, avoids this defect. Because the computation of Bachi's index is somewhat more laborious than Myers' method, no detailed example is given in this manual.

Bachi's method consists, in principle, in a repetitive application of Whipple's method, to determine the relative preference given to each of the ten terminal digits. If the age range is suitably selected, very nearly 10 per cent of all persons within that age range should be expected at ages with each terminal digit. For theoretical reasons, as well as to satisfy practical requirements, the average of total numbers in two age ranges must be related to the number of persons reporting, within these ranges, the ages with a particular terminal digit. The following relations are then established:

1. The per cent relationship of age reports at ages 30, 40, 50, 60, and 70 to the average of all age reports in the age ranges 25-74 and 26-75;
2. The per cent relationship of age reports at ages 31, 41, 51, 61, and 71 to the average of all age reports in the age ranges 26-75 and 27-76;
3. The per cent relationship of age reports at ages 32, 42, 52, 62, and 72 to the average of all age reports in the age ranges 27-76 and 28-77;
4. The percentage formed by age reports at ages 33, 43, 53, and 63, as well as half the numbers at ages 23 and 73, in the average of all age reports in the age ranges 23-72 and 24-73;
5. The percentage formed by age reports at ages 34, 44, 54, and 64, as well as half the numbers at ages 24 and 74, in the average of all age reports in the age ranges 24-73 and 25-74;
6. The percentage formed by the age reports at ages 35, 45, 55, and 65, as well as half the numbers at ages 25 and 75, in the average of all age reports in the age ranges 25-74 and 26-75;
7. The percentage formed by age reports at ages 36, 46, 56, and 66, as well as half the numbers at ages 26 and 76, in the average of all age reports in the age ranges 26-75 and 27-76;
8. The percentage formed by age reports at ages 37, 47, 57, and 67, as well as half the numbers at ages 27 and 77, in the average of all age reports in the age ranges 27-76 and 28-77;
9. The percentage of age reports at ages 28, 38, 48, 58, and 68 of the average of all age reports in the age ranges 23-72 and 24-73.
10. The percentage of age reports at ages 29, 39, 49, 59, and 69 of the average of all age reports in the age ranges 24-73 and 25-74.

Deviations of percentages from 10 per cent can be taken as indices of preference or dislike for each of the ten digits. Bachi's index is then obtained by summing only the positive deviations from 10 per cent (the sum of positive deviations is theoretically equal to the sum of negative deviations). It is, therefore, of the order of one-half of Myers' index, with a theoretical range from 0 to 90. In practice, almost the same results are obtained by both these methods.[16]

The United Nations Secretariat method[17]

This method consists essentially in the computation of sex-ratios and age-ratios for five-year groups of ages, up to age 70.[18] In the case of sex-ratios, successive differences between one age group and the next are noted, and their average is taken, irrespective of sign. In the case of age-ratios, for either sex, deviations from 100 are noted and averaged irrespective of sign. Three times the average of sex-ratio differences is then added to the two averages of deviations of age ratios from 100, to compute the index. This procedure is illustrated, with the Turkish data, in table 12.

This method, unlike the ones described above, is applicable where single-year age data are not available. The resulting index is not very exact and should be regarded as an "order of magnitude" rather than a precise measurement. Various limitations must be borne in mind when applying this method. Exceptions must be allowed for in the case of irregularities arising in certain age groups as a result of real disturbances of the population trend, due to such factors as war casual-

[15] Roberto Bachi, "Measurement of the tendency to round off age returns", *Proceedings of the International Statistical Congress*, Rome, 1953.

[16] For a description of the method and presentation of results obtained, the reader is referred to R. Bachi, "The Tendency to Round off Age Returns: Measurement and Correction", *Bulletin of the International Statistical Institute*, Vol. XXXIII, Part IV, pp. 195-222.

[17] For a full account of this method, the reader is referred to "Accuracy tests for census age distributions tabulated in five-year and ten-year groups", United Nations, *Population Bulletin* No. 2, October 1952, pp. 59-79.

[18] Sex-ratios and age-ratios as defined on p. 22 *supra*.

Table 12

COMPUTATION OF AGE-ACCURACY INDEX BY THE UNITED NATIONS SECRETARIAT METHOD FROM STATISTICS OF THE 1945 CENSUS OF TURKEY

Age-group	*Reported number*		*Analysis of sex-ratios*		*Analysis of age-ratios (males)*		*Analysis of age-ratios (females)*	
	Males	*Females*	*Ratios*	*Successive difference*	*Ratios*	*Deviation from 100*	*Ratio*	*Deviation from 100*
0-4	1,286,705	1,184,799	108.6	—	—	—	—	—
5-9	1,348,446	1,242,281	108.5	— 0.1	104.9	+ 4.9	110.0	+ 10.0
10-14	1,284,952	1,074,080	119.6	+ 11.1	107.2	+ 7.2	98.8	— 1.2
15-19	1,048,701	931,461	112.6	— 7.0	101.1	+ 1.1	105.5	+ 5.5
20-24	789,205	691,680	114.1	+ 1.5	103.0	+ 3.0	89.2	—10.8
25-29	484,328	619,069	78.2	—35.9	63.7	—36.3	89.0	—11.0
30-34	731,283	699,657	104.5	+ 26.3	134.0	+ 34.0	116.9	+ 16.9
35-39	607,377	578,390	105.0	+ 0.5	95.4	— 4.6	92.0	— 8.0
40-44	542,301	558,000	97.2	— 7.8	107.5	+ 7.5	116.6	+ 16.6
45-49	401,379	378,499	106.0	+ 8.8	97.3	— 2.7	76.3	—23.7
50-54	282,856	434,107	65.2	—40.8	98.8	— 1.2	145.2	+ 45.2
55-59	171,162	219,398	78.0	+ 12.8	70.9	—29.1	56.0	—44.0
60-64	199,908	349,207	57.2	—20.8	148.2	+ 48.2	203.0	+ 103.0
65-69	98,664	124,606	79.2	+ 22.0	70.5	—29.5	51.7	—48.3
70-74	80,007	132,964	—	—	—	—	—	—
Total (irrespective of sign)				195.4		209.3		344.2
Mean (total divided by 13)				15.0		16.1		26.5
Index (3 times mean difference sex-ratios plus mean deviations of male and female age-ratios.)						87.6		

ties, temporary birth deficits, or migratory movements involving mainly certain sex-age groups. For small populations, the measurement is also affected by chance fluctuations. The present example, however, illustrates an advantageous use of the method. The population of Turkey is large enough to permit neglecting any effect of chance fluctuations. Moreover, errors in age statements are sufficiently pronounced to overshadow whatever real irregularities of age structure may have existed.

The United Nations Secretariat method has the advantage over the methods of Whipple, Myers and Bachi that the index which is obtained is affected by differential omission of persons in various age groups from the census count and by tendentious age mis-statement as well as by digit-preference and is therefore more truly a reflection of the general accuracy of the age statistics. Also, it provides an indication of accuracy of the data in the form in which they are used for most purposes, that is, in age groups rather than single years. The methods applied to date by single years of age may in some cases show a fairly large amount of age mis-statement which has little influence on the grouped data.

