UNITED NATIONS

ECONOMIC AND SOCIAL COUNCIL

Distr.
GENERAL *
E/CONF.13/293
Meeting No. 3
30 June 1954

ORIGINAL: ENGLISH
(Paper in English)

WORLD POPULATION CONFERENCE
Rome, 31 August - 10 September 1954

Comments on possible national and international standards for the determination and indication of degree of accuracy of demographic statistics

W. Parker Mauldin (U.S.A)

Summary

Relatively little is known about the accuracy of most demographic data, but recently the United Nations has taken several important steps designed to give the user of demographic statistics some indication of the quality of the data it publishes. It has also formulated several tests for evaluating the accuracy of age-sex distributions. Several different methods of evaluating the accuracy of demographic statistics are discussed, including checking of census and other demographic data against : (1) independent aggregates, (2) independent records, (3) special surveys designed to evaluate the quality of data, (4) expected survivors of cohorts reported at an earlier time, and (5) other items included in a survey. Emphasis is placed on developing methods for evaluating accuracy so that rigorous tests may be applied for items such as total population counts, age-sex distributions, and registered births and deaths. Various tests that have been suggested for evaluating total population counts, age-sex distributions, and registered births are enumerated and briefly discussed and some recommendations regarding the use and publication of results of such tests are given.

* General distribution of this document is limited to the introductory summary. Participants who have been invited to take part in the meeting referred to above will receive also the full text of the paper. Other participants in the Conference will receive the full text upon request.

Pour la traduction française voir au verso.

54-18414

18p.

Commentaires relatifs aux normes susceptibles d'être adoptées, tant sur le plan national qu'international, pour la détermination du degré d'exactitude des statistiques démographiques.

W. Parker Mauldin (Etats-Unis)

Résumé. On sait relativement peu de choses sur le degré d'exactitude de la plupart des données démographiques, mais l'Organisation des Nations Unies a récemment adopté plusieurs dispositions importantes destinées à renseigner, dans une certaine mesure, les usagers des statistiques démographiques sur la valeur approximative des données qu'elle publie. Elle a également défini plusieurs méthodes permettant d'évaluer la précision des répartitions selon l'âge et le sexe. On soumet à un examen critique plusieurs procédés permettant la précision des statistiques démographiques, notamment le contrôle des recensements et des autres données démographiques avec : 1) des quantités globales indépendantes; 2) des enregistrements autres que ceux qui ont été utilisés dans l'élaboration des chiffres que l'on vérifie; 3) des enquêtes spéciales destinées à estimer la valeur des renseignements fournis; 4) les survivants présumés des promotions enregistrées à une époque antérieure; 5) d'autres rubriques comprises dans les enquêtes. On souligne la nécessité de mettre au point des procédés qui permettent d'évaluer le degré d'exactitude, afin que des méthodes rigoureuses de contrôles puissent être appliquées, par exemple aux chiffres totaux de population, aux répartitions selon l'âge et le sexe et aux naissances et décès enregistrés. On énumère diverses méthodes de contrôle qui ont été proposées en vue d'évaluer les chiffres totaux de population, les répartitions selon l'âge et le sexe et les naissances enregistrées, et on soumet ces méthodes à un bref examen critique, pour conclure par quelques recommandations relatives à l'emploi et à la publication des résultats ainsi obtenus.

⧈ Seule la présente analyse d'introduction fait l'objet d'une distribution générale. Les participants qui ont été invités à assister à la séance mentionnée ci-dessus recevront en outre le texte intégral du document. Les autres participants au Congrès recevront le texte intégral sur leur demande.

E/CONF.13/293
Meeting No. 3
ORIGINAL: ENGLISH

COMMENTS ON POSSIBLE NATIONAL AND INTERNATIONAL STANDARDS FOR THE DETERMINATION AND INDICATION OF DEGREE OF ACCURACY OF DEMOGRAPHIC STATISTICS

by
W. Parker Mauldin
U. S. Bureau of the Census

Much has been accomplished in recent years in the establishment of standards for items to be included in censuses and vital statistics programs, definition of the items, and tabulations to be made from the collected statistics. This is true in spite of the fact that simple demographic data are unavailable for large segments of the population of the world. Relatively little is known about the accuracy of most demographic data, although the United Nations has taken several important steps designed to give the user of demographic statistics some indication of the quality[1] of the data and also has formulated several tests for evaluating the accuracy of age-sex distributions.[2] Measures such as these are very helpful, but much remains to be done before standards can be adopted and applied to a majority of the population of the world. Objective tests and criteria need to be adopted wherever possible, although in the absence of such tests subjective appraisals of the quality of the data are desirable.

