CHAPTER II. THE COMPLETENESS OF VITAL STATISTICS

A. Introduction

The term "vital statistics" is often used to include births, deaths, marriages and divorces. For the purposes of population estimates, it is mainly the data on births and deaths that are important. Accordingly, in this chapter methods for appraising births and deaths only will be considered. Some of the methods to be described here are adaptable, however, to tests of the completeness of marriage and divorce reporting.

Effects of errors in vital statistics on population estimates

There is an important difference between errors in vital statistics and errors in census enumerations, as regards their effect on population estimates. The difference can be illustrated by referring to those annual estimates of population for the years following a census which are made by adding to the census total the recorded annual excesses of births over deaths. If the census figure is in error—due to under-enumeration, for example—the annual estimates will reflect the same error, but the error will be a constant. The whole series of estimates will be comparable with the census and the picture of population growth since the date of the census will not be invalidated. Errors in the vital statistics, on the other hand, may have a cumulative effect. For example, if there is an under-registration of births every year, so that the recorded annual natural increase is too small, the consequent error in the population estimates will grow larger in each successive year. By the end of a ten-year period, both the estimated size of the population and the estimated amount of its growth since the census date may be considerably below the truth, even though the amount of error in the birth statistics for any one year is not very great.[1]

It is also important to note that the methods of collecting vital statistics are essentially different from those involved in collecting population data by means of a census. A census is taken at infrequent intervals, generally once in ten years. The data are usually collected by a staff of enumerators especially selected for the purpose, who visit the households and obtain the information by direct interview. Vital statistics, on the other hand, are usually recorded continuously by officials who carry this responsibility over a period of years; these officials are not ordinarily expected to visit the people in their areas and make direct inquiries, but only to receive and record the reports of vital events which, by law, are to be brought to their attention. The fact that vital statistics are collected continuously is an advantage in appraising them, since certain tests can be applied which are not possible for a one-time operation such as the census. On the other hand, perhaps the main problem in obtaining complete reporting of vital events is the fact that some one other than the registrar must make a positive effort to report the event.

Coverage

In many countries, registration of births and deaths is not compulsory, or is compulsory for only a part of the population (for example, for "Europeans" in certain African territories), or for certain areas within the country. Obviously, in such cases registration is not complete for the entire country even though it may be nearly so for those parts of the country or the population to which it applies. If the areas or segments not covered by registration are relatively small, then it is possible to estimate how many births and deaths would have been reported if registration were applied to the entire country. Such estimates are sometimes needed if a balancing equation is made. It is then necessary to distinguish between the terms "completeness" and "coverage"; the former means the extent to which all births and deaths are reported where the registration system applies, whereas the latter term means the extent to which the registration system applies to the entire population.

In some areas there is a system of "voluntary" registration. Such a system results in the registration of only a fraction of all births and deaths. There is no known instance of voluntary registration of vital events that has even remotely approached completeness.

Time and place of occurrence

When vital statistics are used for making population estimates, or for the construction of a balancing equation to test the accuracy of demographic data, it is important that they should reflect accurately the numbers of births and deaths occurring in a given time interval. Many countries tabulate and publish their vital statistics according to the time of registration rather than according to the time in which the event occurred. Where there is a very short time lag between time of occurrence and time of registration the effect on annual totals is negligible; but in some countries a lag of two years is not uncommon.[2] Where the number of events is changing significantly from year to year or where an appreciable time elapsed between occurrence and registration, the results may be misleading as to the number of vital events which actually occurred during the specified year. This consideration is particularly important in relation to statistics of births; deaths are ordinarily more promptly reported. Therefore, as a first step in the proper appraisal of vital statistics where tabulations are by date of registration

[1] United Nations, *Manuals on methods of estimating population. Manual I: Methods of Estimating Total Population for Current Dates,* Population Studies No. 10, pp. 38, 39.

[2] See for example, Ricardo, Jimenez J., *Comparative Study of the 1950 Population Census Results and the 1950 Population Estimate of Costa Rica,* San José, Costa Rica, December 1951.

instead of date of occurrence, the factor of delayed registration should be examined.

Vital statistics are often tabulated according to the areas in which the events occurred rather than the areas in which the persons concerned were residents. Very often hospitals serve persons living in a wide radius outside the city in which the hospital is located. Many births and deaths occur in such hospitals, to residents of areas outside the city in which the hospital is located. On the other hand, when vital statistics are used as materials for population estimates, which are usually on a *de jure* basis, that is they refer to the population *resident* in a given area, it is desirable that they should reflect the births and deaths to residents of the areas for which estimates are made. For this purpose, tabulations are required, in which the births and deaths of non-residents are re-allocated to the areas where they reside. The difference between tabulations by place of occurrence and place of residence is especially important where calculations are being made separately for urban and rural areas. In the statistics for an entire country, of course, place of occurrence and place of usual residence will almost always coincide, but the same may not be true of data for component regions.

Characteristics recorded on the certificates

Almost universally when births are registered the sex of the infant and the age of the mother are recorded, and when deaths are registered the sex and age of the deceased are noted. Information on various other personal characteristics is often also contained in birth and death records. For the purposes of population estimates and tests of the accuracy of vital statistics and census counts, reliable data on sex and age are of particular importance.

Sex is probably very rarely mis-stated, either at the census or in vital registers. Mis-statement of sex is perhaps most frequent in the case of birth records, but even in this case is probably rare. If the ratio of reported numbers of male and female births is unusual, the reason is probably not so much mis-statement of sex as a less complete reporting for one sex than for the other.

The ages of parents of a new-born child, or the age at death of a deceased person, may be frequently mis-stated. Considerations relating to the accuracy of age statements are presented in chapter III. If age mis-statements are similar in the vital records and census enumerations, then the censuses and vital statistics are still comparable. It is sometimes believed that ages are more accurately reported in the registration of vital events than in censuses. If the frequency and nature of age mis-statements in vital records differ greatly from those at the census, then the two sets of statistics are to that extent inconsistent and caution is necessary if they are used jointly.

It is thought that other characteristics, such as occupation, for example, may be returned more correctly on the census schedule than on the vital certificate.

Factors affecting the completeness of birth registration

As already pointed out, responsibility for the reporting of vital events rests with the public, usually with the relatives of new-born infants or deceased persons. Hence, more than in a census enumeration, co-operation of the public is decisive in achieving a complete record of births and deaths. The imposition of fees for the registration of births may tend to discourage registration. This may be particularly important among poor people, if the registration fee amounts to a significant proportion of their daily earnings. In countries imposing such fees temporary removal of the registration fee in order to encourage registration may result in a sharp increase in the number of recorded births during the period in which no fees are charged.

Social customs sometimes may lead to non-registration. In areas where a stigma is attached to illegitimate births, there may be a deliberate attempt on the part of the mothers to avoid reporting them, or to report them as legitimate. If the births are reported as legitimate, of course, population estimates and the balancing equation are not affected, but any analyses of marital fertility will be somewhat in error. In some countries also, children of one sex may not be considered quite as important as those of the other sex, in which case there may be a tendency to report less completely births of the less favoured sex. The fear that boys will be conscripted for military service may sometimes lead to the non-registration or late registration of their births.

