Chapter IV

ACCURACY OF ESTIMATION

The extraction of approximate birth and death rates from censuses of varying completeness, in which age, parity and other relevant data are inaccurately or incompletely reported, can scarcely be expected to produce figures of great precision. Moreover, the range of error in the estimates about the true figure cannot in general be determined at all exactly. The purpose of this chapter is to call attention to the imprecision of estimates based on data of poor quality, even when they have been made with the help of elaborate tables and adjustments; to give some rough impression of the magnitude of error that is routinely encountered; and to distinguish the forms of estimation most subject to large errors from those less susceptible.

The sources of error discussed are of two principal kinds: errors caused by a discrepancy between actual and assumed conditions, and inaccuracies in the basic data. Crucial assumptions that may be wrong are that the age pattern of mortality in a given population conforms to a family of model life tables, and that the age composition of a population has a form stable in the sense of Lotka. The kinds of inaccuracy that affect almost all forms of estimation are omission of persons from censuses and surveys and age-misreporting. The effects of these forms of imprecision are illustrated in this chapter principally by synthetic estimates in which each source of error is assumed to operate in the absence of others.

A. Differences between assumed and actual conditions

1. *Errors arising from differences between the actual age pattern of mortality and that embodied in the model life tables*

There is no conclusive way of delineating the errors that might arise because populations with incomplete records may have age patterns of mortality that differ from those with full records. Even when there are recorded data from which age schedules of mortality can be calculated, it is often uncertain whether an extreme pattern of mortality—e.g., unusually low infant mortality, given the prevalent death rates above age one—is genuine, or the product of unusual inaccuracy in the data rather than an unusual pattern of death rates.

In this section the errors that originate in age patterns of mortality different from the model life tables of annex I are illustrated by examples in which estimates based on alternative families of model life tables are compared. The four families are described in chapter I. The distinctive age patterns of mortality they embody are examples of differences found among populations with especially accurate data, and undoubtedly do not nearly exhaust the variety of patterns to be found among all populations. Nevertheless, it can safely be assumed that forms of estimation that yield nearly identical figures in all four families are insensitive to age pattern differences, and are on that account preferable to forms of estimation that yield divergent figures. The divergence itself can be taken as a minimum index of the uncertainty of estimation associated with variations in age patterns; it is obvious that variations *at least* this large do occur.

The effects on various estimates of different age patterns of mortality will be illustrated by utilizing the four families of model life tables to make calculations for a population that is assumed recorded without error. The "test" population is a "West" female stable population with an expectation of life at birth of forty years, a gross reproduction rate of 3.00, and mean age of fertility of twenty-nine years. The various characteristics of this population can be obtained from the appropriate table in the model stable populations. To illustrate the importance of the assumption that the population has a given age pattern of mortality, estimates are made with all four age patterns. The variation among the four figures is the significant result: the fact that the estimate based on the "West" tables always agrees with the "true" figure is of course a purely automatic consequence of how the example is formulated.

(a) *Effects of assumed age patterns of mortality on estimates derived from census survival rates*

In section A of chapter I, the reader can find a description of how to select a model life table consistent with the numbers recorded in two censuses taken at a ten year interval: project the first population by life tables at various levels of mortality and by interpolation find the level that matches the recorded total over ten, over, fifteen etc. It is suggested that the median of the first nine levels so indicated is a sensible choice. By summing the products of the age specific death rates from the life table so determined and of the average intercensal number of persons in the corresponding age groups one obtains an estimate of the death rate. This death rate added to the average annual rate of growth during the intercensal period yields an estimate of the birth rate. The procedure at no point makes use of the assumption of stability.

TABLE 5. EXPECTATION OF LIFE AT AGES 0 AND 5 IN VARIOUS FAMILIES OF MODEL LIFE TABLES PRODUCING A PROJECTED POPULATION OVER AGE x MATCHING THE TEST POPULATION AT THE END OF A DECADE

Age x	*Expectation of life at birth* "*West*"	"*North*"	"*East*"	"*South*"	*Expectation of life at age 5* "*West*"	"*North*"	"*East*"	"*South*"
10	40.0	40.5	36.3	37.2	49.7	49.8	50.1	50.6
15	40.0	38.6	35.5	35.0	49.7	48.6	49.6	49.2
20	40.0	37.7	35.6	34.5	49.7	47.9	49.6	48.9
25	40.0	37.5	35.9	34.3	49.7	47.8	49.8	48.8
30	40.0	37.6	36.2	34.0	49.7	47.9	50.0	48.6
35	40.0	37.7	36.4	33.8	49.7	47.9	50.1	48.5
40	40.0	37.7	36.8	33.8	49.7	47.9	50.4	48.5
45	40.0	37.5	37.4	34.1	49.7	47.8	50.7	48.7
50	40.0	37.3	38.1	34.7	49.7	47.7	51.1	49.0
55	40.0	37.3	39.1	35.6	49.7	47.7	51.7	49.7
60	40.0	37.6	40.3	37.0	49.7	47.9	52.3	50.5
65	40.0	38.1	41.6	38.6	49.7	48.2	53.1	51.5
70	40.0	38.7	42.9	40.3	49.7	48.6	53.8	52.6
75	40.0	39.1	43.8	41.9	49.7	48.9	54.3	53.6
Median of first 9	40.0	37.7	36.3	34.3	49.7	47.9	50.1	48.8

TABLE 6. PARAMETERS ASCRIBED TO THE TEST POPULATION (WEST MODEL STABLE, $^0e_0 = 40$, GRR = 3.00), BY SELECTING THE MEDIAN LEVEL MODEL LIFE TABLE FROM EACH FAMILY FROM AMONG TABLES MATCHING THE PROPORTIONS SURVIVING IN TWO CENSUSES

Estimated parameter	*Estimated by median model life table* "*West*"	"*North*"	"*East*"	"*South*"
$^0e_0$	40.0	37.7	36.3	34.3
$^0e_5$	49.7	47.9	50.1	48.8
l_2	.773	.778	.701	.706
l_5	.725	.704	.654	.629
Death rate	.0234	.0252	.0276	.0290
Birth rate	.0445	.0463	.0487	.0501
Death rate of population under age 5	.0720	.0769	.0976	.1019
Death rate of population over age 5	.0137	.0149	.0137	.0146

Suppose that our hypothetical population ("West" model stable female population, $^0e_0 = 40$ years, GRR $= 3.00$, $\bar{m} = 29$) were enumerated at the beginning and end of a decade, and projections made employing various levels of the four sets of model life tables.