Demographic statistics undoubtedly have been greatly improved over the years, and presumably are very accurate in some countries at the present time. Such an evaluation must, unfortunately, be made on the basis of

54-18414 A

subjective impressions and the application of nonrigorous "tests" to the data collected. Relatively little systematic research has been undertaken designed to determine cost and accuracy characteristics of different survey designs. In general, decisions relating to optimum design -- where cost and degree of accuracy desired are balanced against each other -- are made on the basis of subjective information. This is necessarily so inasmuch as we do not as yet know how much improvement in accuracy may be purchased for a given amount of money. Nor do we always know which survey design gives more accurate results than would other alternatives that might be chosen. It is obvious that the cost of insuring a high degree of accuracy for each subgroup in the population may be disproportionately high. In practice, it is necessary to choose a survey design that presumably gives a high degree of accuracy for most of the population, and for most items being covered, but does not necessarily give a high degree of accuracy for each subgroup in the population. A careful research program designed to measure the accuracy of demographic data should enable us to quantify our impressions in this regard, and should point out the areas of weakness in various methods of collecting information.

One purpose in evaluating the accuracy of demographic data is to apprise the user of data as to how much reliance may be placed on the information. It is not at all clear, however, what degree of precision is needed in the compilation of demographic data. Some rules of thumb may be developed on the basis of how much the user considers it worth to improve a given statistic or set of statistics beyond a specified point. It is of questionable value, however, to make such judgments in the absence of information as to what policy decisions are made on the basis of various demographic data, and how these

decisions might be affected if the statistics were different, or were known to be subject to sizeable error. But in spite of difficulties such as these, some indication of the validity of demographic data is needed, and is useful.

Another purpose served by evaluating accuracy of data is to furnish guides to improve the methods of collection and compilation of these data. Surveys designed to measure the accuracy of demographic statistics generally will also provide information on the sources of error, which, of course, are needed in order to improve the accuracy of the survey design. Some types of checks do not aid in understanding the causes of the errors, nor is this process necessarily an efficient way of learning how to go about improving the data; nonetheless, there are many situations in which it is quite useful in this respect.

Methods of Evaluating Accuracy.--Theoretically, there are several different methods that may be used in evaluating demographic statistics. In practice the methods are quite limited because the nature of the aggregates with which we are concerned -- such as censuses of total population, age and sex distributions, and registration of births and deaths -- is such that they themselves constitute the "best" available data extant. Some of the methods consist of checking demographic data against:

1. independent aggregates
2. individual (administrative) records
3. cohorts reported at an earlier time (survivorship of cohorts)
4. special surveys designed to evaluate quality of data
5. other data in the survey (check for internal consistency)

There is not time to develop each of these methods in any detail but a few illustrations may serve to point out some of their strengths and weaknesses.

1. Checking against independent aggregates.--This technique involves a comparison of a census or other type of statistic with an independent figure purporting to represent the same thing. Normally this technique can not be used for checking the accuracy of the total population count, for example, because only rarely is there an independent figure representing the total population. This method can be used for checking subgroups of the population, as was done in comparing the United States Census figures of males in selected age groups in 1940 with Selective Service Registration statistics.[3]

2. Checking against individual (administrative) records.--This technique consists of checking a given set of data against the records available from some administrative public program. Normally the information pertaining to a subsample of the subjects of a survey are compared on a case-by-case basis with records containing the same information. This technique has been used successfully in checking the accuracy of birth registrations in the United States in 1940 and 1950[4] and in checking age returns in a census of New Zealand.[5] This technique has a major drawback, aside from the very real problem of cost, and that is that the proportion of nonmatches often is so high that the results of matching are inconclusive. Eckler and Pritzker[6] list the following limitations of this technique:

"(a) Incompleteness of the file of records. It is known, for example, that despite marked gains over the years, even today some births are not reported. In some other files, however, similar improvement has not taken place.