A rural population thinly dispersed over a wide area may find it physically difficult to reach the registrar's office in order to report a birth. The consequent tendency to neglect registration may be aggravated if the means of transportation are deficient, if illiteracy is prevalent, if the people see no positive inducement to register the births, and if the births most frequently occur at home without a physician in attendance.

The population living in an urban area or within easy access to the registrar's office is more likely to register births. Indeed, registration is probably most complete among urban, literate populations in which most of the births occur in a hospital and practically none occur without a physician in attendance. It is much easier to ensure that a few hospitals or physicians will report all births than it is to ensure that every individual family will report on its own initiative.

Any system of family allowances based on the number of children should increase the completeness of registration since the payment of the allowance with respect to any child can be conditioned upon the registration of the birth. Any requirement that a person give documentary proof of his age, his parentage, or that he was born in the country or is a citizen of the country for inheritance or other legal purposes will also act to improve the completeness of registration.

The rationing of consumers' goods is another factor tending towards complete registration. Indeed, it may even lead to over-registration. This may be particularly true in those situations where an infant is entitled to as much of the rationed item as an older person, or where the rationed item is particularly needed by children. For example, sugar may be rationed at so many units per person irrespective of age, or milk may be limited exclusively to children; in such a case it may be advantageous to the family to have extra births registered. If the rationed item is for the consumption of adults only, as tobacco or

coffee for example, there may be no incentive to over-register births.

Over-registration may also occur, in some cases, for other reasons; for example, duplicate reporting by the parents through ignorance of the fact that the birth has already been registered without their participation or re-registration when an original certificate of registration is lost. Normal precautions such as birth registration indexes will minimize such duplicate registration, but in some cases the precaution may not be adequate.

Factors affecting death registration

Some of the same factors which operate to reduce the completeness of birth registration may also tend towards incomplete death registration. In particular, among a rural population thinly dispersed over a wide area, in which there are few doctors available so that many of the deaths occur without a physician in attendance, registration is likely to be incomplete. Under such circumstances it is comparatively easy to bury the body privately; if the burial has been accompanied by whatever religious and social ceremonies are deemed important by the people concerned, they may feel no need to register the death.

The requirement that a death must be registered in order to obtain a burial permit is a very important aid to complete death registration, particularly in areas where most bodies are attended to by a relatively small number of undertakers, and where most burials take place in but a very few cemeteries. Death registration may be more complete in urban than in rural areas if for no other reason than that it is likely to be easier to control burials in urban areas.

In areas where most deaths are attended by a physician, complete registration is likely to be more easily achieved than elsewhere. From this point of view also, the situation in urban areas is likely to be more conducive to good registration than that in rural areas. At least the causes of death will certainly be reported more accurately if a physician is in attendance.

Certain legal factors may be conducive to complete registration. For instance, inheritance or other legal claims involving succession or the right of a widow to re-marry may require proof of a death which can be supplied by a death certificate. The State is also interested in deaths in which a crime may be involved. Thus there is some inducement to register deaths in order to avoid entanglements with the law which might result from non-registration.

The rationing of consumers' goods, on the other hand, may have a tendency to encourage non-registration or delayed registration of deaths. If the family of the deceased can continue to collect his rations by failing to report the death, there is obviously some incentive to avoid reporting.

Relative completeness of birth and death registration

For the purposes of population analysis and particularly population estimates, it is highly desirable to know whether the births or the deaths in a given country are more completely reported. If births are reported more fully than deaths, the computed natural increase will be too high, and if the reverse occurs it will be too low. On the other hand, if both are registered with about the same degree of completeness, the recorded natural increase may not be greatly in error even though there is some omission of both births and deaths.[3]

In spite of the importance of this question, the statistical reports of few countries give any indication as to which of the two events are believed to be reported more fully. An indication to the effect that both births and deaths are reported "fairly completely", or that the registration of both events is "markedly deficient", does not suffice to resolve this question.

B. Internal consistency of vital statistics

Some clues as to the possible degree of completeness of reporting of births and deaths are afforded by examination of the data themselves. The use of the procedures to be described in this section will not, by themselves, establish the quality of the registration; they may serve, however, to reveal possibilities of error which deserve further investigation. Procedures for measuring the completeness of registration more exactly will be presented further on in this chapter.

Numbers of reported births, deaths and infant deaths

In some areas, birth rates or death rates cannot be computed because sufficiently reliable population estimates are not available. Sometimes this is the case for an entire country; more frequently, it applies to a country's geographical or civil divisions. Valid comparison can then be made only of the various absolute numbers of vital events. Even where population estimates are realistic enough for a computation of valid birth or death rates, a comparison of the mere numbers of reported births and deaths is a very useful first step in an examination of the accuracy of the statistics.

In countries where birth registration is known to be accurate, male births invariably outnumber female births by a slight margin. For biological reasons, the sex-ratio of births can vary only within rather narrow limits. Usually, there are very nearly 105 male births for every 100 female births, though some variation in this ratio is possible. A ratio lower than 102, or higher than 107, while not impossible, must be regarded as rather unusual. Any significant deviation from these limits should be regarded with suspicion. In unusual circumstances, a slight deviation from this pattern might occur, but if the deviation is at all large it is practically certain that the births of one sex are less completely reported than those of the opposite sex and, very probably, births of both sexes are incompletely reported. If, on the other hand, the ratio of male to female births happens to be very near 105 per 100, birth reporting may still be deficient for both sexes.

Ordinarily, when a population consists of more or less equal numbers of men and women, the deaths of males slightly outnumber those of females. The exact ratio of male to female deaths depends on a variety of

[3] See: United Nations, Population Division, Manual I: *Methods of Estimating Total Population for Current Dates,* Population Studies, No. 10, chapter VI.

factors which cannot be fully considered for the purpose of such a simple comparison. Nevertheless, one should expect that variations in the ratio of male to female deaths among a country's administrative divisions should tend to conform to a systematic pattern; if variations are unsystematic, or if the ratio in some particular areas deviates greatly from the norm without any apparent reason, then the accuracy of death registration is open to suspicion. The comparison of numbers of male and female deaths may also be extended to individual age groups. As a general rule, more males than females die in infancy, childhood, and adolescence, but more females than males die in old age and, at least in some countries, also during the child bearing age.[4] Proceeding from one age group to another, one should expect the ratios of male to female deaths to change gradually and never abruptly. Sharp fluctuations in this ratio between one age group and another may indicate either incomplete recording of deaths of certain sex-age groups, or mis-statement of ages of the deceased (see also chapter III).

Among infants, male deaths almost invariably exceed female deaths. This is not only because slightly more boys than girls are born, but because the boys are generally subject to a higher rate of mortality. The exact ratio of male to female infant deaths may vary in accordance with the circumstances. In countries where infant mortality is high, this ratio is usually, though not necessarily, within the range of 110 to 120 male per 100 female deaths. In countries with low infant mortality, the ratio is often of the order of 130 to 140 male per 100 female deaths. These are only rough standards, but they may help to determine in some cases whether inaccuracy of infant death registration should be suspected. Large and unsystematic differences in this ratio among various subdivisions of a country likewise serve as a warning that registration may be inaccurate at least in some areas.