C. Examination of detailed age statistics from two or more censuses

It is often difficult, as the examples presented above have shown, to determine whether irregularities revealed by tests of the age returns from a single census are due mainly to errors in the data or to real peculiarities of the population structure. When the results of two or more successive censuses are available, it is often possible to clear up these uncertainties even without the use of any more elaborate techniques than were described in the preceding section. For example, if the age statistics from a 1955 census of Turkey were at hand, the possibility of explaining certain irregularities in the 1945 data as the results of birth deficits or war casualties in the periods 1915 to 1925 and 1940 to 1945 would be greatly clarified. If the 1955 figures should show the same peculiarities in the age groups ten years higher, but not in the same age groups in which they appeared in 1945, there would be a strong basis for concluding that these peculiarities reflected the actual facts, rather than errors.

Still more definite information regarding errors can be obtained where data from two or more censuses at intervals of a few years are available, by using balancing equations or analagous calculations with the data for particular cohorts—comparing, for example, the numbers reported at ages 10-14 in an earlier census with those reported at ages 20-24 in a census ten years later. Where data from a series of three or more censuses are available, the returns may be linked in this manner over the entire series, with very illuminating results. For the purpose of explaining the techniques, however, it is sufficient to consider examples of the use of data from two censuses.

For the sake of simplicity in calculations, examples will be taken from censuses taken ten years apart. The balancing equation can be applied to any period of years, but it is arithmetically easiest when the interval is either five or ten years.

Not only the accuracy of age reporting, but also the completeness of enumeration of particular age and sex groups can be tested by proper use of the balancing equation under certain conditions which will be described below. On the other hand, balancing equation tests, unlike the procedures described in the preceding section, cannot often be applied to the age statistics for component areas of the country, because of the lack, in most cases, of adequate information on internal migration. For this reason, the methods to be illustrated here are useful, generally, only for an entire country, unless adequate data on internal migration are available.

In some countries it may be possible to apply these tests to the various ethnic groups separately, if age and sex data are tabulated for such groups and if data are available on immigration and emigration of these groups (or if the groups in question are not substantially affected by international migration.)

For the purpose of this analysis, any particular age group can be defined as a cohort: boys under 5 years of age, women 50 to 54 years, all persons 10 to 19 years of age et cetera, at a given census date. If a second census is taken exactly one decade later, the members of each cohort will be exactly ten years older at the time of the second census. In the meantime their numbers will be reduced by deaths and they may be increased or reduced by the balance of immigration and emigration. Ordinarily, mortality is the main factor; if the migration balance is negligible, the change in numbers can be used to compute a survival ratio analagous to that of a life table. For example, the Japanese census of 1940 reported 3,670,000 women aged 15 to 19 and the 1950 census reported 3,354,000 aged 25 to 29. The survival rate between 1940 and 1950 for this cohort was 3,354,000 divided by 3,670,000 or .913 (on the assumption that the data were accurate and that net migration was negligible).

Computed for one cohort only, such a survival rate often reveals little, if anything, about the accuracy of the statistics. However, a patently absurd result may be obtained, giving clear evidence of error. For example, an increase in the numbers of a cohort, from one census to another, is obviously impossible, unless there has been a substantial amount of immigration. On the other hand, even under conditions of very high mortality, it is unlikely that a cohort aged, say, anywhere between 5 and 60 years at the beginning, will be reduced by one-half within a decade.

More accurate judgement is possible if the survival rates are compared for cohorts of each sex at different ages. Survival rates are functions of age-specific death rates, and, like these, generally conform to more or less the same pattern of variation from age to age whether mortality is high or low. The rate of survival increases after the earliest years of childhood and usually attains its maximum around age 10; thereafter it declines, at first very gradually, but more and more rapidly as advanced ages are attained. Also, at most or all ages females have a somewhat higher rate of survival than do males of the same ages. If the hypothetical survival rates computed for different cohorts deviate significantly from this pattern, and if no explanation, such as migration, can be found, inaccuracy in the statistics must be suspected.

Table 13 presents ten-year survival rates for five-year cohorts, according to the life tables of three countries.[19] The purpose of this table is to indicate the approximate magnitudes and the general patterns of survival rates which should be expected. In passing, we may observe that the survival rate for women aged 15 to 19 according to the Japanese life table of 1947 was 0.935, slightly higher than the rate computed from census statistics for 1940 and 1950. Since the 1940-1950 period includes most of the Second World War, when mortality was probably higher than in 1947, the rate of 0.913 from the census material may be quite consistent with the rate of 0.935 derived from the life table.

Under what conditions can such comparisons of cohorts in successive censuses be made most meaningfully? One condition is either the absence of a substantial net immigration or emigration or full knowledge about the age and sex composition of the migrants. A second condition, analogous to the first, is that of constant boundaries. If the country's boundaries have changed between the two censuses so that considerable numbers of people have been added to or substracted from the population, the age and sex composition of these people must be known, if the cohort analysis is to give an accurate indication of the accuracy of the

[19] Survival ratios computed from a life table are ratios of the approximate L_x-values in the stationary population.

Table 13

TEN-YEAR SURVIVAL RATES FOR FIVE-YEAR COHORTS, BY SEX, ACCORDING TO LIFE TABLES OF THREE COUNTRIES

Age at beginning of decade	*Age at end of decade*	*Thailand 1947/1948*		*Japan 1947*		*Mexico 1930*	
		Males	*Females*	*Males*	*Females*	*Males*	*Females*
0 to 4	10 to 14	—	—	0.952	0.954	0.851	0.851
5 to 9	15 to 19	0.952	0.957	0.971	0.972	0.927	0.929
10 to 14	20 to 24	0.950	0.956	0.952	0.955	0.926	0.930
15 to 19	25 to 29	0.936	0.941	0.925	0.935	0.906	0.913
20 to 24	30 to 34	0.921	0.926	0.915	0.929	0.891	0.899
25 to 29	35 to 39	0.907	0.915	0.916	0.931	0.878	0.887
30 to 34	40 to 44	0.890	0.904	0.914	0.929	0.861	0.877
35 to 39	45 to 49	0.867	0.895	0.903	0.923	0.838	0.865
40 to 44	50 to 54	0.839	0.883	0.878	0.908	0.811	0.842
45 to 49	55 to 59	0.806	0.861	0.834	0.880	0.778	0.805
50 to 54	60 to 64	0.767	0.823	0.762	0.834	0.730	0.747
55 to 59	65 to 69	0.710	0.768	0.660	0.760	0.665	0.658
60 to 64	70 to 74	0.623	0.687	0.538	0.655	0.571	0.543
65 to 69	75 to 79	—	—	0.405	0.519	0.451	0.415
70 to 74	80 to 84	—	—	0.274	0.361	0.329	0.299

statistics. A third condition is identical coverage of the population at the two censuses. For example, if the entire male population is enumerated in one census, but the military are excluded at the second, the age cohorts involving the military cannot be compared without a suitable adjustment, unless the number of the military is negligible. If nationals living abroad are included in one census, and excluded from another, and if the numbers involved are large, especially if they are concentrated in any particular age or sex groups, this type of analysis is invalidated.

In the case of a country of immigration, under certain circumstances a cohort can be compared at two censuses even if migration data are lacking. If the native population (that is to say, persons born in the country) are known not to have emigrated in important numbers, the age and sex data can be tabulated by nativity, and comparisons of the two censuses limited to the native population. This point will be illustrated subsequently with data for the United States.