"(b) The file of records may cover only a certain class of the population. Thus, the Veterans Administration files do not contain the names of non-veterans. Hence, when a person cannot be located

in the Veterans Administration files, there are three possible explanations: he may be a non-veteran; the file of veterans may be incomplete; or there may be an error in the matching process.

"(c) The matching process itself is often quite inconclusive. In addition to errors of filing, spelling and other clerical errors there are risks associated with making a decision as to what constitutes a "match" in a particular case. There are hundreds of John Smiths in New York City for example, and a considerable number born in the same year.

"Thus, while the technique is useful in producing direct measurements on the accuracy of survey results for those cases where matching can be completed, the significant gaps left by unmatched cases seriously limit the inferences that can be drawn."

3. Checking against cohorts reported at an earlier time.--One of the most frequently used methods of evaluating accuracy of demographic statistics is that of comparing cohorts as enumerated in successive censuses. Some additional knowledge normally is presumed, such as a knowledge of the age-sex specific mortality rates, information about the relative magnitude of migration by age and sex, etc. This method is useful for estimating the accuracy of total population counts, assuming that births, deaths, and migration records are reasonably accurate (an assumption not always justified), counts by age, sex, race, etc.[7] Guatemala has indicated that its 1940 census figure for the total population was overstated, or padded, by perhaps as much as 40 percent.[8] The cohort survival technique could be useful in discovering falsifications of this kind.

4. Special surveys designed to evaluate quality of data.--This method is much more flexible and more definitive than the other techniques suggested inasmuch as it is designed for the specific purpose of determining the quality of a given set of data. This procedure has been introduced as a regular part of the census program in some countries, for example, in the United States[9] and in India.[10]

5. Checking against other data in the survey.--Checks for internal consistency are commonly employed, although their value is quite limited. Such checks may serve to alert the user of the data that a problem exists, but they are not adequate for quantifying the extent of any inaccuracies.

Items to be Checked.--Partly because of limits of time and space and partly because of the complexity of checking many items the discussion in this paper will be limited to the following items:

1. Total population
2. Age-sex distribution
3. Births

1. Total population.--Perhaps the most frequently used population statistic is the total population within a specified area. There are two principal methods whereby the accuracy of population counts may be evaluated.

a. The count at a previous census plus births and immigration, minus deaths and emigration should equal the current census count. This is a useful method, but it does not provide an adequate basis for establishing a "standard." If the data don't balance, one only knows that one of the censuses, or births, or deaths, or immigration, or emigration is inaccurate.

If the data do balance it is not a necessary indication of accuracy of the data, although these conditions support the assumption that recorded increments to the population are of approximately the right magnitude, and this may be taken as an indication of moderately high reliability. It is suggested that international agencies consider adopting codes that reflect the consistency of data of this kind. This, as I see it, should supplement the codes currently employed by the United Nations in connection with their evaluation of estimates of total population.

b. A check enumeration, using a probability sample, more intensive training, more detailed procedures, etc., is perhaps the best method now available to us for estimating the accuracy of population counts. One advantage of this method is that the sampling error is known, and can be computed and stated in conventional terms. One disadvantage of the method is that the biases inherent in both the census and the check enumeration are difficult to isolate and to estimate. I shall not duplicate the description of this type of check given by Hurwitz and Pritzker, but should like to recommend that this general procedure be utilized whenever practicable.

At the present time it does not seem to me desirable to set a hypothetical "standard" for accuracy of population counts, but rather it is desirable that those countries carrying out such a survey should publish, preferably in the census volume, a description of the procedures employed and an estimate of the over- and underenumeration, with an indication of the sampling error involved.

Table 2 of the 1953 Demographic Yearbook presents data on the enumerated populations of the various countries of the world, and detailed footnotes alert the reader to boundary changes, classes of persons not enumerated, and in some instances, presents estimates of under- or overenumeration. In some cases the footnote states simply "Believed to be underenumerated," as in the case of Portuguese Guinea for 1940.[11] The figure for Kenya in 1948 excludes groups amounting to more than 4 percent[12] of the cited figure. The 1950 enumerated population for Honduras is given in the Demographic Yearbook, but a footnote indicates that a 10 percent adjustment[13] should be made for underenumeration. Similarly, the 1950 figure cited for Bolivia is 12 percent above the enumerated figure,[14] and the 1940 figure cited for Peru was thought to be too low by about 13 percent.[15] The basis for these figures is not, and probably cannot be, given in a publication of this sort, but some further indication of where one can find supporting evidence for the quantitative figures is desirable.