Other numerical relationships can also be utilized. For example, the ratio of infant deaths to the total of deaths at all ages usually exhibits a considerable stability, at least so far as the various segments of any one country are concerned. This ratio will be high—of the order of one in four, or even one in three—in countries of high fertility and high mortality, and low—about one in ten, or lower still—in countries of low fertility and low mortality but, so far as the various segments of a country or consecutive years are concerned, the ratio should not fluctuate unreasonably. Finally, one may compare the ratio of all births to all deaths. This ratio, which has sometimes been used for purposes of demographic analysis, can vary rather widely, but may also serve to detect irregularities so far as particular areas within a country are concerned.

Levels of rates

Vital rates[5] are known to be functions of the health conditions and other social and economic conditions of the population.[6] Accordingly, the rates for a given country should be examined in the light of the conditions known to exist there, and of the results of researches on the relationships between these conditions and vital rates. Such an examination can indicate very roughly the ranges within which the rates should probably be contained, and point the way to a more precise investigation of possible errors if the recorded rates seem to differ too much from those that are expected. It must be borne in mind in this connexion that erratic rates may be produced by error either in the numbers of births and deaths reported, or in the population figures which form the bases for the rates, or both.

Birth rates

Crude birth rates (number of live births per 1,000 population) much lower than 15 or higher than 50 have rarely been observed where data were known to be accurate. If the observed rate falls below 15, the completeness of registration is open to question; if it is over 50, the two possibilities should be considered: that there was over-reporting of births or that the population figure was understated.

The most probable range of birth rates, given certain economic and social conditions, can be defined more narrowly. Birth rates in the range between 35 and 45 per 1,000 are usually considered as "high", those in the range between 25 and 35 as "moderate", and those between 15 and 25 as "low". These limits should, of course, be interpreted somewhat liberally. High rates are characteristic of most areas of the world where social customs and relationships for the great majority of the population have been little affected by the changes associated with industrialization according to the modern pattern. Low rates are generally found in those countries which have undergone a profound social transformation as a result of urbanization, industrialization, a rising level of popular education, and related factors. Moderate rates tend to prevail in areas where the types of conditions mentioned are intermediate between the two extremes.

This generalized statement, however, is over-simplified; in each case other factors must be considered in determining within what range the birth rate should be expected to fall.

The components of the birth rate, as well as the crude rate for the entire country, should be examined in detail. For example, almost universally among those countries having fairly complete birth registration, it has been found that the birth rate in urban areas is lower than that in rural areas.[7] Therefore, the rates for the urban and rural areas of a country should be studied separately. If the rural rate turns out to be lower than the urban rate (even after allowing for possible differences in age composition and marital status by means of standardizing techniques), the completeness of registration in the rural areas is open to question. In this connexion it will be noted that it is very important that births be

[4] The reason for more numerous female deaths in old age is not a higher female mortality at such ages, but the survival of larger numbers of women than men up to these ages.

[5] The "vital rates" considered here are crude birth rates (live births per 1,000 inhabitants per year), crude death rates (deaths per 1,000 inhabitants per year), and infant mortality rates (infants dying at less than one year of age per 1,000 births during the same year).

[6] United Nations, *The Determinants and Consequences of Population Trends,* Population Studies, No. 17, especially chapters 3 and 4.

[7] For the results of studies on the various relationships mentioned in this and succeeding paragraphs, the reader is referred to the United Nations report mentioned above, *The Determinants and Consequences of Population Trends,* especially chapters 3 and 4.

tabulated according to the usual place of residence of the mother. If many rural women go to urban areas for confinement, and if their births are credited to the urban areas, the urban rate may be higher than the rural even though nearly all births to rural women have been registered.

The birth rates in the various subdivisions of a country should be examined in detail. After they are standardized for possible differences in age composition, the rates in the various subdivisions should be consistent with the social, economic and public health conditions in the various areas.

The birth rates for important ethnic groups or other components of the population (aborigines, nomads, immigrant groups of distinctive culture and living conditions et cetera) should also be studied in the same manner.

It is also useful to compare the level of the birth rate with the census statistics on age composition of the population. In general, a high birth rate is reflected by a high percentage of children in the population and a low rate by a low percentage of children, unless the relation is distorted by major migratory movements.

Consistency of crude birth rates with age structure

It might appear at first thought that the birth rate could easily be checked directly against the number of infants enumerated in the census. Upon closer examination, however, this matter presents considerable complications. Infant enumeration is incomplete in most censuses. Furthermore, the number of reported "infants" (that is to say, those aged less than 12 months) can be much affected by slight mis-statements of age. Finally, only in countries having excellent registration systems is the rate of infant mortality known exactly.

A useful test, called the "forty per cent test", has been devised by Wertheim.[8] As demonstrated by its author, a population of which 40 per cent or more are aged less than 15 years is most likely to have a birth rate of at least 40 per 1,000. The birth rate might be as low as 39, with moderate infant mortality. As is generally known, infant mortality is usually high in populations with very high birth rates.

This test is applicable where only crude age statistics are available. Experience has shown that in most countries where the censuses show categories such as "adults" and "children" but not the exact ages of the populations, the mean age at which persons begin to be reported as "adults" is usually in the vicinity of 15 years (though, as a rule, slightly lower for females and slightly higher for males). Many such countries have, in fact, birth rates of the order of 40 per 1,000; variations around this level are probably reflected—however approximately—by variations in the percentage of persons aged less than 15, at least among population groups which are similar in other respects.

Death rates

Crude death rates based on accurate statistics are seldom much less than 8 per 1,000 population, except among populations of abnormal age composition. This is not to say that a lower rate cannot be correct, but rather that if the rate is lower than 8 and if this fact cannot be explained by an abnormality of age structure, the completeness of registration or the accuracy of the population base figures should be investigated.

In areas where the statistics are accurate, death rates considerably in excess of 30 per 1,000 in normal years are unusual. Normal death rates as high as 35 per 1,000, or even higher, are possible if fertility is so high that the population can be reproduced despite such heavy losses. In years of calamity (for example, famine, epidemic, or mass destruction by war), the death rate may, of course, rise much higher. Except in the case of such catastrophes, a recorded death rate in excess of 35 per 1,000 suggests that the statistics should be examined for possible errors.

As in the case of birth rates, the range of expected death rates in a given area can be specified within narrower limits when the conditions affecting public health in that area are taken into consideration.

One further observation is relevant. High birth rates occur in conjunction with either high or low death rates. Low birth rates, on the other hand, are normally found only in association with rather low death rates. Hence, if the recorded birth rate is found to be of the order of, say, 15 to 25 per 1,000, the death rate would not normally be expected to be much lower than 10 per 1,000 nor higher than perhaps 15 per 1,000. If the recorded birth rate is low and the death rate either extremely low or higher than about 15 per 1,000, the accuracy of the statistics should be examined.

Tests for the completeness of death reporting are substantially similar to those described above for testing birth reporting. The data can be analysed for the various subdivisions of a country, or the various population groups, and the rates related to the social, economic and health conditions of these several segments.

In relating the death rates to these other factors it is necessary to take into consideration the age composition. If a population contains relatively few older people it will have a lower death rate, all other factors being the same, than a population containing relatively many older persons. This can be illustrated by comparing data for Israel and Norway. The crude death rate in the former country was 6.4 (in 1951) which might seem suspiciously low. In Norway (1950) the rate was 9.1. The population of Norway, however, was older than that of Israel. It is possible to standardize the Israel age-sex specific death rates on the Norwegian population so as to find out what the death rate in Israel would be if its population had the same age distribution as that of Norway. Standardizing in this manner yields a death rate of 10.3. In short, the difference in the crude death rate between these two countries was accounted for to a considerable extent by differences in age structure.