Table 5 shows $^0e_0$'s and $^0e_5$'s in the model life tables that produce the numbers over age 10, 15, 20 etc. in the "actual" population at the end of the decade. Table 6 shows various parameters that would be ascribed to the test population by assuming that the median life table indicated in the last row in table 5 represented the population's mortality schedule.

The differences in age pattern among the four families produces estimates of over-all mortality (estimated death rate and expectation of life at birth) and mortality under age five (l_2, l_5, and death rate under five) that are much more divergent than the estimates of mortality in the population 5 and over $^0e_5$ and death rate over five). The uniformity diminishes among estimates based primarily on survival at the older ages. As a consequence of the differences in the estimated population death rates, arising primarily from differences in the estimated mortality under age five, the estimated birth rates differ by some 5.6 points or 11 per cent of the largest estimate. Note that when there is no evidence of the age pattern of mortality nor separate indications of child mortality, the assumption of the "West" pattern of mortality produces low (or conservative) estimates of birth and death rates.

(b) *Effects of assumed age patterns of mortality on estimates derived from stable populations chosen on the basis of* C(x) *and* r

If the test population conforming exactly to the West model stable with $^0e_0 = 40$ years and GRR $= 3.00$ were enumerated twice, the intercensal rate of increase could be calculated, and model stable populations found matching the given population in the proportion under age x and in the rate of increase. Table 7 shows the expectation of life at birth and at age five in model stable populations

TABLE 7. EXPECTATION OF LIFE AT AGES 0 AND 5 IN MODEL STABLE POPULATIONS BASED ON VARIOUS FAMILIES OF MODEL LIFE TABLES THAT MATCH THE TEST POPULATION IN PROPORTION UNDER AGE x AND IN THE RATE OF INCREASE

	Expectation of life at birth				*Expectation of life at age 5*			
Age x	*"West"*	*"North"*	*"East"*	*"South"*	*"West"*	*"North"*	*"East"*	*"South"*
5	40.0	42.2	36.8	39.5	49.7	50.9	50.4	52.1
10	40.0	41.3	36.1	37.9	49.7	50.3	50.0	51.1
15	40.0	40.4	35.7	36.9	49.7	49.7	49.7	50.5
20	40.0	39.6	35.5	36.3	49.7	49.2	49.6	50.1
25	40.0	39.2	35.4	35.8	49.7	48.9	49.6	49.7
30	40.0	38.9	35.5	35.3	49.7	48.7	49.6	49.4
35	40.0	38.7	35.5	35.0	49.7	48.6	49.6	49.2
40	40.0	38.5	35.6	34.7	49.7	48.5	49.7	49.0
45	40.0	38.4	35.8	34.5	49.7	48.4	49.8	48.9
50	40.0	38.2	36.1	34.5	49.7	48.3	50.0	48.9
55	40.0	38.1	36.6	34.6	49.7	48.2	50.3	49.0
60	40.0	38.0	37.3	35.1	49.7	48.2	50.7	49.3
65	40.0	38.0	38.2	35.9	49.7	48.2	51.1	49.8
Median of first 9	40.0	39.2	35.6	35.8	49.7	48.9	49.7	49.7

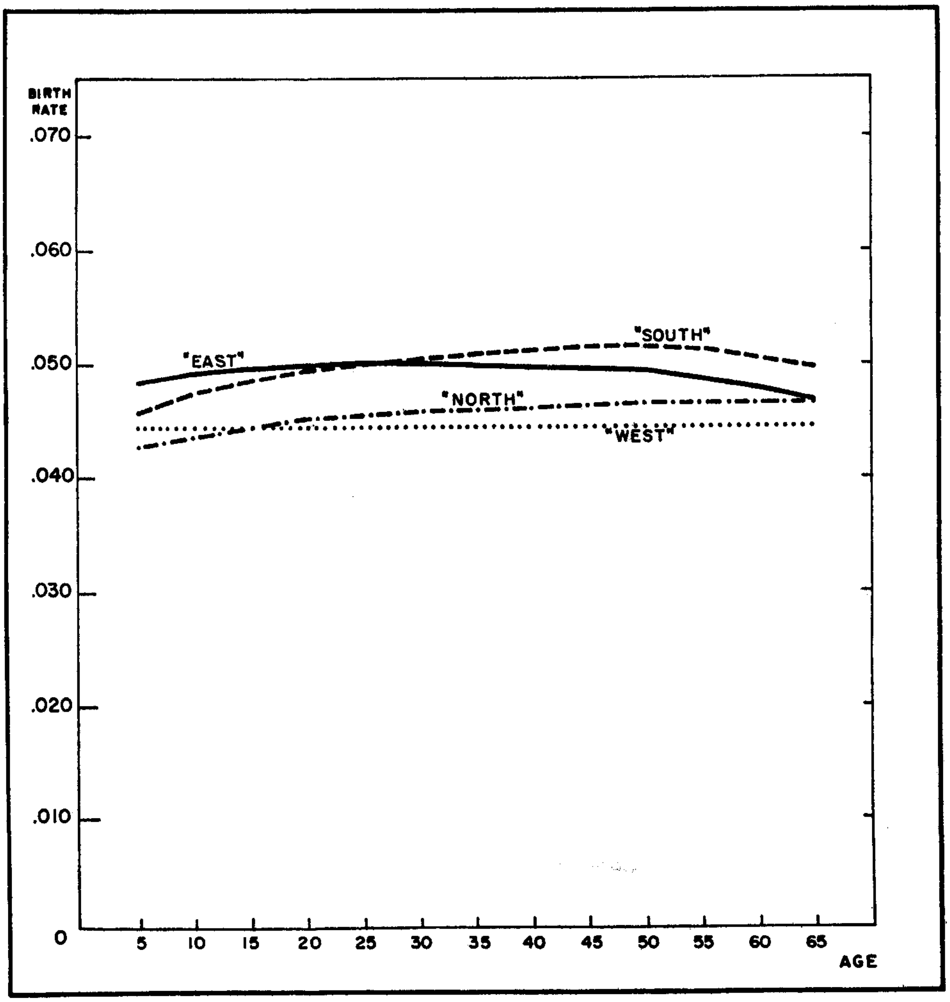

Figure XVII. Birth rates in the test population derived by the stable population method from $G(x)$ and r assuming various patterns of mortality as the appropriate one

TABLE 8. PARAMETERS ASCRIBED TO THE TEST POPULATION BY SELECTING THE MODEL STABLE POPULATION FROM EACH FAMILY WITH THE MEDIAN BIRTH RATE FROM AMONG THOSE WITH THE SAME r AND C(5), C(10), ..., C(45)