It should be noted that such survival rates can be calculated over any age span and time interval. Those shown in table 13 were calculated for a ten-year span of ages; but they can be calculated for single years, five-year, fifteen-year, or twenty-year periods, or whatever length of time is desired. A ten-year period was selected for the above illustrative purposes because the actual census survival rates, which will be discussed below, were obtained from censuses taken ten years apart.

To illustrate the application of this test, six countries were selected, in each of which two censuses had been taken a decade apart, and the data had been tabulated by five-year age groups and sex, so that it was arithmetically simple to calculate ten-year survival rates for each of the age and sex groups. If the censuses had been less than ten years apart, or if the data had been tabulated by age groups of varying size (for example, some groups of five years, some in single years, and some of ten years) the arithmetic would have been much more complicated.

Example of Egypt, 1937 to 1947

Table 14 and figure 3 show survival rates computed from the age distributions of the Egyptian censuses of 1937 and 1947. The first observation is that in four cases for the males, and in two for the females, the survival rate is higher than 1.00. The following possibilities must be considered: (*a*) net immigration into Egypt in those particular age and sex groups; (*b*) errors in the 1937 census, and (*c*) errors in the 1947 census. Actually, there is no evidence of immigration into Egypt such as would explain these ratios, and the conclusion follows that the 1937 or 1947 census data, or both, for the cohorts mentioned were considerably in error.

Table 14

COMPARISON OF AGE DISTRIBUTIONS ACCORDING TO CENSUSES OF EGYPT, 1937 AND 1947

(Population in thousands)

Age group (years)	Males 1937	Males 1947	Males Survival rate	Females 1937	Females 1947	Females Survival rate
(*a*) *Five-year groups*						
TOTAL	7,966.8	9,391.7	—	7,954.0	9,575.0	—
0 to 4	1,021.9	1,279.6	—	1,085.7	1,305.3	—
5 to 9	1,107.9	1,208.9	—	1,101.0	1,191.2	—
10 to 14	1,030.9	1,142.3	1.117	878.2	1,071.2	0.985
15 to 19	713.2	983.9	0.888	633.1	917.4	0.833
20 to 24	539.7	677.8	0.657	565.2	706.1	0.804
25 to 29	616.7	685.7	0.961	692.9	786.5	1.242
30 to 34	557.9	620.1	1.149	634.5	689.5	1.220
35 to 39	600.4	659.3	1.069	540.6	653.6	0.943
40 to 44	474.8	569.1	1.020	472.2	566.3	0.893
45 to 49	345.1	428.5	0.714	313.2	415.2	0.768
50 to 54	330.3	421.2	0.887	335.5	448.5	0.950
55 to 59	144.7	171.1	0.496	134.3	173.1	0.553
60 to 64	201.7	252.0	0.763	231.3	298.8	0.891
65 to 69	72.3	83.8	0.579	72.8	82.1	0.611
70 to 74	100.9	107.8	0.534	122.1	136.7	0.591
75 to 79	26.9	23.4	0.324	28.9	24.0	0.330
80 to 84	39.0	34.8	0.345	58.6	52.6	0.431
85 and over	24.3	17.3	0.192	34.8	23.5	0.192
Not reported	18.2	25.1	—	19.2	33.3	—
(*b*) *Ten-year groups*						
TOTAL, age 5 and over	6,944.9	8.112.1	—	6,868.3	8,269.6	—
5 to 14	2,138.8	2,351.2	—	1,979.2	2,262.4	—
15 to 24	1,252.9	1,661.7	0.777	1,198.3	1,623.5	0.820
25 to 34	1,174.6	1,305.8	1.042	1,327.4	1,476.0	1.232
35 to 44	1,075.2	1,228.4	1.046	1,012.8	1,219.9	0.919
45 to 54	675.4	849.7	0.790	648.7	863.7	0.853
55 to 64	346.4	423.1	0.626	365.6	471.9	0.727
65 to 74	173.2	191.6	0.553	194.9	218.8	0.598
75 to 84	65.9	58.2	0.336	87.5	76.6	0.393
85 and over	24.3	17.3	0.192	34.8	23.5	0.192

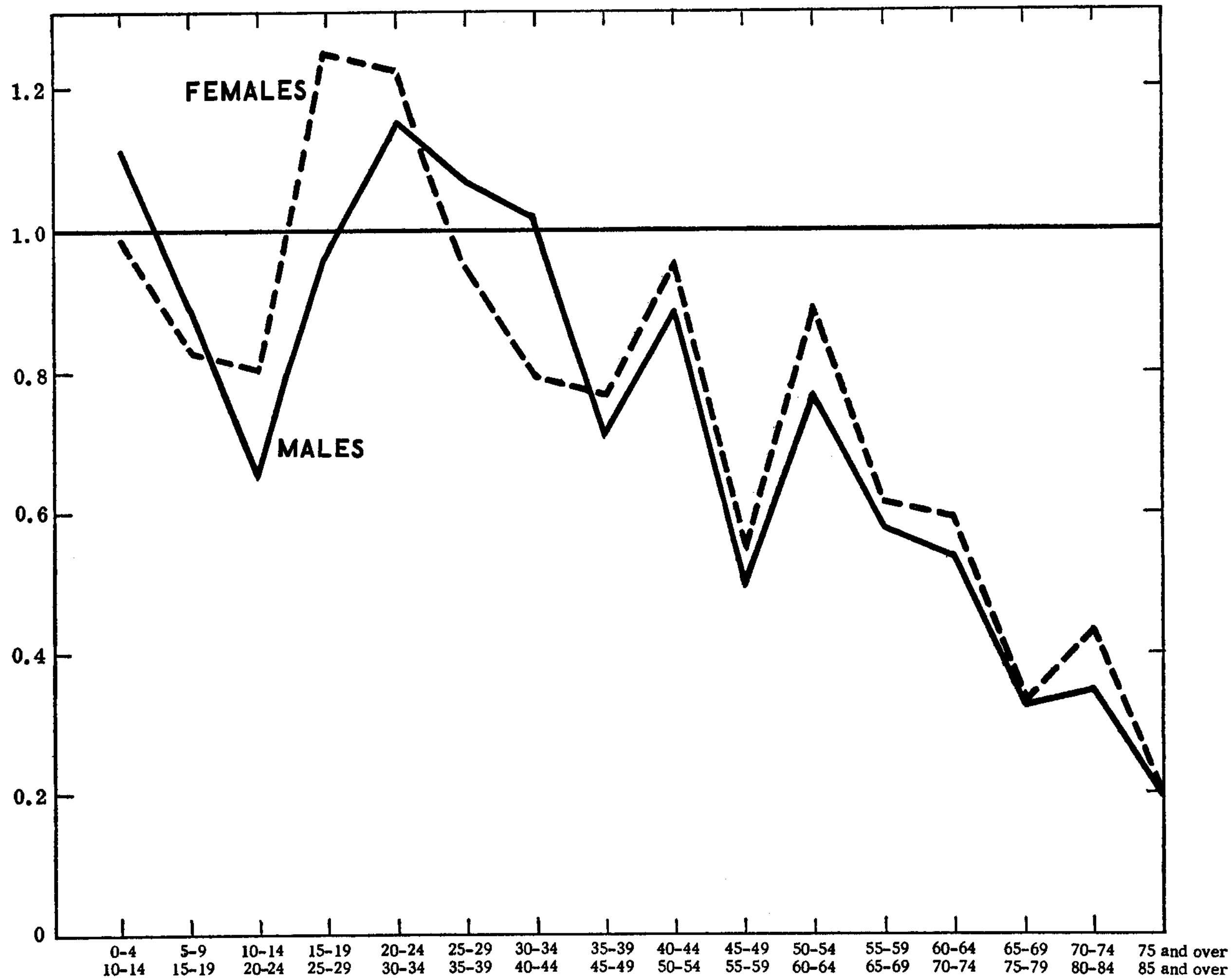

Figure 3. TEN-YEAR SURVIVAL RATES FOR MALE AND FEMALE COHORTS COMPUTED FROM EGYPTIAN CENSUS DATA FOR 1937 AND 1947

Examination of the entire pattern of survival rates reveals considerable irregularity. For example, among males 5 to 29 years old the rates were:

Between ages 5 to 9 and 15 to 19	0.888
Between ages 10 to 14 and 20 to 24	0.657
Between ages 15 to 19 and 25 to 29	0.961

Furthermore, between ages 50 to 54 and 60 to 64, the survival rate (0.763) was higher than the rate between the ages of 10 to 14 and 20 to 24 (0.657) and higher than between the ages of 35 to 39 and 45 to 49 (0.714).

Among females also the survival rates form an irregular pattern.

Comparison of the survival rates for males and females reveals that in four instances the rates are higher for the males. Although not impossible, this is improbable. As stated previously, higher mortality rates, and conversely lower survival rates, for females than for males may occur during the child-bearing ages in some countries having very high birth rates and high maternal mortality rates. Examination of the four instances in which the female survival rates were lower than those for males, reveals that only two of these were in the child-bearing ages; other explanations would have to be sought at least for the other two deviations; errors in the data are strongly suggested.

In table 14, the rates are also shown for ten-year cohorts. The results reveal some smoothing of the irregular pattern previously observed. Nevertheless, for two male cohorts, and one female cohort, the rate is over unity. For one cohort—between ages 25 to 34 and 35 to 44—the rate for females is lower than that for males; this is during the child-bearing period.

In summary, it must be concluded that the patterns of survival rates obtained by comparing the two censuses are so irregular, particularly in comparison with the patterns observed in other countries, as to imply a high degree of inaccuracy in the data. Further investigation would be required to arrive at definite estimates of the amounts of the errors and explanation of their causes (age mis-statements or under-enumeration in each census).

Example of Honduras, 1940 to 1950

Survival rates, calculated from census data by the same method as in the case of Egypt, are shown in table 15 for Honduras, the Philippines, Portugal, Thailand and Turkey. The rates for four of these countries are also presented in figure 4.

For Honduras, as for Egypt, the first observation is that for four cohorts for the males and six for the fe-

Table 15

SURVIVAL RATES OF 5-YEAR AND 10-YEAR COHORTS COMPUTED FROM CENSUS DATA FOR FIVE COUNTRIES

Five-year cohorts		*Honduras 1940-1950*		*Philippines 1939-1948*		*Portugal 1940-1950*		*Thailand 1937-1947*		*Turkey 1935-1945*	
Age at beginning of decade	*Age at end of decade*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*
0- 4	10-14	0.966	0.918	0.988	0.979	0.952	0.971	0.928	0.927	0.898	0.828
5- 9	15-19	1.016	1.034	0.783	0.875	0.945	0.991	0.940	0.980	0.866	0.826
10-14	20-24	0.973	1.078	0.935	1.048	0.928	0.963	0.883	0.965	0.930	0.927
15-19	25-29	0.913	0.958	0.894	0.908	0.894	0.920	0.871	0.905	0.867	1.274
20-24	30-34	0.912	0.885	0.718	0.753	0.833	0.877	0.876	0.871	0.976	1.093
25-29	35-39	1.050	0.984	0.860	0.821	0.916	0.941	0.912	0.877	0.977	0.802
30-34	40-44	1.000	1.020	0.824	0.822	0.926	0.952	0.833	0.826	0.928	0.869
35-39	45-49	0.977	0.977	0.815	0.780	0.914	0.934	0.813	0.858	0.825	0.743
40-44	50-54	0.933	1.046	0.689	0.742	0.901	0.940	0.751	0.814	0.874	0.916
45-49	55-59	0.985	1.031	0.588	0.618	0.853	0.905	0.753	0.788	0.705	0.697
50-54	60-64	1.115	1.147	0.847	0.766	0.812	0.875	0.682	0.666	0.885	0.908
55-59	65-69	0.870	0.910	0.639	0.612	0.754	0.842	0.565	0.572	0.609	0.638
60-64	70-74	0.660	0.755	0.456	0.509	0.600	0.672	0.357	0.472	0.456	0.447
65-69	75-79	—	—	0.415	0.441	0.496	0.588	0.290	0.410	0.303	0.370
70-74	80-84	0.718	0.797[a]	0.538	0.565	0.323	0.415	0.256	0.337	0.266	0.303[b]
75 and over	85 and over	—	—	0.385	0.351	0.149	0.211	0.230	0.257	—	—
Ten-year cohorts											
5-14	15-24	0.995	1.055	0.847	0.947	0.937	0.977	0.914	0.974	0.892	0.866
15-24	25-34	0.913	0.924	0.808	0.835	0.866	0.901	0.873	0.889	0.929	1.171
25-34	35-44	1.026	1.000	0.845	0.821	0.921	0.947	0.875	0.853	0.953	0.843
35-44	45-54	0.956	1.009	0.764	0.764	0.908	0.937	0.786	0.839	0.845	0.826
45-54	55-64	1.043	1.083	0.691	0.683	0.834	0.891	0.722	0.732	0.792	0.813
55-64	65-74	0.773	0.838	0.542	0.559	0.680	0.760	0.469	0.528	0.529	0.523
65-74	75-84	0.718	0.797[c]	0.465	0.498	0.425	0.515	0.279	0.383	0.280	0.322[d]
75 and over	85 and over	—	—	0.385	0.351	0.149	0.211	0.230	0.257	—	—

[a] Ratio of numbers aged 75 and over in 1950 to those aged 65 and over in 1940.

[b] Ratio of numbers aged 80 and over in 1945 to those aged 70 and over in 1935.

[c] Ratio of numbers aged 75 and over in 1950 to those aged 65 and over in 1940.

[d] Ratio of numbers aged 75 and over in 1945 to those aged 65 and over in 1935.

males, the survival rates are above unity. The question arises, whether immigration during this decade of persons in these particular age and sex groups could have produced this result.

The second observation is that the survival rates are all very high in comparison with those for the other countries shown. If there was considerable immigration at all ages or the death rate was unusually low in Honduras, these rates might be correct.

Comparison of the rates for the two sexes reveals three ages at which the males have higher survival rates. Two of these are in the child-bearing period.

The above short analysis is presented as an illustration of the procedure for detecting suspicious figures. Actually, the explanation for these unusual survival rates is that the 1950 census data were adjusted for under-enumeration whereas the 1940 figures were not so adjusted. Comparison of the unadjusted 1950 census data with those of 1940 reveals a much more plausible pattern of survival rates. More details are given in appendix C.

Example of the Philippines, 1939 to 1948

In only one instance is the survival rate above unity, namely, for females between the ages of 10 to 14 and 20 to 24. The pattern of decreasing survival rates with increasing age is not as regular or smooth as would be expected in a life-table population, but nevertheless, more regular than in the case of Egypt.

Comparison of the rates for males and females reveals seven cohorts in which the males had higher rates.

