MISSING AND MISPLACED PERSONS: THE CASE OF CENSUS EVALUATION IN DEVELOPING COUNTRIES

James F. Spitler and Eduardo E. Arriaga, U.S. Bureau of the Census

It is generally recognized that data collected by censuses and surveys are subject to error, and without knowledge of the magnitude and direction of this error, results are of questionable usefulness. This lack of knowledge is particularly acute in many developing countries where census and survey data provide the major sources of information on the demographic processes.

The purpose of the present paper is to discuss the utility individual record checks and aggregate comparisons offer in population census coverage evaluation. Selected developing countries of Asia and Latin America are taken as examples. Since each of the evaluative methods are themselves subject to error, particular attention is given to cases where a combination of individual record checks and aggregate comparisons have been utilized to derive estimates of census coverage error.

EVALUATION METHODS

Individual record checks. This method refers to the checking of individual census returns against records which are independently obtained in an effort to measure the degree of consistency between the two sets of information. For the purpose of this paper, two types of record checks are identified: matching studies and postenumeration surveys (PES).

Matching studies involve the matching of individual census records with those obtained from an independent source such as previous population census, censuses of housing and agriculture, birth and death registers, church records, tax rolls, school enrollment records, records on old age benefits, etc. The results of the matching process give estimates of gross differences (erroneous omissions and inclusions), as well as net differences. Furthermore, this method may be used to obtain a listing of the population which is more complete than either the census or the independent source.

In general, postenumeration surveys (PES) are special household surveys (taken shortly after the census enumeration), conducted for the purpose of evaluating censuses through an individual matching process. As such, they generally provide estimates of both net and gross coverage error.

Aggregate comparisons. This method of evaluation pertains to the critical analysis of the internal consistency of the census results; to the manner in which these results relate to known demographic social, political, and/or natural occurrences; and to the relationship between independent estimates (derived from direct or indirect estimation techniques) for the components of demographic change and the size, distribution, and characteristics of the population. As such, the method may utilize any or all of the commonly known techniques of demographic analysis; e.g., balancing equation, lexis diagram, cohort analysis, age and sex ratio analysis, forward and/or reverse survival. Comparisons at the aggregate level only give indications of the net differences between the estimates.

Obviously, the method (or methods) employed to evaluate the census for a single country is dependent upon the availability and detail of the necessary data. Many countries of the world do not undertake individual record checks as a method of census evaluation. For those countries which do conduct such checks, there is often a lack of published cross-tabulations of sufficient detail to provide a meaningful evaluation beyond an aggregate comparison of the published results. The failure of many countries to publish sufficient detail to adequately assess the quality of results obtained from individual record checks in terms of variance and bias also hampers the researcher engaged in the evaluation of published census data. Census evaluation is further complicated in many instances by the absence of independent demographic or vital event data to apply alternative direct and/or indirect demographic estimation techniques in an effort to conduct an evaluation based on aggregate comparisons. This is particularly true where migration is an important factor in the process of demographic change.

In the sections which follow, discussion centers on the results obtained from an application of various evaluative methods of available data in selected countries of Asia and Latin America. Particular emphasis is given to the limitations of data and methods, and to the considerations involved in ascertaining the "final" estimate of coverage error.

CENSUS EVALUATION IN SELECTED COUNTRIES

The census evaluations discussed below draw upon the experiences of the demographic situation in a number of developing countries. The examples were selected after considering the availability of several different types of evaluative methods, as well as the numerous approaches to census evaluation given specific method combinations. As can be seen in table 1, methods in the selected countries cover a wide spectrum; from Mexico with no individual record checks, to Peninsular Malaysia with a PES tabulated by age, sex and race.

Mexico. Due to the lack of individual record checks, the evaluation of the 1960 and 1970 censuses had to rely upon aggregate comparisons. These comparisons were further hampered by the non-availability of reliable international migration data. Thus, after considering the probable impact migration had had upon the census age-sex structure, it was felt that the most prudent course of action would be to concentrate upon the evaluation of the population under age 10 and to accept the reported figures for the total population ages 10 years and over for each sex.

An extensive demographic analysis of data from the vital registration system led to the conclusion that birth and death registration was relatively complete and could, with "minor" adjustments, be used to construct Lexis diagrams and obtain adjusted populations under 10 years of age for each sex. Attention next focused on the elimination of probable age misreporting in the accepted population over 10 years of age. This was accomplished by accepting the enumerated population for both sexes combined in each 10-year age group, and subsequently splitting them into 5-year age groups with a mathematical formula and applying a smoothed set of sex ratios. The resulting distributions were adjusted back to the enumerated totals for each sex.

Jamaica. Although the Jamaica evaluation also relies upon aggregate comparisons, it differs from

Mexico in two important aspects: 1) a reenumeration of selected areas was undertaken in 1970 and the results were incorporated into the published figures; and 2) information on external migration was judged to be "more reliable" (based on a comparison of reported emigration and immigration data for the major receiving countries — Canada, United Kingdom, and the United States), than in the case of Mexico.