Infant mortality rates

The level of the infant mortality rate, in general, is related to the level of the total death rate. Countries having high general mortality rates also have high rates of infant mortality. It is impossible, however, to assign any specific value to the infant mortality rate which

[8] W. F. Wertheim, "La population de l'Indonésie et le test des 40%", *Population*, 1954, No. 4, pp. 655-674.

should accompany a given crude death rate. A preliminary appraisal of the infant mortality rate can be made by considering its level in the light of the social, economic and public health conditions of the country. If a rate seems unusual in relation to the known conditions, the accuracy of the vital data can be considered suspect.

The infant mortality rate is calculated by relating the number of infant deaths, generally under one year of age, to the number of births. Hence, it is affected by whatever errors there may be in the registration of both these events.

In this connexion it is well to note that in some countries infants who do not survive a minimum length of time, as one day or two days or three days, are not included in either the birth or death registration data. The effect of such a procedure is to lower the reported infant mortality rate, and make appraisal of it somewhat more difficult. For example, let us assume that 1,000 births are reported of which 100 died in infancy. The infant mortality rate is then calculated as 100 (100 deaths per 1,000 live births). If there were 20 infants born who died during the first day and were not registered, so that 20 should be added to both the numerator and the denominator, the rate becomes 117.6 per 1,000. Fortunately, for the purpose of population estimates, this error is of little importance since the additional number of births is almost instantly cancelled by the same additional number of infant deaths. In an investigation of birth rates and infant death rates, however, this circumstance can produce misleading results.

The infant mortality rate may also be affected if the age of the infant is not correctly reported. It is possible that the deaths of some children who are not quite one year old may be reported as deaths above the age of one year, but it is also possible that some children who have died after their first birthday may still be reported as infant deaths. If either of these two errors is more frequent than the other, the number of infant deaths is accordingly under-reported or over-reported.

A useful test consists in an examination of numbers of infant deaths by months of age, if such statistics are available. Deaths are by far most frequent during the first month, and especially the first few days, of life; in subsequent months mortality declines continuously. Ordinarily, one should expect the number of deaths in the first month to exceed significantly the number in the ensuing five months which again should be substantially larger than the number of deaths between 6 and 12 months of age. Beginning with the second month, the progression should be a fairly smooth one.

The sources of error and ambiguity in statistics relating to infant mortality, and methods of appraising the quality of the data, have been described in another publication.[9]

Trends in crude rates

It is expected that there will be year-to-year variations in the birth, death and infant mortality rates. Where these yearly variations are very large, however, there is reason to suspect the accuracy of the statistics. In Bolivia, for example, the recorded birth rate rose from 14.7 in 1940 to 28.5 in 1941 and has fluctuated around 30 per 1,000 ever since; the death rate as recorded rose from 4.9 in 1940 to 19.4 in 1941 and has since been within the range of 15 to 20 for most years. Such extreme variations normally do not occur; in this case they are explained by the fact that Bolivia introduced compulsory registration in 1940. It can be deduced that there was considerable under-reporting prior to 1941; how complete the reporting has been from 1941 to date has yet to be determined.

Sometimes, however, a rather sharp change in a vital rate may occur as a result of a major change in the relevant conditions of life. An example is found in the statistics of Japan. In 1946 the reported crude birth rate was 25.3 and in 1947, 34.3. The Japanese birth rate is known to have fallen during the years of the Second World War, and in particular during those years when many of the men were in the armed forces outside the country. With the termination of the war in 1945 and the repatriation of the armed forces as well as of many other Japanese men who had been abroad, the birth rate rose rapidly.

Another example is found in the Netherlands, where the death rate was 11.8 in 1944, 15.3 in 1945, and 8.5 in 1946. In this case, mortality had been rising gradually during the years of the Second World War, as living conditions deteriorated. At the end of 1944 and the beginning of 1945, the Netherlands were the scene of military operations which had devastating effects, causing a considerable temporary increase in mortality. By 1946, conditions were again relatively normal and the death rate returned to its pre-war level.

It should be noted that where the births and deaths are tabulated by year of registration rather than year of occurrence, large fluctuations in the rates may result from variations in the numbers of delayed registrations. One further point should be noted. The obtaining of fairly complete reporting of all births and deaths which occur is a slow process. After a compulsory registration law is introduced which applies to all the inhabitants, and even after the necessary administrative mechanism is instituted, several years may be required before satisfactory reporting of all the events is ensured. A steadily rising trend in crude birth or death rates over a period of several years is then likely to indicate improvement in coverage rather than a rise in fertility or mortality.[10]

Patterns of death rates, by age and sex

Another clue as to the possible completeness of registration of deaths is afforded by examination of the patterns of death rates by age and sex. In those countries in which all the evidence has proven quite conclusively that registration of deaths is virtually complete, there appears an almost standard pattern of rates by age and sex. This typical pattern is illustrated with data for Sweden (see table 3 below). The essential points to note are as follows:

(1) The rates for males, in each age group, are higher than those for females;

[9] United Nations, Population Branch, *Foetal, Infant and Early Childhood Mortality*. Volume I. *The Statistics*. ST/SOA/Series A/13, New York, 1954.

[10] See also United Nations, *Demographic Year Book, 1952*, p. 27.

(2) The rates are quite high in the age group under one year, after which they rapidly decrease to a low point after age 10. From this minimum, the rates rise slowly at first and at the oldest ages increase very rapidly. Thus there is a *steady and smooth progression* describing a U-curve.

In reviewing the age-sex pattern of any country, if deviations from the above typical pattern are observed they should be explained in terms of known peculiarities; otherwise errors in the statistics are to be suspected. For example, in some populations the death rate for women is higher in the reproductive ages than it is for men in the same age groups, because of unusually high maternal mortality associated with a very high birth rate. In some countries, tuberculosis affects the survival of adolescents and young adults to such an extent that the death rate rises fairly rapidly until about age 20, to remain at a fairly constant level until about age 30. Any deviations that are not explainable imply either that the death reporting (number of deaths and/or age of deaths) is deficient, or that the age data obtained from the population census (or population estimates) are in error. Tests for attempting to determine the accuracy of the age data will be shown in chapter III.

Some of the kinds of deviations from the normal pattern of sex-age specific mortality rates which may be found in the statistics for various countries, are illustrated by the figures for Egypt and the Moslem populations of pre-partition Palestine in table 3.