When the data are combined into ten-year groups, the patterns become much more regular; with increasing age the survival rates become progressively smaller. There is only one exception, among men between the ages of 15 to 24 and 25 to 34; this rate is lower than that of the preceding or following ages. One possible explanation for this lower rate is that it may have resulted from military casualties during the Second World War. There was considerable fighting in the Philippines and many men were killed; the group of men 15 to 24 years of age in 1939 probably would have been more involved in such fighting than other age groups.

Comparison of the rates for the 10-year cohorts of males and females reveals three cases in which the males have higher survival rates. This fact suggests that there may have been differences in the completeness of enumeration of males and females at either or both of the censuses. Such differences, of course, may have come from mis-statements of age. They could have resulted also from:

(*a*) More under-enumeration of males than of females in 1939;

(*b*) More under-enumeration of females than of males in 1948;

(*c*) Over-enumeration of males in 1948 or of females in 1939;

(*d*) Any combination of the above.

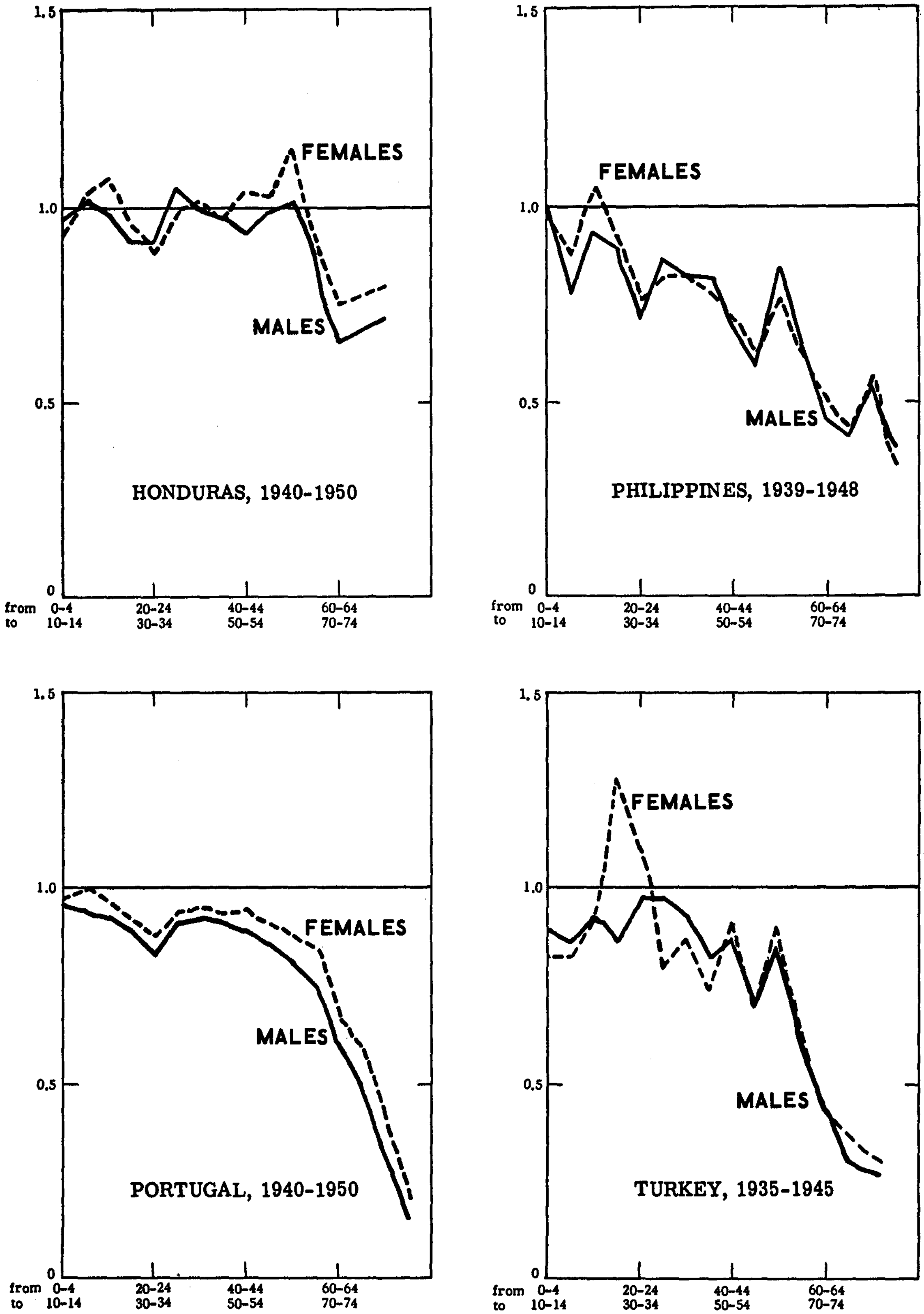

Figure 4. TEN-YEAR SURVIVAL RATES FOR MALES AND FEMALES COMPUTED FROM CENSUSES OF HONDURAS, THE PHILIPPINES, PORTUGAL AND TURKEY

Example of Portugal, 1940 to 1950

There are no instances in which the survival rates are over unity, and no instances in which the rates for the males are higher than those for the females. In general, the survival rates for each sex decrease with advancing age, except for slight deviations in the middle ages.

Examination of the data for 10-year cohorts suggests that there is only one cohort which deviates sufficiently to attract attention; this is the cohort aged 15 to 24 in 1940 and 25 to 34 in 1950. The survival rates for this group, both for males and females, are lower than those for the preceding or following age groups. This could have resulted from emigration of members of these particular cohorts, or from the errors in the census. It is desirable to investigate the former possibility by reference to migration data.

In summary, Portugal is a good illustration of a country in which the census survival rates follow the expected life-table pattern reasonably well. Additional tests of the Portuguese age statistics can be made; one of these tests is the application of the balancing equation to be described subsequently.

Example of Thailand, 1937 to 1947

Examination of the survival rates for Thailand by five-year age groups reveals approximately the same degree of regularity observed in the case of the Philippines. There are no instances in which the survival rates are above unity. On the other hand, in five instances the rates for the males are higher than those for the females.

Combining the ages into ten-year categories smoothes out some of the irregularities. In only one cohort is the rate lower for females, the cohort which was 35 to 44 years of age—still in the child-bearing period—in 1947.

Pending further investigation it can be concluded tentatively that the age data for this country are perhaps not quite as accurate as those for Portugal, but perhaps somewhat better than those of the Philippines.

Example of Turkey, 1935 to 1945

There are two instances in which the five-year survival rates are over unity, both among females. Also, there are eight instances in which the survival rates for males are higher than those for females.

The rates fluctuate from one age cohort to the next more than they do in the cases of Portugal and Thailand.

Combining the data into ten-year groups does not smooth the patterns of survival rates very much. There is still one cohort with a rate over unity, and four in which the rates for males are higher than those for females.

Comparisons of the Turkish patterns of census survival rates with those observed in some of the other countries, and with the expected patterns, indicates that there were large errors and that the quality of the data may be similar to that of the Egyptian material. Both mis-statements of age and differences in the completeness of census enumeration in 1945 and 1935 may be involved.