In terms of current population estimates, the procedure now being followed by the United Nations in the publication of the Demographic Yearbook is warmly endorsed. Possibly the codes could be revised slightly to indicate whether or not the statement "substantially accurate statistics of births and deaths" is based on objective studies, or on the basis of knowledge acquired by on-the-spot researchers.

2. Age-sex distribution.--Age and sex distributions are among the most useful demographic data that can be made available for a given population. These distributions are basic to computing age-specific fertility and mortality rates, labor force participation rates, etc. Also, far more attention has been devoted to ways of evaluating such distributions than to any other statistic or group of demographic statistics. Durand has proposed four tests for errors in a given

age-sex distribution, two for children and two for adults, and five tests for inconsistencies in age-sex distributions from two censuses by cohort survival comparisons -- two for children and three for adults.[16] Three tests of age and sex ratios are proposed in the United Nations Population Bulletin.[17] Many other tests have been proposed by a variety of authors, including Myers' "blended" test for heaping,[18] Whipple's index of concentration,[19] test of errors by minus differences, or the "coefficient of error,"[20] comparison of ungraduated age returns with adjusted distributions,[21] the ratio of survivorship of two adjacent cohorts as determined from census data divided by the same ratio as determined from mortality data.[22] There are many variations of the above "tests," and undoubtedly other kinds of tests have been proposed and used by other demographers, but this brief listing is indicative of the fact that a variety of relatively simple tests now are available for evaluating age-sex distributions.

The various international agencies concerned with demographic data could select from among tests such as those listed two or three for general use. Some additional work needs to be done in determining what "standards" should be adopted or recommended. It is easier to arrive at such standards for a test which evaluates a given age-sex distribution than for one which utilizes the cohort-survival technique. Durand, however, has suggested some standards for comparison of data from two censuses, and these could serve as an excellent starting point for this purpose. A wider choice of codes could be adopted so as to indicate somewhat more precisely the consistency or lack of consistency in the distributions. Also, countries should be requested to publish unadjusted age-sex distributions rather than only a "corrected" distribution.

It should be noted that some of the age-sex tests not only reflect the accuracy of the age and sex distributions but also are useful in evaluating the size of the total population. For example, in the United States the relatively low sex ratios observed in the young adult ages without a correspondingly high ratio in the adjacent age groups is indicative of underenumeration among young adult males. This low sex ratio is particularly noticeable among the nonwhite population, which suggests that underenumeration among nonwhite young adult males is higher than among the white groups. Similarly, an examination of the age distribution within the 0 to 4 group may indicate underenumeration of infants. In some countries, an unusually high sex ratio, particularly among children, is indicative of underenumeration of females. In view of the fact that the sex ratio at birth is virtually constant, it is possible to make adjustments for underenumeration of a given sex, assuming that mortality rates are known.

3. Birth statistics.--Birth and death data are among the most important demographic statistics collected, and, accordingly, it is useful to have some measure of their accuracy. Due to lack of time and space only birth statistics will be discussed, although some of the observations apply to both types of statistics.

Among the most common errors in connection with birth statistics are (1) failure to differentiate among stillbirths and live births, (2) failure to register births, and (3) delayed registrations. Each of these types of errors has been given considerable attention through the years and from country to country, although precise methods of measuring the extent of error have been utilized only infrequently. Problems pertaining to accuracy of information regarding parents of the children, e.g., age and usual place of residence of mother at time of birth, parity, duration of marriage, etc., have been virtually ignored -- and are beyond the scope of this paper.