Table 3

DEATH RATES BY AGE AND SEX, FOR SWEDEN (1948), EGYPT (1937) AND PALESTINE MOSLEMS (1944)

(Deaths for 1,000 persons in specified age-sex class)

	Sweden		*Egypt*		*Palestine Moslems*	
Age	*Males*	*Females*	*Males*	*Females*	*Males*	*Females*
All ages	10.0	9.7	29.5	25.0	17.3	16.8
Under 1 year	27.4	19.4	257.6	211.2	112.8	108.5
1 to 4 years	1.5	1.2	85.5	72.7	33.0	38.3
5 to 9 years	0.7	0.5	8.5	6.7	4.6	3.6
10 to 14 years	0.6	0.4	5.1	3.9	3.4	2.3
15 to 19 years	1.3	0.7	5.8	4.2	3.6	2.5
20 to 24 years	1.8	1.1	8.1	4.5	4.3	4.7
25 to 29 years	1.7	1.1	8.8	5.4	8.0	7.6
30 to 34 years	2.0	1.5	10.5	7.3	6.8	5.8
35 to 39 years	2.4	2.1	10.5	7.2	6.5	6.5
40 to 44 years	3.2	2.5	13.7	9.6	7.1	4.8
45 to 49 years	4.9	3.9	14.0	8.2	10.0	5.2
50 to 54 years	8.0	6.2	21.9	13.9	8.8	5.8
55 to 59 years	12.2	9.1	22.1	12.5	10.9	9.3
60 to 64 years	19.1	15.2	37.9	22.1	22.1	13.5
65 to 69 years	30.8	25.9	45.3	25.3	30.2	24.3
70 to 74 years	48.7	44.5	76.4	50.5	44.4	37.8
75 to 79 years	81.1	74.4	103.7	74.7	—	—
80 to 84 years	136.3	129.5	192.6	144.9	—	—
85 and over	235.8	222.5	566.4	578.1	—	—
75 and over	—	—	—	—	103.9	119.7

The death rate for all ages in Egypt is about double that of Sweden. As in Sweden, the rates for males are higher than those for females, age for age, except in the very oldest age group. The rates for males in the successive age groups exhibit no deviations from the expected steady and smooth progression of a U-curve. In the case of female rates, minor deviations are noted; between the ages of 30 and 65 the progression is not smooth. Whereas the expected pattern is that of a steadily rising rate, the Egyptian data show:

Age	*Rate*
30 to 34 years	7.3
35 to 39 years	7.2
40 to 44 years	9.6
45 to 49 years	8.2
50 to 54 years	13.9
55 to 59 years	12.5
60 to 64 years	22.1

Intensive testing is necessary to determine whether the dips at ages 35 to 39, 45 to 49 and 55 to 59 reflect primarily errors in the death statistics or in the population figures by age used for computing the rates.

The data for the Palestine Moslems (table 3 above) reveal greater discrepancies. The male rates are lower than those for females at three age groups—1 to 4 years, 20 to 24 years, and 75 years and over; only the deviation at the 20 to 24 year group might be explainable in terms of high maternal mortality. The patterns by age, for both males and females, are much more irregular than those observed in Egypt. Among both sexes the patterns between ages 25 and 59 are highly irregular. For example, it is most implausible that women aged 45 to 54 years have lower mortality than women aged 25 to 29 years, or that men aged 40 to 44 years have lower mortality than men aged 25 to 29 years. The age-sex rates for this population strongly suggest the need for intensive investigation into the completeness of death reporting, as well as the accuracy of the population figures.

C. The use of balancing equations

The use of balancing equations has been described in chapter I. It was noted that the balancing equation alone does not permit a final appraisal of any one of

the sets of statistics involved, that is, census counts and statistics of births, deaths and migration.

In the present chapter, balancing equations are made for certain segments of a population, thereby concentrating attention on the errors which arise from inaccuracies in the vital statistics. Again it is to be noted that any one such equation, by itself, is insufficient to determine exactly from which set of statistics the major part of the discrepancy originates. Nevertheless, as the terms of the equation are changed, additional evidence is gathered and the conclusions, which were at first tentative, may be either confirmed or disproved, at least to a certain extent. Every such appraisal is relative only, and the degree of accuracy of each set of statistics can in no case be determined with absolute certainty by these equations alone. Absolute appraisal would be possible only if at least one set of statistics were known to be perfectly accurate or its error were known precisely. Even so, balancing equations can often determine with sufficient assurance whether a given set of statistics is probably reasonably accurate or not.

Appraisal of death statistics by means of balancing equations

It is obvious that the population above a certain age enumerated at a recent census must have already been alive at the time when the census previous to it was taken. This part of the population would have aged by the number of years which have passed between the two census dates and, during the interval, would have been diminished by a certain number of deaths and otherwise modified as a result of migration. Births which occurred during the interval do not affect this segment of the population,

For reasons to be explained in chapter III, the comparison of the two populations is likely to be more precise if children under 5 years of age at the earlier census are not included. Let it be assumed that censuses have been taken in 1940 and 1950. The balancing equation then becomes:

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

P_1 = population *15* years of age and over in 1950, P_0 = population *5* years of age and over in 1940, D = number of deaths in 10 year period to persons who were 5 years and over in 1940, I = number of immigrants who were 5 years of age and over in 1940, E = number of emigrants who were 5 years of age and over in 1940.

By setting up the equation in this way, the factor of births is eliminated so that the balancing equation can be used to appraise the death statistics, the migration data, and the census enumerations at ages above the stated limits. Thus a better appraisal of the death reporting is obtained than could be had if the balancing equation also included births.

It would be preferable also to eliminate the migration factor and this is possible for a country in which the census and vital statistics are presented separately for the native population, if it is known that few of the native-born population emigrate; in such a case the equation becomes:

$$P_1 = P_0 - D$$

where P_1, P_0 and D refer to native-born individuals only.

In some countries it may be known that net migration of both native and foreign-born is very small. In this case it is not necessary to tabulate the data by nativity. The balancing equation can be applied using only the total count of persons 5 years of age and over in the earlier census, the number 15 years of age and over in the following census, and the total number of deaths during the decade to persons who were 5 years of age or over at the time of the earlier census.

The number of deaths to persons aged 5 and over at the time of the earlier census is obtained by adding together the deaths at ages 5 and over in the census year, those at ages 6 and over in the following year, those at ages 7 and over in the subsequent year, and so forth. A refinement of the calculation is possible if the exact date of the census is taken into account. If the census was taken at the beginning of the year, then some of the persons dying at age 5 in that year were less than 5 years old at the date of the census. If the census was taken near the end of the year, only a small portion of the deaths of that year could have occurred after the date of the census.

Great refinement of the computations, however, is unnecessary, unless the statistics are very nearly accurate. Otherwise, the slight modification in the estimated number of deaths which results from a refinement of the calculation can only be negligible as compared to the probable magnitude of the residual in the balancing equation. For most practical purposes, it is sufficient to utilize death statistics tabulated by 5-year age groups. It is enough to assume that all deaths occurring at ages 5 to 9 during the first year after the census was taken were the deaths of persons aged 5 to 9 at the census date; for the next year, it may be assumed that four-fifths of the deaths in this age group were to persons aged 5 to 9 at the census; two years later three-fifths; three years later, two-fifths; and four years later only one-fifth. In subsequent years, the deaths of persons aged 5 to 9 are no longer to be considered, but of the 10 to 14 year age group all deaths should be taken into account five years after the census, four-fifths of the deaths six years after, and so forth.