Estimated parameter	*Estimated from median model stable population*			
	"West"	*"North"*	*"East"*	*"South"*
$^0e_0$	40.0	39.2	35.6	35.8
$^0e_5$	49.7	48.9	49.7	49.7
l_2	.773	.790	.694	.719
l_5	.725	.720	.646	.646
Death rate	.0234	.0244	.0287	.0287
Birth rate	.0445	.0455	.0498	.0498
GRR ($\bar{m}$ = 29)	3.00	3.11	3.36	.342
Death rate of population under age 5	.0720	.0724	.1030	.0992
Death rate of population over age 5	.0137	.0144	.0136	.0138

(based on various families of model life tables) that duplicate the rate of increase and the proportion under age 5, 10, 15 etc. in the test population. Figure XVII shows the birth rate in these stable populations. Table 8 shows various parameters that would be ascribed to the test population by assuming that the model stable population with median fertility shown in figure XVII was representative of the test population.

The same features of variation in estimation are seen in these tables as in the preceding two. Again estimates of mortality above age five are insensitive to differences in age pattern while mortality estimates below age five and other measures strongly affected by infant and early childhood mortality, such as the over-all death rate or the expectation of life at birth, are markedly influenced.

(c) *Effects of assumed age patterns of mortality on estimates derived from reported child survival in combination with records of the age distribution originating from one census, or from two or more censuses*

Assume that the test population has been enumerated in a census that includes questions about the number of children ever born to each woman, and the number still alive at the time of the census, tabulated by sex of the child and age of the woman. By methods described in section B of chapter II, proportions surviving to age two and age three (l_2 and l_3) can be estimated for females, and (as outlined in section B of chapter III) these values can be used to select a model life table, and with C(5), C(10) etc., to select model stable populations at the indicated mortality level. Table 9 shows l_5, $^0e_0$, and $^0e_5$ selected on the basis of l_2 according to each of the "regional" patterns of mortality. Figure XVIII indicates the birth rate in the stable population based on these regional patterns with the given l_2 and the proportion under age 5, 10, 15 etc., in the test population. Table 9 also includes various parameters of the "median" model stable populations selected on the basis of the fertility shown in figure XVIII. Note that the patterns of estimated birth rates shown in figure XVIII are much more uniform than in figure XVII, and that, therefore, the estimation of the birth rate from $C(x)$ and l_2 is much more nearly independent of mortality pattern than when the estimates are based on $C(x)$ and r. On the other hand parameters shown in table 9 that measure mortality above age five are directly dependent on the assumed age pattern of mortality, since the only observed quantities relate to child mortality. Thus estimates of $^0e_5$ and the death rate of persons over five have a much larger range than in tables 6 and 8.

Now assume that the test population had been enumerated in two censuses a decade apart, and that the second census included the data needed to calculate l_2. It is now apparent that l_2 gives estimates of the death rate under age five that are less sensitive to age patterns of mortality than estimates based on age composition, and that on the other hand, estimates of the death rate over age five based on age composition (either census survival or $C(x)$ and r) are less sensitive to differences in mortality patterns. It is therefore recommended (cf. section C of chapter III) that parameters related to child mortality be estimated from l_2 and those related to adult mortality from the age composition; from census survival in the general case or from the intercensal growth rate and the (stable) age distribution when stability may be assumed. To illustrate the advantages of this procedure in terms of increased precision, and in particular to illustrate the increased insensitivity of the resulting estimates to differences in mortality patterns, the following calculations were made:

(a) *Death rates estimated from census survival (ages five and over), and from* l_2 *(ages under five), and* b *estimated as* d + r. A model life table was selected on the basis of best agreement between the projected population and the population recorded in the second census (the median level in table 5), and the $_5m_x$ values in this model table were applied to the recorded (test) population above age five to obtain the death rate over five. Death rates under age five ($_1m_0$ and $_4m_1$) from the life table selected on the basis of l_2 (cf. table 9) were applied to the test population at these ages. The expectation of life at birth was calculated as $_5L_0 + {}^0e_5 \times l_5/l_0$ where $_5L_0$ and l_5 were taken from the life table associated with l_2 and $^0e_5$ was taken from the life table obtained for the population over age five.

(b) *Death rates estimated from stable population (ages over five), and from* l_2 *(ages under five),* b *estimated as* d + r.