The approach used to evaluate the 1970 Jamaican census was essentially a forward survival of an officially adjusted 1960 census, using adjusted registered births and deaths, and reported emigration data for the intercensal period. Because of the reliance on reported emigration data, the net coverage error in 1970 implied by this approach (see table 2) may overstate the "actual" coverage error to the extent that emigration has been misspecified.

Thailand. The postenumeration survey conducted shortly after the 1970 census resulted in a low estimate of net under-coverage (see table 2). It was also possible to obtain various aggregate comparisons. This was hindered, however, by the lack of adequate vital registration data. Thus, it was necessary to indirectly estimate the levels and trends for each of the demographic components using various estimation techniques before an evaluation of the census could be undertaken.

The processes of evaluating the census consisted of obtaining an adjusted 1960 census based on cohort analysis, age and sex ratio analysis, and reverse survival utilizing estimated levels and trends for the components of change during the 1950-60 period. Once the evaluated 1960 census age-sex distribution had been obtained, it was survived to 1970 using estimated levels and trends in fertility and mortality for the 1960-70 intercensal period.

Honduras. Two types of individual record checks were used to evaluate the 1961 census; a reenumeration of selected areas, and a matching of census schedules with the birth register for the month prior to the census. Results from these two procedures indicated net underenumeration of 8.9 percent (all ages) for the selected areas and 3.6 percent (under age one) for the reenumeration and matching procedures, respectively; the combined estimate of net underenumeration was 5.3 percent (Honduras, 1962, table 1). No individual record check was conducted after the 1974 census.

As an initial step in the aggregate evaluation of the 1961 and 1974 censuses, cohorts for each sex were analyzed to ascertain the degree of consistency of cohorts between the two censuses. The results suggest that there was either a considerably larger underenumeration in the 1974 census (relative to the 1961 census), or there had been a sizable amount of emigration from Honduras during the intercensal period. After investigating the available evidence for emigration, it appeared that the discrepancies in the cohorts at the beginning and end of the intercensal period were more likely due to the greater extent of underenumeration in the 1974 census. Therefore, the aggregate evaluation concentrated on obtaining an adjusted 1961 age-sex distribution which would be survived to 1974 using intercensal estimates of fertility, mortality, and migration obtained by various demographic estimation techniques applied to data from numerous sources.

The first step in the 1961 census evaluation was to smooth, for each sex, the reported population in 10-year age groups and split the resulting estimates into 5-year age groups to lessen the effects of age misreporting. At this point a sex ratio analysis was undertaken, and the smoothed and split age-sex distribution was adjusted to an expected pattern of sex ratios. These adjustments implied a total net underenumeration which was less than that obtained by the individual record checks. Therefore, the adjusted age-sex distribution obtained by the age and sex ratio analysis was proportionally inflated to the total population figure implied by the total net underenumeration estimated from the individual record check (5.3 percent). A final step was to obtain an evaluation of the population under 5 years of age for each sex. First, the total births for 1956-61 were obtained by using an estimated set of age-specific fertility rates, the adjusted 1961 female population and a female population for 1956 (reverse survived from 1961). Second, these births were subsequently survived to 1961, resulting in adjusted population under 5 years of age which implied a net underenumeration of 9.45 percent for both sexes.

Pakistan. In the case of Pakistan, a PES was undertaken after the 1961 and 1972 censuses. Results from the 1961 PES were reported only for the total urban and rural population, and indicated no significant net coverage error (see table 2). The 1972 PES was not only directly used for establishing the undercount in particular ages, but also formed the basis for the aggregate evaluation.

An extensive age and sex ratio analysis indicated that the results from the 1972 PES could be accepted for the overall estimate of net underenumeration, for each 10-year age groups (over age 19) for each sex, and for the total net error estimated for all ages under 20 years. The problem, therefore, amounted to obtaining an estimate of the age-sex distribution under age 20 which would be consistent with the overall net coverage error found by the PES for the age group 0 to 19, and with past trends of fertility and mortality. This was accomplished by: 1) inflating the broad age-sex distribution reported in the census by the net coverage error found by the PES; 2)splitting the adjusted 10-year age groups into 5-year age groups; 4) reverse surviving the age-sex distribution over age 19 to 1952; and 5) projecting this age-sex distribution to 1972 based on estimated levels and trends in fertility and mortality during the 20-year period. The age-sex distribution under age 19 resulting from this projection was accepted. (For more detailed examples of the procedures, see U.S. Bureau of the Census, 1980).

Malaysia (Peninsular). This example is similar to Pakistan in that the 1970 PES provided information for net coverage error by age and sex (see table 3). The major difference lies in the approach to the estimation of coverage error for under age 10. Rather than having to rely on a reverse survival and projection process, independent estimates were derived through the use of adjusted vital registration statistics and a Lexis diagram technique.