Many devices have been used in an effort to estimate the proportion of births that are registered. In the United States, for example, shortly after the birth registration area was established, the names of newborn infants collected from a variety of sources such as newspaper listings, infant death records, school censuses, etc., were matched against birth certificates as a means of estimating the proportion of births registered within a given State. A more systematic procedure for obtaining a representative list of births was adopted in the early twenties. Postcards were distributed to every household in the State by the postal delivery service. Each family was requested to fill out the card if a child had been born during a specified period of time. This procedure was considered to be more accurate than any other that could be devised at that time, although returns usually amounted to only between 30 and 50 percent of the number of birth records on file.[23]

A much better procedure was adopted in 1940 and 1950 when birth registrations and census enumeration of infants were matched in an effort to estimate the proportion of births registered in each State of the United States.[24] These tests indicated that 92.5 percent of births were registered in the United States in 1940, and about 98 percent in 1950. This type of test has one major defect. Failure to register the birth of an individual and failure to enumerate an infant in the census probably are not random events, but, rather, it is likely that these are positively associated occurrences. If this assumption is true, a bias of unknown magnitude results, although from other kinds of information it seems probable that the bias is relatively small in terms of the total number of births. The effect of this factor taken alone is to overstate the proportion of births that are registered.

There are at least two other well-known methods of arriving at estimates of the accuracy of registration of births. The first of these was developed by Whelpton and may be described as follows. It is assumed that birth registration is complete in certain areas. The ratio of children under one, calculated on the basis of birth and death registrations, to those enumerated in the census is assumed to apply to all other areas. Thus, the enumerated number of children under one is adjusted on the basis of this ratio, and this calculated number of births is divided into the number of births registered during the previous year to obtain an estimate of the percentage of births that were registered.[25] It should be pointed out that the ratio calculated in the manner just mentioned may not apply to all areas within a given country. Differential underenumeration by section of the country, or extensive in- and out-migration of infants between date of birth and date of census enumeration,[26] etc., could invalidate this procedure.

Another procedure consists of comparing the birth registration statistics with births calculated from census and mortality statistics. For example, the number of persons 10 to 14 years of age enumerated in a census in 1950, divided by the survival rate for a ten-year period, gives an estimate of the number of persons 0 to 4 in 1940. This estimate of persons 0 to 4 years of age can be converted into an estimated number of live births. The estimated number of live births divided into the registered births gives an estimate of the proportion of births that were registered.[27]

At the present time the United Nations uses the symbol "C" for data stated to represent a complete, or virtually complete, coverage of births (and/or deaths) occurring each year. The symbol "U" is used for data stated to represent an incomplete coverage or to be subject to considerable irregularity of registration.[28]

These codes are assigned on the basis of information obtained from the national statistical offices and from relevant official publications. These symbols are useful guides to the quality of the data, but, of course, do not provide measures of accuracy, nor can they be regarded as "standards." This dichotomization of the data into the categories "complete" and "incomplete" is a start in the right direction. It would be still more useful to expand this code, after appropriate consultation with experts from various countries, so that the "complete" codes could be quantified in a rough way. For example, "A" might represent countries which state that 95 percent or more of births are registered within 6 months of occurrence, with this statement being based on a comparison of birth registrations with a census enumeration. "B" could represent 95 percent or more of births registered within 6 months of occurrence, with this statement being based on a comparison of birth registrations with a list of infants prepared from various sources other than census enumeration. "C" might represent the same percentage completeness as calculated using the Whelpton method. Other codes could be used for 90 to 94 and for 85 to 89 percentage completeness of birth registrations.

Tentatively, I should like to suggest that codes of this sort be utilized if 85 percent or more of births are said to be registered within a given period. A separate code should be used for those countries which report their births as virtually complete but with such statements being based on nonquantitative knowledge. Similarly, the current "U" code should be retained. I realize that many problems exist in putting such a scheme into operation. The time reference creates a problem because countries which have determined that birth registrations are virtually complete are not likely to undertake detailed studies to "check" their feeling that the registration system has not deteriorated appreciably.

Also, there may be considerable variation in the estimated percentage completeness of registration from time to time or from area to area if nonrigorous "tests" are applied. Objections of this sort are equally applicable to the other tentative suggestions made in this paper.