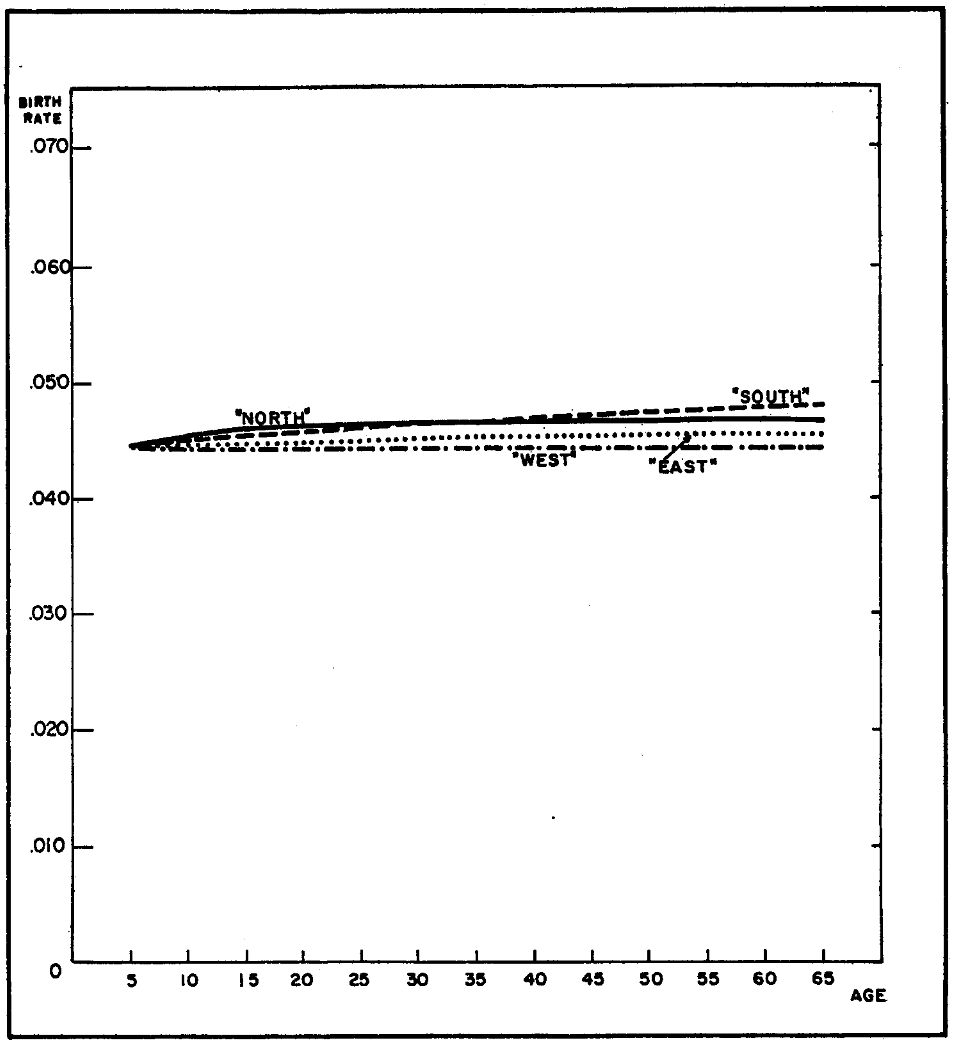

Figure XVIII. Estimated birth rates in the test population derived by the stable population method from C(x) and l_2 assuming various patterns of mortality as the appropriate one

TABLE 9. PARAMETERS ASCRIBED TO THE TEST POPULATION BY SELECTING THE MODEL STABLE POPULATION FROM EACH FAMILY WITH THE MEDIAN BIRTH RATE FROM AMONG THOSE WITH THE SAME l_2 AND C(5), C(10), ..., C(45)

	Estimated from median model stable population			
Estimated parameter	*"West"*	*"North"*	*"East"*	*"South"*
$^0e_0$	40.0	37.0	44.0	42.6
$^0e_5$	49.7	47.5	54.4	54.1
l_5	.725	.697	.736	.716
Death rate	.0234	.0265	.0207	.0221
Birth rate	.0445	.0465	.0451	.0461
Growth rate	.0211	.0201	.0244	.0240
GRR ($\bar{m}$ = 29)	3.00	3.18	3.07	3.19
Death rate of population under age 5	.0720	.0797	.0697	.0730
Death rate of population over age 5	.0137	.0154	.0109	.0113

The death rates for ages over five were obtained by selecting the life table underlying the stable population determined from $C(x)$ and r (the median level in table 7); otherwise the same procedure was followed as in (*a*).[1]

(c) *Birth rate estimated from stable population, death rate as* $b-r$. The birth rate was obtained from $C(x)$ and l_2 (table 9), and the death rate was calculated as the difference between this birth rate and the reported intercensal rate of increase.

Table 10 shows the population parameters calculated by these methods. Method (*a*) is applicable whether the population is stable or not, but (*b*) and (*c*) can be employed only with stable populations. The most striking feature of this table is the small variability in the estimate of such over-all characteristics as the birth and death rates and the expectation of life at birth caused by differences in age pattern of mortality when mortality above age five is derived from the age distribution and intercensal change and mortality under age five from reported child survival. It is clear how valuable is the supplementary information provided by data on proportions surviving among children ever born.

2. *Errors caused by non-stability of a population assumed to be stable*

If the age distribution of a population conforms closely to that of a stable population, estimation of many characteristics is greatly simplified, especially through the use of tabulated model stable populations. The methods of selecting an appropriate model population are given in chapters I and III; and in the preceding discussion in this chapter the errors that may arise because of variations in age patterns of mortality are explained. Another source of error is that the population in question may not in fact have the age distribution of a stable population because of age-selective migration, or past variations in fertility or mortality.

[1] In a less artificial example of the application of the stable population method the life table death rates would be applied not to the reported age distribution but to the *stable* age distribution as explained in section C.2 of chapter III.

Any extended discussion of the deviations in stability that can and do occur is beyond the scope of this *Manual*. Only a few comments about general principles and frequently encountered cases will be attempted.

Ideally, stable estimation should be employed only for a closed population with constant mortality during the preceding 25-30 years, and constant fertility for some two generations. A useful practical test is the absence of substantial change in age composition and of intercensal rate of increase in three consecutive quinquennial or decennial censuses. For example, examination of the age distributions in Turkey from 1935 on reveals clearly (in spite of conspicuous distortions caused by age–misreporting) that fertility was greatly reduced during certain periods since 1910: the evidence is a low point in the age distribution that does not remain at the same age from one census to the next as it would if age–misreporting were the cause of the low point, but rather advances by five years in each subsequent quinquennial census. In contrast, the Indian age distributions from 1891 to 1911, also irregular, are much the same in form, indicating that the irregularities are caused by age–misreporting, and that the underlying age composition was essentially constant (see figure XIX). Stable analysis is appropriate for India in 1911, but not for Turkey in the years shown.

Few populations for which estimation is necessary have been enumerated in an extended series of censuses of comparable quality, and it is often impossible to apply the suggested criterion of an essentially unchanging age distribution and rate of increase. The assumption of stability must often be made without much direct evidence in its support.

In general, stable methods of estimation should be attempted only in populations where there is no widespread use of birth control, since where the practice is common there are usually pronounced variations or

TABLE 10. ESTIMATED PARAMETERS OF THE TEST POPULATION CALCULATED BY VARIOUS METHODS BASED ON POPULATION AGE DISTRIBUTIONS FROM TWO CENSUSES AND FROM REPORTS ON CHILD SURVIVAL

		Estimated parameter					
Method of estimation	*Pattern of mortality*	*Death rate under age 5*	*Death rate over age 5*	$^oe_0$	$^oe_5$	*Birth rate*	*Death rate*
(*a*) Death rates from census survival (ages over 5) and from l_2 (under 5); $b = d+r$	"West"	.0720	.0137	40.0	49.7	.0445	.0234
	"North"	.0791	.0149	37.3	47.9	.0466	.0255
	"East"	.0697	.0137	40.8	50.1	.0441	.0230
	"South"	.0734	.0146	38.8	48.8	.0454	.0243
(*b*) Death rates from $C(x)$ and r (ages over 5) and from l_2 (under 5); $b = d+r$	"West"	.0720	.0137	40.0	49.7	.0445	.0234
	"North"	.0791	.0142	38.0	48.9	.0460	.0249
	"East"	.0697	.0140	40.5	49.7	.0443	.0232
	"South"	.0734	.0139	39.5	49.7	.0448	.0237
(*c*) Birth rate from $C(x)$ and l_2; $d = b-r$	"West"					.0445	.0234
	"North"					.0465	.0254
	"East"					.0451	.0240
	"South"					.0461	.0250