Computation of survival rates from age 5 and over to age 15 and over

It is useful mainly as a check on the completeness of enumeration, to compare the population aged 5 and over at the earlier census with those aged 15 and over at the later census, in the six cases which have been chosen as examples. In the absence of migration the latter group should be the survivors of the former. These ratios, by sex, are presented below:

Country	*Survival rate over decade*	
	Males	*Females*
Egypt	0.830	0.874
Honduras	0.959	0.989
Philippines	0.785	0.819
Portugal	0.850	0.875
Thailand	0.822	0.852
Turkey	0.850	0.844

One observation is that the rates for Honduras are unusually high in comparison with the other countries. In fact, they are so high that they are impossible, unless there was a large volume of immigration.

Among the other countries the survival rates, for each sex separately, do not differ greatly. Among males the highest rate is 0.850 in Portugal and the lowest is 0.785 for the Philippines. Among women the range is from 0.875 to 0.819.

There are no hard and fast rules regarding the amount of variation that is possible for this survival rate given accurate data. Purely as a rule of thumb, the rates in most cases can be expected to fall within the range of about 0.800 to 0.900. If a rate is very much above or below these limits, it can be taken as an indication of probable errors in the census which should be investigated further.

Use of the balancing equation

Various forms of the balancing equations have been considered in chapters I and II. In those instances, the terms of the equation were so arranged as to bring most clearly into relief the effects of inaccurate census enumeration, incomplete birth registration, or errors in the registration of deaths, as the case might be. Another type of balancing equation can be used as a refinement of the comparisons of survivors in a given cohort which have been discussed in the preceding pages. Instead of computing an implied survival rate and attempting to evaluate it, in the case of each cohort, actual statistics of deaths are introduced and so a better basis for comparing numbers at the two censuses is obtained. For each cohort, the balancing equation can then be expressed by the formula:

$$P_1 = P_0 - D + I - E$$

where P_0 is the number in the given cohort at the earlier census, P_1 is the number in the same cohort at the more recent census, D is the number of deaths to members of the cohort during the census-interval, I is the number of immigrants of relevant sex and age, and E that of emigrants of such sex and age, during the interval.

Actually, few countries have the required migration data by age and sex. Accordingly, in most cases it is necessary to ignore the migration components and reduce the equation to:

$$P_1 = P_0 - D.$$

The remarks made previously regarding the limitations of cohort analyses where migration has taken place are applicable here.

In addition to the effects of migration not taken into account, failure of the equation to balance can be due to any of the following factors or combinations of them:

1. Differences in completeness of the two census enumerations, with respect to the cohort in question,
2. Age mis-statements at one or both of the censuses,
3. Incompleteness of death registration,
4. Mis-statement of ages at death.

Knowledge and good judgment are important in determining the relative weight of these factors in each instance. The results of other tests applied to the census enumerations and vital statistics are especially helpful. For example, if a direct check of the death statistics shows that registration is nearly complete and that the reporting of ages at death is not subject to such errors as would greatly affect the balancing equation, any sizeable errors can be attributed to the census figures alone. In the case of incomplete death registration or erratic reporting of ages at death, if estimates of the amounts of error from these sources can be made, on the basis of results of checks or otherwise, the figures for deaths can be corrected before they are used in the balancing equation. The same remarks apply, of course, to known errors in the census figures, where the equations are used to check death registration in the various age groups.

There are also some general considerations which are helpful in identifying the sources of errors. In the first place, errors due to both under-enumeration in the census and under-registration of deaths will usually conform to a predictable pattern of differences from age to age and between the sexes. The probability of omission from the census is likely to be particularly great in the case of very young children, and there may be some other sex-age groups for which the probability of omission is also high for reasons that will be known to persons familiar with the census operations and the reactions of respondents in the country concerned (young men subject to military service, young women of marriageable age, et cetera). With these exceptions, the amount of under-enumeration is likely to be similar for different sex-age groups, though it may well vary somewhat with sex and age, and the changes from one age group to the next will ordinarily not be very large. The same generalizations apply to under-registration of deaths. Consequently, if the discrepancies shown by the balancing equations change abruptly from one cohort to the next, mis-statement of ages, rather than under-enumeration, is likely to be the main cause of these variations. Age mis-statements are especially indicated if the discrepancies for successive cohorts are alternately positive and negative.

There is likely to be a relationship between mis-statements of ages in the census records and on death certificates. Although the frequency of mis-statements may not be the same in the two types of records, the pattern of their incidence in various age categories is likely to be similar, at least in the cases of mis-statements due to negligence or ignorance of the correct age. However, any errors in the reporting of ages on death certificates will usually have much less effect on the balancing equations than errors in the census returns. Even within a period of ten years, only a minor fraction of each cohort, unless of advanced ages, is expected to die. If completeness or accuracy of age statements in death statistics is comparable to completeness and age-accuracy at the census count, the errors which the death statistics introduce into the balancing equations are fractional as compared to the census errors.

The effects of migration depend on the character of the migration. Migration for economic motives often affects particularly the cohorts of young men. Different types of migration may, however, vary as to their incidence by sex and age. The probable sex-age pattern of migration can usually be determined roughly, even without a complete statistical measure, by analysing available sources of information on the characteristics of the migrants and reasons for their migration.

Example of Portugal

Data for Portugal will be used to illustrate the use of balancing equations in the test for the accuracy of age reporting, for selected five-year age cohorts. These data are the age and sex distributions from the 1940 and 1950 censuses, and the deaths during this decade, as reported by age and sex.

The discussion of the accuracy of death registration in Portugal[20] seemed to show some indication that death reporting was not entirely complete. For the purpose of this example, however, no attempt has been made to correct the death statistics. Also, it was assumed for this illustration that there was no migration, although it is known that there was some net out-migration during the decade. The method is illustrated with the following data for the cohort of males aged 5 to 9 years in 1940:

Number reported in 1940 census	428,000
Number 15 to 19 years reported in 1950 census	404,400
Apparent inter-censal decrease	23,600
Number of deaths during decade reported to cohort	10,600
Residual (excess of inter-censal decrease over number of reported deaths)	13,000

Table 16 presents the results of such calculations for selected age groups, for males and females separately.

There is considerable variation in the size of the residuals for the various age-sex cohorts shown. For example, the residuals are relatively small for the cohorts aged 50 to 54 years in 1940; among males the residual amounts to only 2.1 per cent of the apparent inter-censal decrease, and among women to 8.3 per cent. On the other hand, among females aged 5 to 9 in 1940, the residual is greater than the inter-censal decrease, and among males aged 5 to 9 in 1940, the residual amounts to 55.0 per cent of the reported inter-censal decrease.

How important are these residuals? Are they so small that they can be ignored for most analytical purposes? What are the sources of the errors? There is no simple and decisive answer to these questions. In arriving at any final judgements it is necessary first of all, to introduce any data available on the numbers and age and sex composition of migrants during this decade. It might be, for example, that the residual of 15,800 males

[20] See chapter II, pages 26-7.