Example of Thailand

Age data are available from the censuses of 1937 and 1947, and deaths by age for the inter-censal period together with an estimate of the net volume of migration for the period, are also available. These data can be used to illustrate the above-described form of the balancing equation as follows:

Population 15 years of age and over according to census of 1947	10,068,000
Population 5 years of age and over according to census of 1937	12,027,000
Apparent inter-censal decrease	1,959,000
Number of reported deaths of persons who were 5 years of age and over in 1937	1,530,000
Migratory inward balance of persons who were 5 years of age and over in 1937	100,000
Total expected decrease (deaths minus immigrants)	1,430,000
Apparent inter-censal decrease	1,959,000
Population decrease accounted for	1,430,000
Residual unaccounted for	529,000

The residual amounts to about 5.3 per cent of the population 15 years of age and over in 1947, and 27.0

per cent of the apparent inter-censal decrease. This amount is large enough to suggest that further investigation into the accuracy of the component parts is required.

If the residual were to be attributed to any one factor, this single factor could have been a greatly excessive enumeration in 1937, a greatly deficient enumeration in 1947, a large emigration not accounted for, or a considerable deficiency in death registration.

Excessive enumeration at one of the censuses does not seem very probable. Census enumerations may have been deficient, but there is no obvious reason why the 1947 enumeration should have been much less complete than the earlier one. In this connexion, it may be stated that examination of the census data by age and sex (see chapter III) yields no clues with regard to any major irregularity in the census enumeration. Whether, despite the available estimate of an inward migratory balance, there has in fact been a large emigration, cannot be decided in the absence of further evidence, but it would seem unlikely. Therefore—although some of these errors may have been present—it is probable that the major part of the error should be attributed to incomplete registration of deaths.

One estimate of the completeness of death registration is arrived at by assuming that the two censuses were either complete, or deficient to the same extent, and that the migration balance is substantially correct. In this way the total number of deaths which occurred is estimated as the number reported plus the residual; or a total of 2,059,700, of which 1,530,000, or 74 per cent were reported.

On this basis, the average crude death rate of Thailand for the period 1937-1947 inclusive should be raised from the recorded level of 16.6 to 22.4. This rate is comparable with an average death rate of 21.0 in Malaya during 1932-1940, when death registration is believed to have been fairly accurate.

On the assumption that the censuses differed with respect to completeness of enumeration, other estimates of the completeness of death registration would be obtained. If it were assumed that there was a larger under-enumeration in the 1947 census, the residual attributed to under-registration of deaths would be reduced, and the conclusion would be that death registration was somewhat more than 74 per cent complete. On the other hand, if there were thought to have been a larger amount of under-enumeration in 1937 than in 1947, the residual due to incomplete death registration would be larger and the estimate of completeness of registration would be less than 74 per cent. By making various assumptions as to the completeness of the two censuses, based on whatever information can be obtained on that score, it is possible to arrive at maximum and minimum estimates of the completeness of death registration.

Example of Puerto Rico

The procedures can also be illustrated with data for Puerto Rico as follows. Censuses were taken on 1 April, 1940, and 1 April, 1950. For the intervening years data are available on deaths and migrants.

Population 15 years of age and over according to census of 1 April 1950	1,255,300
Population 5 years of age and over according to census of 1 April 1940	1,588,900
Apparent inter-censal decrease	333,600
Reported deaths of persons who were 5 years and over in 1940	157,800
Migratory outward balance of persons who were 5 years of age and over in 1940	172,000
Expected decrease (deaths plus emigration)	329,800
Apparent inter-censal decrease	333,600
Decrease accounted for	329,800
Residual	3,800

The residual would be less than 3,800 if deaths which occurred among men in the armed forces had not been excluded from the death statistics employed here. Even the residual of 3,800 amounts to only 0.3 per cent of the enumerated 1950 population 15 years of age and over.

These results suggest that the death statistics, as well as the two population censuses and the migration data, are fairly accurate. It is possible, of course, that an under-reporting of deaths was compensated by faulty migration data. This result could have been produced by an over-count of emigrants or an under-count of immigrants. However, since the migratory balance is computed from statistics of total civilian arrivals and departures (so that no problem of distinguishing between migrants and other travellers, such as visitors, is involved), neither of these possible errors is likely to be important. Puerto Rico is an island, and most of the migration occurs by air via the one central airport; it is unlikely that many persons leave from or arrive at the island without being recorded, and the possibility of duplication of records in any large number of cases is remote.

With respect to the remaining two elements in the balancing equation, the census counts and the death statistics, the fact that the equation does balance out with but a very small residual suggests that both these elements are quite accurate also. This is not proven beyond doubt, since there could be compensating errors among these two sets of data. If the size and direction of any errors in the census were the same at both dates, it follows that the reporting of deaths was quite complete. If the 1950 census was more complete than that of 1940, it follows that deaths were under-reported by a corresponding amount. This possibility could be checked in part, by making the same comparisons for the 1940 and 1930 censuses, and for the 1950 and 1960 censuses when the results of the latter become available. If the equation should balance out as well for the other two inter-censal periods as it did for the decade 1940 to 1950, the evidence would be quite conclusive that the censuses and death reporting were nearly complete.

Example of Portugal

The censuses of 1940 and 1950, together with mortality data and estimates of net migration for the inter-censal decade, permit the construction of this version of the balancing equation for Portugal:

Population 15 years of age and over according to census of 1950	5,953,100
Population 5 years of age and over according to census of 1940	6,890,500
Apparent inter-censal decrease	937,400
Reported deaths of persons who were 5 years and over in 1940	787,900
Migratory outward balance of persons who were 5 years of age and over in 1940	64,000

Expected decrease (deaths plus emigration)	851,900
Apparent inter-censal decrease	937,400
Decrease accounted for	851,900
Residual	85,500

In this case the residual amounts to 1.4 per cent of the population 15 years of age and over enumerated in 1950, and 9.1 per cent of the inter-censal decrease. This residual is smaller than that observed in Thailand and greater than that found in Puerto Rico. Further investigation of the components parts is required before any definitive statement can be made about the completeness of death reporting. Perhaps the migration data are in error; perhaps the 1950 census was more nearly complete than the 1940 census; perhaps death reporting was incomplete.

If it is assumed that the two censuses were equally complete and the migration data substantially correct, the residual is accounted for by unreported deaths. On these assumptions, the true number of deaths which occurred is estimated at 873,400, of which 787,900 or about 90 per cent, were reported. Other estimates can be obtained in the same manner as was shown for Thailand previously.

Appraisal of birth statistics by means of balancing equations

Another form of the balancing equation is particularly suited to direct attention to an evaluation of birth statistics. In this instance, the procedure consists essentially in a comparison of the number of children below a certain age enumerated at the census with the number of births during the years immediately preceding the census, allowance being made for the deaths and migratory movements by which this child population has been affected prior to the census. The balancing equation can then be stated in the form:

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

where P_1 = number of children enumerated, under a specified age (for example, under 5 years), B = number of births reported during an equivalent number of years prior to the census, D = number of deaths among children born during these years, up to the time of the census, I = number of immigrants (children born during the specified period), E = number of emigrants (children born during the specified period).

In computing the number of deaths which have occurred to children born during the given period, statistics of deaths by age are required. It is of some importance to refine this calculation because deaths are far more frequent to children in their first year of life than in subsequent years. Usually more than two-thirds of the deaths under one year of age occur during the first six months of life, and a disproportionately high percentage of these deaths occur to children aged less than one month. This consideration has a bearing on the ratio which should be selected in order to estimate that number of infant deaths which should be attributed to children born in the earliest year included in the calculation.