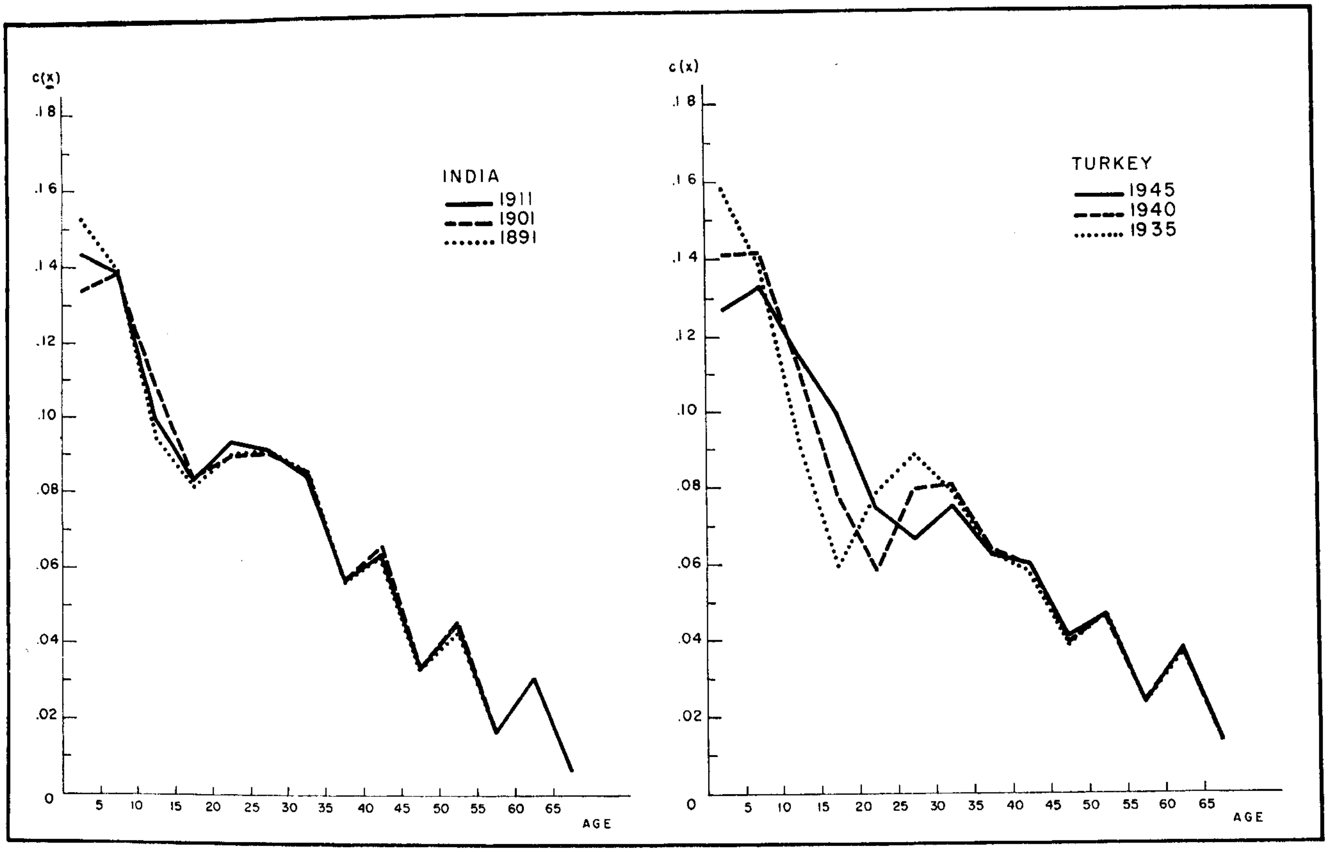

Figure XIX. **Distribution of the female population by age in five-year intervals as recorded in selected censuses in India and in Turkey**

trends in fertility. Stable analysis should be avoided in populations in which migration has had a pronounced influence on age composition—as is typical of the populations of many cities in developing countries. One reason for the emphasis on analysing female age distributions in this *Manual* is that female distributions are very often less affected by migration than are male. With some conspicuous exceptions (e.g., rural-urban migration in Latin America) female migration is usually less than male, and because women are usually accompanied by their children, the effect on the age distribution is sometimes negligible, even when migrants form a substantial fraction of the population. It has been shown for example that a constant stream of female immigrants, constituting annually 4 to 5 per cent of the receiving population, consisting of young adults and their children, does not produce an age composition markedly different from what would exist in the absence of immigration.[2]

If mortality and fertility have fluctuated rather than being constant, but without a regular trend, the stable population that has the the same ogive up to age fifteen, twenty or thirty (and the same rate of increase) has fertility and mortality close to the average during the past few decades. What may produce poor estimates are either major swings in fertility producing one or more consecutive small five-year cohorts, or a sustained trend in either fertility or mortality. A marked and continued decline in mortality has occurred in many populations that appear to have essentially constant fertility, and chapter I includes a section (section C) that shows how estimates based on stable populations can be adjusted for the effects of a history of falling death rates. But these adjustments can be made only if there are clues indicating approximately how long and how rapidly mortality has fallen. When there is no reliable basis for detecting the downward course of mortality, the adjustments given in table III.1 in annex III can be interpreted as indications of the errors that occur from stable population calculations made in ignorance of a downtrend in mortality. For example, suppose that a stable population is chosen with $C(30) = .7$, and ten-year intercensal $r = .020$, leading to an estimated b of .04912 and GRR_{29} of 3.312. But if in fact because of falling mortality population growth has increased from 0.144 in the next earlier intercensal decade and about .009 in still earlier decades the approximate value of k would be .01, and the value of t about 20 years. It can be seen, then, in table III.1 that the true value of the birth rate would be about 8.5 per cent higher, and the true value of the gross reproduction rate about 9.6 per cent higher than the estimates arrived at on the basis of the false assumption of stability. In other words the correct estimates for the birth rate and the GRR in this example are .0533 and 3.63, respectively.

[2] See J.M. Boute, S.J. , *La démographie de la branche indo-pakistanaise d'Afrique*, Études morales, sociales et juridiques, (Louvain, 1965). On this topic see also Léon Tabah and Alberto Cataldi, "Effets d'une immigration dans quelques populations modèles", *Population*, No. 4, 1963, pages 683-696.