Table 16

Application of balancing equation to selected age-sex cohorts, for Portugal, 1940 to 1950
(Numbers in thousands)

Age at 1940 census	Age at 1950 census	Number enumerated 1940	Number enumerated 1950	Inter-censal decrease	Number of deaths reported	Residual: Number	Residual: Per cent of inter-censal decrease	Residual: Per cent of number enumerated in 1950
Males								
5 to 9	15 to 19	428.0	404.4	23.6	10.6	13.0	55.0	3.2
10 to 14	20 to 24	409.5	380.1	29.4	13.6	15.8	53.7	4.2
25 to 29	35 to 39	298.3	273.2	25.1	16.0	9.1	36.2	3.3
50 to 54	60 to 64	154.5	125.5	29.0	28.4	0.6	2.1	0.5
Females								
5 to 9	15 to 19	410.3	406.6	3.7	9.5	5.8	156.8	1.4
10 to 14	20 to 24	396.3	381.6	14.7	11.7	3.0	20.4	0.8
25 to 29	35 to 39	312.5	294.1	18.4	12.3	6.1	33.2	2.1
50 to 54	60 to 64	192.8	168.7	24.1	22.1	2.0	8.3	1.1

observed in the cohort aged 10 to 14 years in 1940 would disappear if emigration were taken into account; or it might become negative. On the other hand, it is also possible that the introduction of migration statistics into the equations would increase the residuals for cohorts aged 50 to 54 in 1940.

It should be noted that for some cohorts the residual is only one or two per cent of the number counted in the census. For most purposes to which the census data might be put, such differences are small enough to be ignored. Residuals which amount to four or five per cent or more, on the other hand, may have a significant effect upon the results of some types of analyses.

Example of Thailand

Several times in this manual reference has been made to the fact that children under 5 years of age are almost universally under-counted in censuses. For an illustration of the use of balancing equations to investigate under-enumeration in this age group, the following data on male children in *Thailand* will be used:

Number under 5 enumerated in the census of 1937....	1,230,500
Number 10 to 14 enumerated in the census of 1947..	1,142,000
Apparent inter-censal decrease	88,500
Reported number of deaths during decade to cohort under 5 years of age in 1937.....................	118,400
Residual (excess of deaths over apparent inter-censal decrease)	29,900

The true excess of deaths over the inter-censal increase was probably greater than that shown above, since the previous analysis (in chapter II) indicated the probability of under-registration of deaths. One estimate[21] was that only 74 per cent of the deaths during the decade had been registered. Applying this proportion to 118,400 provides an estimate of about 160,000 for the true number of deaths during the decade to children who were under 5 years of age in 1937. If this were correct, the residual—the excess of deaths over inter-censal decrease—would be about 71,500.

What other errors could account for the residual? They may be listed as:

(1) Undercount in 1937 census of children under 5 years of age;

(2) Overcount of children 1947 census aged 10 to 14;

(3) Age mis-statements in 1937 so that some children under 5 were returned as 5 to 9 years old;

(4) Age mis-statements in 1947 so that some children under 10 or 15 and over were returned as aged 10 to 14.

More than one of these errors probably occurred to some extent, but what is the likelihood that the first error was the most serious and of the largest size? Further testing beyond the scope of this manual would be required to establish the size of each type of error. Such testing should take into consideration all available data, both published and unpublished, which the technicians in the country can obtain, together with any additional knowledge of whatever social and economic factors may be particularly important in the country concerned.

Example of the United States

Data are available for the United States for the decade 1930 to 1940 which are useful for illustrating apparent discrepancies in the two census counts. These data and the relevant calculations (shown in table 17) are for native white males. By limiting the analysis to the native group any possible distortions which might be introduced by migration are largely avoided since, by and large, very few natives emigrated. There were no boundary changes and no war to disrupt the comparison.

Instead of death statistics, which are known to have been somewhat incomplete, a life table constructed by the national statisticians was used; this life table includes corrections for under-reporting of deaths. The number of deaths in each cohort during the decade was estimated by calculating life-table survival rates on the basis of the L_x figures. These values (shown in column (*b*) of table 17) were obtained by dividing the sum of L_x for each five-year age group by the sum of L_x for the group 10 years younger.

The findings are as follows:

(1) There was an under-enumeration of children under 5 years of age in 1930. The number enumerated in 1930 was 5,036,900 whereas the number 10 to 14 years in 1940 was enumerated as 5,231,900. Altogether

[21] See chapter II, page 26.

Table 17

APPLICATION OF BALANCING EQUATION, BY MEANS OF LIFE-TABLE SURVIVAL RATIOS, TO CENSUS AGE DISTRIBUTIONS OF UNITED STATES NATIVE WHITE MALES, 1930 AND 1940.[a]

(Numbers in thousands)

Age (years)	Enumerated in 1930 (a)	Survival rate 1930-1939 (b)	Expected number in 1940 (c)	Enumerated in 1940 (d)	Discrepancy	
					Absolute (e)	As per cent of enumerated (f)
0 to 4	5,036.9	—	—	—	—	—
(0 years)	(965.2)	(0.96272)	—	—	—	—
(1 year)	(960.3)	(0.97707)	—	—	—	—
(2 years)	(1,023.0)	(0.98124)	—	—	—	—
(3 years)	(1,046.6)	(0.98332)	—	—	—	—
(4 years)	(1,041.9)	(0.98439)	—	—	—	—
5 to 9	5,497.3	0.98432	—	—	—	—
10 to 14	5,265.8	0.97838	4,926.0	5,231.9	+305.9	+ 5.8
15 to 19	4,907.3	0.96131	5,411.1	5,433.5	+ 22.4	+ 0.4
20 to 24	4,346.9	0.96805	5,152.0	5,014.7	—137.3	— 2.7
25 to 29	3,731.8	0.96131	4,769.9	4,698.4	— 71.5	— 1.5
30 to 34	3,408.6	0.94920	4,208.0	4,230.3	+ 22.3	+ 0.5
35 to 39	3,278.8	0.93063	3,587.4	3,724.2	+136.8	+ 3.7
40 to 44	2,771.5	0.90232	3,235.4	3,338.4	+103.0	+ 3.1
45 to 49	2,411.9	0.86278	3,051.3	3,025.7	— 25.6	— 0.8
50 to 54	2,092.8	0.80819	2,500.8	2,568.4	+ 67.6	+ 2.6
55 to 59	1,670.6	0.73120	2,080.9	2,054.2	— 26.7	— 1.3
60 to 64	1,305.3	0.62691	1,691.4	1,659.2	— 32.2	— 1.9
65 to 69	944.8	0.49318	1,221.5	1,314.2	+ 92.7	+ 7.1
70 to 74	690.0	0.33944	818.3	873.2	+ 54.9	+ 6.3
75 to 79	376.6	0.19590	466.0	487.8	+ 21.8	+ 4.5
80 to 84	165.8	0.09118	234.2	250.4	+ 16.2	+ 6.5
85 to 89	57.0	0.03263	73.8	81.8	+ 8.0	+ 9.8
90 to 94	11.7	0.00772	15.1	17.3	+ 2.2	+12.7
95 to 99	1.9	—	1.9	2.9	+ 1.0	+34.5
100 and over	0.3	—	0.1	0.3	+ 0.2	+66.7
Unknown	36.5	—	—	—	—	—
Totals						
All ages	48,010.1	—	43,445.1	44,006.8	+561.7	+ 1.3
20 years and over	—	—	33,108.0	33,341.4	+233.4	+ 0.7

[a] From A. Jaffe, *Handbook of Statistical Methods for Demographers,* U. S. Bureau of the Census, Washington 1951, p. 91.

an estimated 305,900 native white male children appear to have been omitted from the 1930 census.