Example of Thailand

The above balancing equation can be applied to Thailand with the following results:

Population under 5 years of age according to census of 1947	2,644,000
Births reported in 5 years preceding census date	2,513,000
Deaths reported to children born during these 5 years	386,000
Expected survivors under 5 years of age to be enumerated in the 1947 census	2,127,000
Number actually enumerated	2,644,000
Residual	517,000

This calculation indicates that there was under-reporting of births. It is possible, but very unlikely, that the census count exaggerated the number of children. In fact, in almost every census, young children are under-enumerated at least to some extent, and if the Thailand census figure for this age group was in error, it almost certainly erred on the low side. It is hardly possible, on the other hand, that several hundred thousand young children immigrated; some of the residual, but probably only a small part, could be accounted for by migration.

How incomplete might the birth reporting have been? If it is assumed that the number of children reported by the census was correct, and the number of child deaths was fully reported, one estimate of the completeness of birth reporting can be computed as follows: to the number of births as reported, is added the amount of the residual, giving an estimated total of 3,030,000 births, and an estimate of 83 per cent for the completeness of birth registration.

Another estimate can be arrived at in the following manner: The number of children under 5 years of age reported in the census is probably too small; let it be assumed, purely for illustrative purposes, that 90 per cent of the children were enumerated. Then the true number of children under 5 living in the country at the census date would be 2,930,000. With reference to deaths, the calculations earlier in this section of the chapter suggested as one possibility that death reporting was 74 per cent complete. If it is assumed that the same percentage applies to the deaths of children under 5 years of age, the true number of deaths for this age group is estimated at 521,000 (386,000 divided by 0.74). Adding the estimated number of children to the estimated number of deaths gives 3,451,000 as an estimate of the number of births, which implies that birth registration was 73 per cent complete (2,513,000 divided by 3,451,000 = 0.73).

On this basis, the average crude birth rate for Thailand for 1942-1946 would be raised from the recorded 30.7 to an estimated 42.1. The latter rate is comparable with that for Malaya during 1947-1951, namely 42.6, where birth registration is believed to have been very nearly complete.

This result cannot be taken as a reliable estimate of the completeness of birth reporting. The calculations were made only to illustrate the methods; better estimates might be obtained by investigating other available information bearing on the completeness of the census enumeration and of death registration, or by applying direct checks. The probability is very high, however, that not all children under 5 years of age were enumerated in the census of 1947, and that not all deaths to infants born in the 5 years preceding the census were registered. Therefore it is not likely that birth regis-

tration was as much as 83 per cent complete—the figure first derived.

Example of Puerto Rico

For illustrative purposes, let us use the balancing equation to test the completeness of registration of births in the year preceding the census date. The calculations are:

Population under 1 year of age according to census of 1 April 1950	71,000
Births reported during preceding 12 months	85,500
Deaths reported to children born during this period	4,600
Expected surviving children under 1 year of age to be enumerated in the census	80,900
Number actually enumerated	71,000
Residual	9,900

It should be noted that data on migration are missing from this balancing equation. It is known that there was considerable net outward migration from Puerto Rico, but the number of babies born in the period 1 April, 1949 to 1 April, 1950 who emigrated before the latter date is unknown; it is probably reasonable to assume, however, that this number was very small and could have accounted for but a very small portion of the residual of 9,900.

There might have been a significant under-reporting of infant deaths. This balancing equation proves nothing about the quality of death statistics. However, the results of the balancing equation used earlier for estimating the completeness of death reporting suggested that deaths were reported quite completely for the population which was 5 years of age and over in 1940. The presumption is therefore strong that reporting of infant deaths was also at least fairly complete.

The above reasoning leads to the conclusion that the census enumeration of infants was probably deficient, and that birth registration may very well have been nearly complete. The census deficiency could have taken two forms: complete omission of some infants from the head count and enumeration of some as one year of age or more. In either case, the result is an understatement of the number of infants. The question becomes: how much under-enumeration was there? If the under-enumeration were of about the size of the residual, it would follow that birth reporting was very nearly complete. However, if the amount of under-enumeration were considerably greater than the residual, it would be clear that a significant proportion of the births had not been registered. The under-enumeration of infants under one year of age in Puerto Rico in the 1950 census can be estimated, when the 1960 census has been taken, by the methods to be explained in chapter III.

APPLICATION TO THE SUBDIVISIONS OF A COUNTRY

As has been noted in chapter I, balancing equations may be applied to the population of parts of a country as well as to the population of the entire country. The usefulness of balancing equations is diminished, however, by the fact that migration between the various areas of a country is usually important and that satisfactory statistics on internal migration can rarely be found. Hence, in many circumstances, the balancing equations for parts of a country provide insufficient evidence, particularly on the quality of death statistics. As a matter of fact, in countries where the census figures and birth and death statistics are supposed to be of fairly high quality, such equations are sometimes used for the purpose of estimating the balance of migration rather than detecting errors in the vital and census statistics. On the assumption that the errors are negligible, the discrepancies are ascribed to the factor of migration.

The limitations of the use of the equation are not so strict if the census provides data on the population not only in accordance with the present place of residence, but also with residence at the time of the previous census. In that event, the earlier population within each administrative division can be compared with essentially the same population, at a more recent date, as modified by deaths in the intervening period and by migration beyond the country's national boundaries. Census tabulations of the type mentioned are rare, however. More commonly, censuses provide tabulations of the population according to place of birth, or according to place of residence at some specified earlier time (but not the date of the last previous census). In that event, the exact balance of the population between the two census enumerations cannot be established, but the census data furnish a certain amount of evidence regarding the direction and volume of migratory movements within the country between the two census dates.

The complications relating to internal migration are at a minimum when the balancing equation is used to test the completeness of birth registration for subdivisions of the country, in the same manner as explained above for the whole country. When births in the year preceding the census enumeration are compared with the number of children under one year of age enumerated in the census, the conclusion is not necessarily invalidated because the relatively small migration of infants, during a single year, can be disregarded.[11] However, this procedure suffers from the disadvantage that the completeness of infant death registration in the various subdivisions cannot be determined by the balancing-equation methods; unless some independent information on this score can be obtained, the estimates of variations in completeness of birth registration from area to area may be subject to considerable error.

The direct checks to be discussed in section D of this chapter are the best means of determining the completeness of vital statistics registration. If the application of direct checks has been limited, because of the cost, to certain parts of the country, it is still possible to estimate the completeness of birth and infant death reporting in those areas which were not directly checked by a procedure analogous to that of the balancing equation. The calculations are shown in table 4 below, with a hypothetical example. It is assumed that direct checks were made in subdivision A but not in subdivision B; therefore, the problem is to ascertain the probable com-

[11] If it can be assumed that death reporting is nearly complete in all the subdivisions, it is possible to estimate and make allowance for internal migration. In this way it is possible to compute the completeness of birth reporting by subdivisions using the numbers of children under 5 years of age enumerated in the census, and allowing for internal migration. See, A. J. Jaffe, *Handbook of Statistical Methods for Demographers*, Washington, United States, Bureau of the Census, 1951, pp. 145-148.