Table III.1 can also be used as an approximate indication of the errors caused by the assumption of stability when fertility has recently followed a rising trend such that the GRR has risen by one per cent annually—been multiplied by $(1+k)$ in each of the last t years, where $k = 0.01$. The reason that table III.1 indicates the effect of a history of rising fertility is that the age distribution is displaced from the stable in an almost identical manner by a recent trend of falling mortality on the one hand, or of rising fertility on the other. If the sign of the adjustment factors in table III.1 is reversed, the factors then indicate the errors in stable estimates if in fact fertility has been *falling* by one per cent annually for t years.

It does not seem advisable to attempt to extend the use of table III.1 or to construct similar or more complicated tables, to cover estimation for populations in which the voluntary control of fertility is widespread. The application of stable analysis should probably be confined to populations that apparently do not practise deliberate contraception or abortion. But fertility may be subject to prolonged, if limited, upward and downward movements even when conscious birth control is rare. In tropical Latin America there appears to be little contraception or abortion except in some urban populations, but in many countries fertility was reduced in the depressed 1930s, and rose during the 1940s and 1950s.[3] The source of the variation was changes in marital status reflecting somewhat later marriage in the 1930s. In other populations proportions married increase when mortality falls because of a reduced incidence of widowhood, and in still others fecundity is changed by the spread or the conquest of pathological sterility. Table III.1 shows the orders of magnitude of the errors in stable estimates introduced by such trends, if not allowed for explicitly.

B. Errors caused by faulty data

The discussion of mistaken estimation caused by mistakes in the basic data will be confined to two important forms of defective information: omission of persons who should have been included in a census or survey (or the opposite mistake of erroneous inclusion), and misreporting of age. Distortions in the recorded age distribution are caused by age selective omissions as well as by age-mis-statement, but it is not generally possible to determine which of these factors has caused a given irregularity in age composition, and it will be implicitly assumed that omissions affect primarily the total number of persons enumerated, and that distorted age composition is caused by misreporting of ages.

1. *Differential rates of omission in consecutive censuses*

Consider a population enumerated in two censuses a decade apart, and suppose the coverage of the second census is 2 per cent more complete. What is the effect on estimates of mortality and fertility?

It is assumed (to maintain comparability with the earlier discussion) that the population is the same "test" population used before—the "West" female model stable, with $^{0}e_0 = 40$ years, and $GRR_{29} = 3.00$. Table 11 shows the levels of mortality required to project the earlier

TABLE 11. EXPECTATION OF LIFE AT BIRTH IN "WEST" MODEL LIFE TABLES PRODUCING A PROJECTED POPULATION OVER AGE x THAT MATCHES THE TEST POPULATION AT THE END OF A DECADE ASSUMING A 2 PER CENT RELATIVE OVERCOUNT OF THE TEST POPULATION AT THE LATTER DATE

Age x	$^{0}e_0$	*Age x*	$^{0}e_0$
10	44.5	50	42.4
15	44.6	55	42.1
20	44.2	60	41.7
25	43.8	65	41.5
30	43.5	70	41.2
35	43.2	75	41.0
40	43.0	80	40.9
45	42.7		

population to the later one at ages over 10, 15, 20 etc. The assumed conditions imply that a mortality level must be chosen producing a proportion surviving that is 2 per cent above the true numbers. It requires a bigger difference in mortality level to increase survival of the whole population by 2 per cent than of the population over fifty or sixty. Thus the biggest mistake in selecting a level of mortality when there is a change in the completeness of coverage occurs, as is evident in table II, in levels based partly on survival of the population at younger ages. However, as will be seen later, age-mis-statements make the projected number of persons derived from the population over forty especially unreliable.

TABLE 12. STABLE POPULATION ESTIMATES OF THE BIRTH RATE IN THE TEST POPULATION DERIVED FROM CORRECTLY REPORTED PROPORTIONS UP TO AGE x AND FROM THE OBSERVED INTERCENSAL GROWTH RATE WHEN THE LATTER IS DISTORTED BY A 2 PER CENT RELATIVE OVERCOUNT IN THE SECOND OF TWO CENSUSES TAKEN TEN YEARS APART

Age x	*Birth rate*	*Age x*	*Birth rate*
5	.0429	40	.0432
10	.0425	45	.0434
15	.0425	50	.0436
20	.0426	55	.0438
25	.0428	60	.0440
30	.0429	65	.0442
35	.0431	Test population	.0445

Now suppose that the estimate is made by stable population methods. It has been assumed that the improvement in coverage has not affected the reported age distribution, but the intercensal rate of increase is overestimated by about 2 per thousand. Table 12 shows the estimates of b obtained from this erroneous r and C(5), C(10) etc. Note that the error in estimating b varies little from about ages five to forty and that its absolute

[3] Cf. O. Andrew Collver, *Birth Rates in Latin America, New Estimates of Historical Trends and Fluctuations* (Institute of International Studies, University of California, Berkeley, 1965).

TABLE 13. EXPECTATION OF LIFE AT BIRTH AND VITAL RATES IN THE TEST POPULATION, AND ESTIMATES OF THESE QUANTITIES BASED ON THE CENSUS SURVIVAL METHOD AND THE STABLE POPULATION METHOD. BIAS OF ESTIMATES REFLECTS AN ASSUMED RELATIVE OVERCOUNT IN THE SECOND OF TWO CENSUSES OF THE TEST POPULATION, TAKEN TEN YEARS APART

	$^0e_0$	*Birth rate*	*Death rate*	*Growth rate*
Test population	40.0	.0445	.0234	.0211
Census survival method	43.5	.0434	.0203	.0231
Stable population method	44.3	.0429	.0198	.0231

magnitude is larger than when obtained with ogives up to higher ages. Once again, however, due to age-mis-statements at such ages, this observation is of no consequence in selecting a stable estimate of the birth rate from a reported age distribution.

Table 13 summarizes the values of various parameters based on the census survival method and on the stable population method i.e., obtained by accepting the median model life table and the median stable population among the first nine in tables 11 and 12, respectively. In the former case the m_x values of the selected life table are combined with an estimated mean age distribution to derive the (crude) death rate; the estimate of the birth rate is obtained by adding to this death rate the observed intercensal growth rate. In the case of the stable population method, the selection of the median stable population naturally implies the acceptance of all other stable parameters of that population. Note that with the census survival method a more complete enumeration in the second census causes an appearance of higher rates of survival and hence causes the estimated death rate to be too low; of course, the calculated intercensal rate of increase is too high, and the two errors are partially compensating when the birth rate is calculated. In the case of the stable population method, an upward-biased r in combination with $C(x)$ causes b to be underestimated—in the case of the median stable population this error is approximately equal to the error in r. Since its sign is opposite, however, the estimate of d is about twice as far removed from the true value as the estimate of b.