(2) In 1940 there appears to be a deficit in the cohort 20 to 29 years old at that time and a surplus in the cohort 30 to 39 years (or 30 to 44 years). This suggests that young adult males may have been missed in both censuses, an observation which fits in with other evidence and which could be explained on the theory that many young men, at the age at which they leave their paternal home to seek employment and perhaps to migrate, and before they have established their families, do not have a well-established place of residence where they will be reported by other household members to the census enumerators. The reasoning is as follows. If, in the United States censuses there is a tendency to omit males about 20 to 29 years old, but not of men 30 to 39 years old, the 1930 count for ages 20-29 is presumably too low, but the 1940 count for ages 30-39 should be approximately correct, showing a surplus such as that observed over the expected survivors of this cohort. In the 1940 census on the other hand, there would be again a deficit in the age range of 20 to 29 years. On the other hand, young men may have tended to overstate their ages, with the result that unduly small numbers were reported in their twenties and unduly large numbers in their thirties. If this happened frequently on the occasion of both censuses, the same results would follow.

(3) At the older ages beginning about 65 there appears to be a tendency to exaggerate age in the census reports, a tendency which seems to increase with advancing age. Its appearance about 65 is probably related to the fact that at this age many persons become eligible for old-age benefits.

Comparison of the entire native white male population enumerated in 1930 with that 10 years of age and over in 1940 reveals on the whole but minor discrepancies, as follows:

Number enumerated in 1930	48,010,100
Number 10 and over enumerated in 1940	44,006,800
Apparent inter-censal decrease	4,003,300
Estimated number of deaths during decade to persons alive in 1930	4,565,000
Residual (excess of deaths over apparent decrease)	561,700

Of the total residual, over one-half seems to be accounted for by the presumed under-enumeration of children in 1930. If the balancing equation is applied to the population 10 years of age and over in 1930 and 20 years of age and over in 1940, the residual is reduced to 233,400, as follows:

Number aged 10 and over enumerated in 1930......	37,475,900
Number aged 20 and over enumerated in 1940......	33,341,400
Apparent inter-censal decrease	4,134,500
Estimated number of deaths during decade to persons 10 years of age and over in 1930............	4,367,900
Inter-censal decrease	4,134,500
Residual (excess of deaths)	233,400

The last residual is about 5.6 per cent of the inter-censal decrease, and only 0.7 per cent of the enumerated 1940 population—too small to invalidate the data for most purposes.

On the basis of a comparison of the 1940 and 1950 censuses the United States Census Bureau reported finding substantially the same types of errors as described above in the 1950 age and sex statistics.[22]

D. Evaluation of the accuracy of statistics of deaths by sex and age groups

Errors in statistics of deaths classified by sex and age may be caused by the same two factors noted in the case of census returns, namely, mis-statements of age and omissions from the records that are more frequent in some sex-age categories than in others. In the case of infant deaths there is an additional factor, which may be of considerable importance in many areas, namely, failure to follow exactly the prescribed statistical definition of an infant death as distinguished from a still-birth. The latter error commonly takes the form of reporting as still-born, if not failing to report in any way, infants who are born alive but die very shortly after birth.

The balancing equations and analogous computations described in preceding sections of this chapter can be used to test the accuracy of death statistics, as well as population census returns, by sex and age groups. Valid conclusions with regard to the accuracy of the death statistics, however, will only be possible if the census data are very nearly accurate, or if the errors in them have been measured by means of direct checks. In the case of infant deaths it is also necessary to have accurate data on the number of births in order to establish an equation which will yield a useful test of death registration, and this requirement is not likely to be met where the registration of infant deaths is seriously deficient.

Various other methods of investigating the reliability of statistics of death classified by sex and age groups, and of data on infant deaths, have been described in chapter II.

E. Direct checks

Direct checks on the completeness of census enumeration and registration of vital statistics such as have been discussed in chapters I and II can be used also to test the accuracy of census and death statistics classified by sex and age groups, and to provide a basis for correction of any errors which may have resulted either from mis-statement of ages or from omissions varying in frequency within different sex-age categories. This can only be done, of course, if the independent records, against which the census returns or death registers are checked, contain reliable information on ages. When the independent record is established by means of a special field survey, it is well worth-while to include information on ages in this record and to take whatever steps are feasible to ensure accurate age returns. In a sample check enumeration of the population, specially selected and trained enumerators conducting more intensive interviews can obtain more accurate age reports than are obtained in the census; likewise, careful investigation in a house-to-house canvass of deaths may yield more accurate reports on the ages of decedents than are to be found on the death certificates. Not only the accuracy of age returns, but also the reliability of reports on other characteristics of the population and of decedents can be investigated in this manner.

On the other hand, no matter how carefully information on ages may be obtained in the check enumeration or survey, it is always subject in some degree to errors of the same kind which affect the data being tested. In particular, willful mis-statements of age and errors due to utter ignorance of the persons concerned may affect the results of even the most carefully conducted check.

A better method of checking age returns is to match the census records or death certificates with the birth certificates of the persons concerned. In a large country with a mobile population this procedure may be very laborious and expensive, even for a small sample of the records. Nevertheless, this is the most objective check on the accuracy of age statements which can be made. In the birth registers, the date of birth will rarely be mis-stated and, if so, probably only by a few days at the most. The age of an individual, as computed from the date of his birth according to these records, is for all practical purposes exact. This type of check was conducted on the occasion of the 1921 census of New Zealand. The following is an excerpt from the report:[23]

> "...For the first time an attempt has been made to supply a more or less reliable measure of the extent and incidence generally of this assumed inaccuracy (i.e. of age statements). The *modus operandi* was of the simplest character, comprising little more than the checking of stated ages with birth registration records. A selection of names and persons of both sexes was chosen, following certain age-divisions. The method employed naturally restricted operations to native-born New Zealanders. Only those cases were chosen where the names and other particulars were such that certain identification was possible.
>
> "Altogether 2,524 cases were investigated, resulting in the actual testing of 2,219; the remaining 305 cases either could

[22] United States, Bureau of the Census. 1950 Census of Population, *United States Summary,* Bulletin P-C 1, p. xiii, Washington, 1953.

[23] New Zealand, Census and Statistics Office, *Results of a Census of the Dominion of New Zealand taken for the night of 17th April 1921. General Report.* Wellington, 1925, pp. 93-96.

not be located, or, for one reason or another, could not be identified beyond doubt.

"The actual sample (2,219) is not a large one, consisting as it does, of only 0.24 per cent of the total native-born. That it is not larger is due to the considerable labour and inconvenience which larger numbers would have involved. Nevertheless the results show much consistency and are of no little value.

"......

"The following summary gives some of the facts in brief and shows that there exists a slight tendency to overstate the age of minors, whether male or female, and that mis-statement is much more common with adults. Adult females are not greatly less accurate than males in the matter of stating ages; where, however, there is no defined tendency in the male figures, those for females show a decided preference for understatement."

	Adults		*Minors*	
	Number	*Per cent*	*Number*	*Per cent*
Males:				
Correctly stated	392	76.41	567	94.82
Understated	62	12.09	3	0.50
Overstated	59	11.50	28	4.68
Females:				
Correctly stated	352	68.75	568	95.30
Understated	115	22.46	7	1.18
Overstated	45	8.79	21	3.52

The ages reported in the census death records may also be compared with the ages or dates of birth stated in other records, for example, social insurance records. Likewise, for a sample of persons enumerated at one census, it is possible to examine ages returned at another census, either prior or subsequent. This was done in New Zealand for persons reporting advanced ages, and it was found that a large proportion of them had exaggerated their ages.