The application of uniform tests to the collection and compilation of demographic data obviously cannot be accomplished everywhere at the present time. Nonetheless, the steps recently taken by the United Nations and the studies undertaken by countries such as India, Guatemala, Costa Rica, the United States of America, and Yugoslavia, to name just a few, not only will help the users to evaluate the strengths and weaknesses of the data, but also will serve as a stimulus to improve the completeness of coverage and the accuracy of the data collected. It is preferable to express estimates of completeness of coverage, errors in age, etc., in terms of percentage deviation from the estimated total for a given subclass, and to give the sampling error of such estimates. It does not seem likely that such a practice will soon be feasible for many countries or for many items. The utilization of descriptive codes such as those employed by the United Nations at the present time is an excellent device and it is hoped that this practice not only will be continued but that its use will be expanded. In particular, the methods employed in evaluating the data should be indicated.

[1] United Nations, Demographic Yearbook, 1953, New York, pp. 8-16.

[2] United Nations, "Accuracy Tests for Census Age Distributions Tabulated in Five-Year and Ten-Year Groups," Population Bulletin, No. 2, October 1952, pp. 59-79.

[3] Daniel P. Price, "A Check on Under-Enumeration in the 1940 Census," American Sociological Review, v. 12, no. 1, February 1947, pp. 44-49.

[4] Sam Shapiro, "Development of Birth Registration and Birth Statistics in the United States," Population Studies, v. IV, no. 1, June 1950, pp. 94-95; Sam Shapiro and Joseph Schachter, "Methodology and Summary Results of the 1950 Birth Registration Test in the United States," Estadistica, v. X, no. 37, December 1952, pp. 688-699.

[5] New Zealand, Results of a Census of the Dominion of New Zealand, 17th April 1921, General Report, Wellington, 1925, pp. 90-96.

[6] A. Ross Eckler and Leon Pritzker, "Measuring the Accuracy of Enumerative Surveys," Bulletin of the International Statistical Institute, v. XXXIII, Part IV, p. 12.

[7] Ibid., pp. 9-10, consider the "inflow-outflow" technique for estimating as a different technique from the survivorship of cohorts procedures. These two techniques are considered together in this paper.

[8] United Nations, Demographic Yearbook, 1953, p. 86; Estadistica, no. 1, September 1950, p. 6.

[9] U. S. Bureau of the Census, U. S. Census of Population: 1950, v. I, 1952, pp. xii-xiv; William N. Hurwitz and Leon Pritzker, "The 1950 Census of Population of the United States: Accuracy of Statistics and Reliability of Methods," paper prepared for this conference.

[10] Census of India, Paper No. 1, Sample Verification of the 1951 Census Count, 1953.

[11] Demographic Yearbook, 1953, op. cit., p. 84.

[12] Ibid., p. 84.

[13] Ibid., p. 86.

[14] Ibid., p. 90.

[15] Ibid., p. 90.

[16] John D. Durand, "Adequacy of Existing Census Statistics for Basic Demographic Research," Population Studies, v. IV, no. 2, September 1950.

[17] United Nations, "Accuracy Tests for Census Age Distributions Tabulated in Five-Year and Ten-Year Groups." Population Bulletin, No. 2, October 1952, pp. 59-79.

[18] Robert J. Myers, "Errors and Bias in the Reporting of Ages in Census Data," Transactions of the Actuarial Society of America, v. XLI, part 2, no. 104, October-November 1940.

[19] U. S. Bureau of the Census, Thirteenth Census of the United States, v. I, Population, 1913, p. 291.

[20] Hugh H. Wolfenden, Population Statistics and Their Compilation, Actuarial Studies No. 3, The Actuarial Society of America, 1925, p. 40.

[21] Ibid., p. 41.

[22] Ansley J. Coale, "The Population of the United States in 1950 Classified by Age, Sex, and Color -- A Revision of Census Figures," to appear in the Journal of the American Statistical Association.

[23] Sam Shapiro, Ibid., pp. 94-95.

[24] Ibid., and Sam Shapiro and Joseph Schachter, Ibid., pp. 688-699.

[25] P. K. Whelpton, "The Completeness of Birth Registration in the United States," Journal of the American Statistical Association, v. XXIX, no. 186, June 1934.

[26] Abram J. Jaffe, Handbook of Statistical Methods for Demographers, U. S. Government Printing Office, Washington, D. C., 1951, p. 141.

[27] Ibid., pp. 140-144.

[28] United Nations, Population and Vital Statistics Reports, Statistical Papers, Series A. See, for example, v. VI, no. 2.