Variations in completeness of coverage tend to compromise the adjustments for declining mortality. For example, if the middle of three decennial censuses is especially incomplete, mortality in the earlier decade is overstated and in the later decade understated, creating the impression of declining mortality when death rates were actually constant, or exaggerating the extent of a real decline.

2. *Age-misreporting in censuses or surveys*

A large proportion of the ages recorded in most censuses and demographic surveys in less developed countries are inaccurate. In tabulation of the population by single years of age, peaks at ages ending in zero and five, and to a less extent at two and eight, are usual, with deficits at the other digits. Numbers reported at forty which are several times bigger than those at forty-one are not uncommon, for example. In section B.3 of chapter I comparisons with stable populations were used to demonstrate the existence of typical distortion in the ogives of age distribution, and in distributions by five-year age intervals, caused by characteristic forms of age-misreporting. The most conspicuous systematic distortions are found in populations in which apparently the ages of many persons are estimated by the interviewer rather than the respondent, and rules were suggested for selecting ages at which ogives are likely to be most reliable. Of course such rules can do no more than minimize errors that are unavoidably substantial, and the question remains concerning the range of unavoidable variation introduced into estimates of birth and death rates by age-mis-statements.

(a) *Age-mis-statement and mortality estimation by census survival*

The effects of age-mis-statement on the estimation of mortality by finding the model life table that duplicates observed census survival values cannot be simply summarized. There are two principal ways in which age-mis-statement affects the level of mortality that gives a projected population over $x+10$ agreeing with the enumerated population. First is the effect of age-mis-statement on the reported numbers over $x+10$ in the later census relative to the effect on those reported as over x in the earlier census. Both numbers may be inflated or both deflated without causing an error in the estimated level of mortality, provided relative inflation or deflation is the same. But if the increase in the proportion over thirty in the later census by age overstatement is less than the increase in the proportion over twenty in the earlier census, the estimated survival rates will be too low, and estimated mortality too high. Secondly, the estimated level of mortality is affected by age errors that exaggerate or understate the proportions at ages of high mortality. The typical exaggeration of the age of persons past fifty or sixty means that too many old persons are reported in the earlier census. This upward shift of age reduces the expected number of survivors in a projection by any given life table, and therefore requires overstated survival rates (too low mortality) to produce expected survivors equal to the actual. Experiments show that in the high-fertility populations for which estimates are usually needed, overstatement of age by the aged tangibly affects only the estimates of mortality based on projections of the population forty-five and over. Because of the complexity of the effects of age-mis-statement on census survival ratios, no preferred ages are suggested for estimating the level of mortality. Selection on the basis of projection of the population at ages above forty-five cannot be trusted, and the median of the first nine estimates is a neutral choice.

(b) *Age–mis-statement and stable population analysis*

The use of model stable populations to estimate various parameters involves the selection of a stable age distribution matching the recorded age distribution in some way. In this *Manual* the recommended procedure is to select a stable population with the same proportion under some age – the selection of *what* age depending on the pattern of apparent distortion of the recorded age distribution. With the extreme distortions characteristic of certain Asian and African censuses, it is recommended that the stable ogive matching the given female population at age thirty-five be used. For age distributions subject to less distortion such as found in the Philippines and Latin America use of the male distribution was recommended, specifically the selection of the median ogive among those matching the census at 5, 10, ..., 45.

Whatever procedure is followed, it can only avoid extreme errors, and cannot insure that the proportion recorded as under the age prescribed by the rule is exact. The value of $C(x)$ in the model population selected is thus generally somewhat above or below the true value of $C(x)$, because age–mis-statement has caused a net transfer of persons across age x. How much does an error in $C(x)$ affect the estimated birth rates, death rates, and other parameters ?

TABLE 14. VALUES OF $\Delta b/\Delta C(x)$ FOR THE TEST POPULATION WHEN THE BIRTH RATE IS ESTIMATED BY THE STABLE POPULATION METHOD GIVEN $C(x)$ AND r, AND $C(x)$ AND l_2

Age x	$\Delta b/\Delta C(x)$ *given* C(x) *and* r (*a*)	$\Delta b/\Delta C(x)$ *given* C(x) *and* l_2 (*b*)	*Col. a/b*
5	.568	.304	1.868
10	.388	.189	2.053
15	.319	.155	2.058
20	.293	.146	2.007
25	.283	.145	1.952
30	.289	.153	1.889
35	.305	.171	1.784
40	.336	.192	1.750
45	.379	.225	1.684

Table 14 shows the ratio of errors in the estimation of b to errors in the recorded value of $C(x)$ (x = 5 to 45) when $C(x)$ is combined with r on the one hand and l_2 on the other. These calculations apply to the same test population ("West" model female, $^{o}e_0$= 40, GRR_{29} = 3.00) employed before. The meaning of the entries in this table is illustrated by this example: Suppose $C(35)$ were recorded as .7627 rather than the correct figure of .7527, or that there were an error of .01 in $C(35)$. If r were known, the error in estimating b would be (.01) (.305) or .00305, and if l_2 were known, the error in estimating b would be .00171. (Since the "true" birth rate of the test population is 44.48 per thousand, the erroneous estimates would be 47.53 and 46.19, respectively.) Note that knowledge of l_2 yields more "robust" estimates of the birth rate than does knowledge of r—estimates less sensitive to errors in the reported age distribution as well as possible differences in age pattern of mortality. Estimates of d are in error to the same extent as is b when r is known. Because knowledge of l_2 implies (within the context of a family of model life tables) that the level of mortality is known, errors in $C(x)$ would generally produce trivial errors in the estimate of d, and errors in the estimate of r essentially equal to those in estimating b.