Table 4

PROCEDURE FOR ESTIMATING LIMITS OF COMPLETENESS OF BIRTH REPORTING FOR A COUNTRY'S SUBDIVISIONS (hypothetical data)

	Subdivision A	*Subdivision B*
1. Reported number of births, in year preceding census	110	350
2. Correct number of births as revealed by a direct check	120	(no check)
3. Reported number of infant deaths	14[a]	35
4. Expected number of survivors at census (uncorrected figures, item 1 less item 3)	96	315
5. True number of survivors at census date in subdivision A (item 2 less item 3)	106	—
6. Children under one year of age enumerated at census	98	313
7. Ratio of true number to number enumerated, subdivision A	1.082	—
8. Expected number of children under one year of age at census date in subdivision B, assuming equal completeness of enumeration in both subdivisions (that is to say, 313 times 1.082)	—	339
9. Estimated number of births, subdivision B, assuming infant death reporting was complete (item 8 plus item 3)	—	374
10. Estimated completeness of birth registration in subdivision B (item 1 as percentage of item 9) (per cent)	—	94

[a] It is assumed that the direct check revealed this figure to be correct.

pleteness of birth and infant death reporting in subdivision B. It must be emphasized that this procedure only provides approximate answers, and that a direct name-for-name check affords a more accurate appraisal of the vital statistics.

The simple assumption used in table 4 can be varied. For example, on the basis of the calculations for several areas where the direct checks of birth statistics were carried out, it may be possible to estimate a range within which the correction factor for census under-enumeration of children (item 7) in subdivision B is almost certainly contained. If this range is, say, between 1.06 and 1.10, the expected number of children (item 8) falls in the range between 332 and 344. Adding 35 (item 3) we obtain the range from 367 to 379, and the completeness of birth registration (item 10) is then estimated in the range of 92 to 95 per cent.

Furthermore, the results of the checks of infant death registration in various areas may provide a basis for estimating its completeness in subdivision B. If it is estimated in this manner, for example, that between 80 and 90 per cent of infant deaths in subdivision B were registered, there would be, instead of 35, some 39 to 44 actual infant deaths, to be added, in item 9, to the range of 332 to 344 estimated for item 8. Thus the estimated number of births is placed in the range from 371 to 388, indicating that birth registration was 90 to 94 per cent complete.

D. Direct checks on completeness of vital statistics registration

The method of direct checks consists of securing an independent list of children born, or persons who died, during a particular period and in a particular area, and then checking name for name with the registrar's files to determine whether these events were registered. The simplest, but not necessarily the most accurate estimate of under-reporting is that proportion of the births or deaths on the independent list, for which no certificate of registration can be found. The procedures for a good estimate of the completeness of registration may be more complicated in some circumstances.

Lists of names of births and deaths can be secured from a wide variety of sources including church records, names reported in newspapers and other publications, and house-to-house inquiry. These names can be collected from the entire population or from a sample; in the latter event it is important that the sample be representative of the entire population living in the area. Any effort to cover the entire population of a country in order to determine the number of births and deaths which have occurred during a specified year, for example, is very expensive; indeed, its cost may be as great as that of a population census. The minimum sample which will provide reliable results is most economical and practical.

So far as births are concerned, the names of infants listed on the census schedules constitute a valuable basis for a check on registration.

In many respects these procedures are very similar to those described in chapter I for a direct check of the accuracy of a census enumeration, but a check on vital statistics can be carried on at any time whereas direct checks of the census enumeration are best conducted immediately after the enumeration is finished.

It is evidently desirable that the independent list be as nearly complete as possible; however, as will be explained below, it is not essential that the list be perfectly complete in order to obtain good estimates of the completeness of registration.

In matching the names on the independent list with the names of registered births or deaths, the events are tabulated in three categories:

(A) Matched events, that is to say, those recorded on both lists;

(B) Events found on the independent list but not registered;

(C) Events registered but not found on the independent check list.

In addition, unless the independent check list is known to be perfectly complete, a fourth category (which we shall designate as category D) must be considered, namely, events not included in either list, which actually occurred and should have been registered.

The sum of A + B represents all events found on the check list; the sum of A + C all events found in the register; and the sum of A + B + C + D all events which in fact occurred. A perfect measure of the completeness of registration would be the ratio of A + C to A + B + C + D.

If the independent list is very nearly complete, very few (if any) events should be found in category C and it may be presumed that category D is very small. The categories C and D can then be neglected, and the ratio of A to A + B is a good estimate of the completeness of registration.

If a sizeable number of events of both categories B and C is found, it may be presumed that the number of category D is also substantial. This is the situation which is usually to be expected. One estimate of the magnitude of D can be made on the assumption that, in the check list, the probability of inclusion of an event is the same for registered events as for those which escaped registration. On this assumption, the ratio of C to C + D would be the same as the ratio of A to A + B, or of A + B to A + B + C + D. The ratio of A to A + B then still constitutes a good estimate of the completeness of vital registration.

As a rule, however, the probability of omission from the check list will be greater for unregistered than for registered events. The reasons why certain events do not reach the registers are likely to apply also in some degree to the omissions from the coverage of the house-to-house inquiry or other source of the check list. The ratio of C to C + D will therefore usually be smaller than the ratio of A to A + B, so that the latter is a maximum estimate of the completeness of registration. The error of the estimate is likely to be the greater, the larger the numbers of cases in categories B and C. If the probable size of category D is large enough to have a substantial effect on the estimate of the completeness of registration, refined methods are necessary in order to arrive at a realistic estimate of completeness.[12]

[12] For a description of refined procedures, as applied in a test of the birth and death registers in an area of Bengal, where vital registration is known to be greatly deficient, the reader may refer to C. Chandra Sekar and W. E. Deming, "On a Method of Estimating Births and Death Rates and the Extent of Registration," *Journal of the American Statistical Association*, Vol. 44, No. 245, March 1949, pp. 101 ff.

Birth registration test in Puerto Rico, 1950

The above procedures can be demonstrated with data obtained in a test of birth registration which was carried out in Puerto Rico. At the time of the 1950 population census special cards were filled out for all infants enumerated in the census, who had been born during the preceding three months. These cards contained the name of the infant, the names of the parents, and such other information as would be useful for identification purposes. The cards were then matched with the records of births registered during the same period.[13]

The figures for the municipality of Adjuntas, for example, are as follows:

Infants' cards matched with births registered.....	200
Infants' cards for which there were no births registered..	15
Total	215

The completeness of birth registration was thus estimated at 93 per cent (200 divided by 215).

The results of this test for the entire island of Puerto Rico revealed that in the three months preceding the census of 1 April 1950, birth registration was about 95 per cent complete. The presumption is that birth registration had been approximately as complete for some time in the past, and would probably continue to be at least as complete in the future. In comparing this finding with that obtained by using the balancing equation (as described earlier in this chapter), it is seen that the tentative conclusion arrived at by means of the equation are substantiated by means of the direct check. The direct name-for-name check not only provides a more precise appraisal of completeness for the island as a whole, but also permits estimation of completeness of registration in each of the 77 municipalities, which is not at all feasible with the balancing equation.

It should be noted that this type of check can be carried out any time after a census has been taken, so long as the census schedules and the registration certificates are available.

[13] For further information on the steps involved in matching the records see Sam Shapiro and Joseph Schachter, "Methodology and Summary Results of the 1950 Birth Registration Test in the United States", *Estadística* (Journal of the Inter-American Statistical Institute), December 1952, Vol. 10, No. 37.