In the earlier discussion of typical pattern of age–mis-reporting, the hypothesis was advanced that when age reporting is not subject to the gross distortions seen in the censuses of Africa, India, Pakistan and Indonesia, populations have an approximate knowledge of age, accounting for the relatively orderly form of the ogive, and even of the five-year age distribution, even though age heaping is extensive. This hypothesis suggests that gross overstatement or understatement of age is not the norm and that age–mis-statement is usually the result of small errors, with a preference for round numbers. In Latin America and other areas where there appears to be knowledge of the approximate age, it is possible to argue that ogives to ages divisible by five have a systematic downward bias, because the cumulative age distribution always stops just short of the highly preferred ages ending in zero or five. To make the point in a different way: cumulative proportions under age 11, 16, 21, 26, 31 etc. would indicate higher fertility than ogives to age 10, 15 etc., because the first set would always just include, rather than just exclude, an age containing a greatly exaggerated number. It appears probable that the ogives barely including the ages divisible by five are too large because they contain persons really 11 and 12, 16 and 17, 21 and 22 etc., reported as 10, 15, 20 etc. On the other hand, ogives ending just short of ages divisible by five are possibly too small, because some persons 8 or 9 are reported as 10, 13 or 14 as 15 etc. To test the extent of the possible bias caused by this effect of age heaping, a special analysis was made of the census of Mexico in 1960, when age heaping was extensive. The numbers recorded in three-year intervals (9-11, 14-16, 19-21, etc.,) around each age divisible by five were reapportioned so as to have the sequence expected in an appropriate stable population, and the ogive recalculated. The effect was to increase the proportion below each age divisible by five (except below age fifteen) because some of the persons reported at these preferred ages were reassigned to the younger quinquennium. Despite the considerable age heaping, the differences between the birth rates estimated on the basis of the adjusted $C(x)$ values and those calculated from the unadjusted data were at all ages well within one per thousand population.

(c) *Age–mis-statement and the estimation of fertility and mortality from special questions on past experience*

In chapter II there is a description of methods of approximate calculation of fertility and mortality from the data obtained in asking women about the number of children they have ever borne, and the number of these surviving. It was noted, because of a tendency towards omission in the answers given by older women, that reported parity could be accepted as about correct only for women under thirty. However, it is a plausible hypothesis that the age pattern of fertility is correctly indicated by responses to questions about births in the preceding year. In the method of fertility estimation devised by William Brass, a comparison at ages 20-24 or 25-29 of cumulated fertility indicated by births reported for the preceding year with average numbers of children ever born to these ages provides an appropriate adjustment to reported births for the last year that can give a good estimate of the fertility schedule. This method, valid in principle when age is accurately reported, also works with tolerable accuracy in the presence of large and frequent errors in age, provided the errors are not systematic. But in the Asian-African populations, where age is often estimated by another person, it appears that about half the women in some of the child-bearing age groups are reported in the next higher group, which means that the cumulated experience of those reported as 15-19 does not in fact correspond to the history of those reported as twenty. It is not possible to analyse the resultant biases here,[4] but merely to warn that the method can be used only with the possibility of a wide margin of error in populations where massive systematic age-misreporting is apparent. It is much more promising, should the appropriate questions be asked in a census or survey, in populations subject to milder age distortions, such as in the Philippines or in Latin America.

Age-misreporting also affects the estimation of child mortality from proportions dead among children ever born to young women. Here again, accurate estimates can be expected only in populations not subject to gross transfers of women among ages 15-19, 20-24, 25-29 and 30-34. But the sensitivity of the estimated levels of child mortality to such transfers is not great, and one can be confident, for example, if responses about live and dead children are approximately correct, that the adjusted proportion dead reported by women 20-24 is greater than ${}_1q_0$ and less than ${}_3q_0$ even if not equal to ${}_2q_0$.

C. Suggestions for best estimation

A number of rules of thumb have been offered in this *Manual*, representing, however, not any definitive best procedure in each form of estimation, but a preliminary distillation of the authors' experience with each method. In the first application of an unfamiliar method, it may be best simply to follow these rules; but in situations when many alternative calculations are feasible (surely the usual situation in making estimates for a single national population), all of the possibilities here discussed should be examined, and a final estimate made only after a critial examination of these alternatives. The analyst should be sensitive to patterns of age-misreporting, to evidence that the population is not stable, and to the possibility of systematic omission of events or persons with certain characteristics.

A basic feature of demographic analysis underlying many of the procedures suggested here, but not always emphasized, is that there are logically necessary or biologically inevitable interrelations among population parameters, and these relations should be fully exploited in examining the consistency of estimates. For example, in populations in which circumstances permit a separate valid estimate of the average number of male and female births in a given decade, it is important to compare the estimated numbers to see if they lie within tolerable limits of the expected sex ratio at birth. Almost all populations of non-African origin have 105-107 male births for every 100 female births, subject, of course to sampling fluctuations. Populations of African origin have a sex ratio at birth of perhaps 102-104. Thus if a male birth rate (male births/male population) and a female birth rate have been calculated from the age distribution of each sex, it is essential to see whether the estimates are consonant or inconsistent. If the former, there is reason for added confidence, if the latter, the data or the methods are inconsistent (a judgement subject to the crucial reservation that sampling variance must be allowed for).

One of the principles to be borne in mind in making such checks is the importance of noting what aspects of the two estimates being confronted are independent. Thus for example a comparison of the level of mortality indicated by census survival during a decade, and by $C(x)$ in the second census plus the intercensal rate of increase is a comparison of two inferences from essentially the same basic data. Deficiencies in the censuses or special features in the age pattern of mortality would affect the two estimates similarly, and perfect agreement would not be as reassuring as agreement between estimates with wholly independent bases. An example of a more nearly independent pair of estimates is the male birth rate calculated indirectly from the female age distribution plus the sex ratio of the population and an assumed normal sex ratio at birth, and directly from the male age distribution.

The comparison that is most acceptable as a confirmation of valid estimation is between figures derived from data of wholly different kinds. For example, in a situation where fertility is constant, data on children ever born by age of woman, if themselves internally consistent, take on added persuasiveness if verified by comparison with the cumulation of current fertility rates based on reported births last year. But suppose now that death rates over five are calculated from census survival, and under five from proportions surviving among children ever born, that these rates are combined to form an estimated population death rate, to which the intercensal rate of increase is added to provide a figure for the birth rate. If now the birth rate thus obtained agrees closely with that derived by the methods outlined earlier in this paragraph, the confirmation is a strong one. Of course it is always possible that the agreement is fortuitous. An additional principle in judging the importance of agreement is its consistency. As a final element in this hypothetical example, suppose that the comparison has been made in each of a moderate number—ten or more—of subdivisions of the same data system, such as regions or provinces enumerated in the same national censuses. If

[4] See the discussion in Brass *et al.*, op. cit.

the various forms of independent estimation yield fertilities that agree in the geographical differences they show, the reality of the differences is more or less conclusively established.

The last suggestion for making the best of inaccurate and incomplete data is always to seek the form of estimation least sensitive to unknowable uncertainties—to age patterns of mortality when age specific death rates are not recorded, to systematic age–misreporting etc. This principle leads to preferring l_2 to r as an adjunct to $C(x)$ in estimating the birth rate.