As is shown in table 3, the PES estimates for the sex ratios suggest the possibility of rather severe age misreporting (although less than in the enumerated census) for ages 30 years and over. It also suggests that the PES estimates of underenumeration for the population under age 10 are probably too low, while those for ages over 70 are too high, based on the experiences found in most developing

The independent analysis of the population less than 10 years of age gave credence to the observation that the PES estimates were relatively low. While no indirect independent estimate could be made for the population ages 70 and over, the reported census figures appeared more reasonable in light of the pattern of sex ratios obtained by splitting the PES estimates for the age groups 10 to 69 years.

Taking the aforementioned results into consideration, the combined estimates shown in table 3 are comprised of: 1)the results from the Lexis diagram to obtain the adjusted population under age 10; 2) the PES results by 10-year age groups for ages 10 to 69 smoothed to account for age misreporting; and 3) the acceptance of the enumerated census population 70 years of age and over. (For an example of combining results from individual record checks and aggregate comparisons for the Republic of Korea, see Marks and Finch, 1977).

CONCLUSIONS

This paper discussed methods frequently used in developing countries for evaluating enumerated census populations by age and sex. The methods were classified into individual record checks based on postenumeration surveys and matching procedures and aggregate comparisons based on analytical demographic procedures. Several examples of particular approaches to evaluation were presented (3 Asian and 3 Latin American countries).

The point was made the reliability of results from individual record checks should be evaluated in terms of the survey design, statistical error and confidence intervals of the estimates, and the matching process. The data and information necessary for this are, however, often not available from published sources. Similarly, it was noted that the results from aggregate comparisons should be evaluated in relation to the reliability of the demographic estimates accepted and the validity of assumptions made in the process of evaluating the age-sex distribution of the population.

The tentative conclusion to be reached is that no single evaluative approach or procedure can be universally recommended. Furthermore, no recommendation can be made as to which method (individual record checks or aggregate comparison) may provide "more reliable" results under differing circumstances. Currently, indications tend to support the contention that individual record checks provide "more acceptable" results when evaluating the population over age 10. Aggregate comparison, on the other hand, tend to provide a "more acceptable" evaluation of the population under 10 years of age. In all cases, combining both approaches appear to produce a "more acceptable" evaluation. That is, complementing an individual record check with aggregate comparison--through a demographic analysis--produce results which are more consistent with existing knowledge about the demographic characteristics of the population and the components of demographic change--mortality, fertility and migration. Probably the only general rule that can be offered is that in all cases, a combination of evaluative techniques should be applied to the available data and the results of each taken into consideration.

It is hoped that continued research by international, governmental, and private organizations into all aspects of census evaluation will be encouraged. Specific attention in the case of individual record checks, should be drawn to the areas of developing and evaluating the usefulness of alternative survey designs, estimating and evaluating the effects of correlational bias, and of investigating problems associated with the matching process. With regard to aggregate comparisons, research efforts should focus upon investigations into the reliability and validity of estimates for the components of demographic change derived through the application of various indirect estimation techniques. Such investigations should include the validity of underlying assumptions and the consequences deviations from the assumptions have upon the resulting estimates. Only through the continued research efforts of all concerned, can a more complete understanding of each evaluative method be reached and more conclusive recommendations be made regarding the evaluation of census data for developing countries.

MAJOR SOURCES

Arnold, Fred and Mathana Phananiramai. 1975. *Revised Estimates of the 1970 Population of Thailand.* Research Paper No. 1, National Statistical Office. Bangkok.

Honduras, Dirección General de Estadística y Censos 1962. *Post-Enumeration Study, Census of Population and Housing of Honduras: 1961.* Tegucigalpa, D.C.

_____. 1977. *Anuario Estadístico: 1975.* Tegucigalpa, D.C.

Malaysia. Department of Statistics. 1973. *An Interim Report on the Post Enumeration Survey.* Kuala Lumpur.

_____. 1975. *1970 General Report — Population Census of Malaysia.* Vol. 2. Kuala Lumpur.

Marks, Eli S. and Glenda Finch. 1977. "Developments in Techniques of Census Evaluation". *Proceedings of the 41st Session of the International Statistical Institute, Bulletin of the International Statistical Institute.* Vol. 47, Book 4, pp. 318-321.

Pakistan. Office of the Census Commissioner. no date a. *Census of Pakistan.* Vol. 1. *Population 1961 - Pakistan.* Karachi.

_____. no date b. *Census of Pakistan.* Vol. 3. *Population 1961 - West Pakistan.* Karachi.

_____. 1974. *Census Evaluation Survey, Population Census 1972.* Karachi.

United Nations. 1970. *Demographic Yearbook: 1969.* New York.

_____. 1971. *Demographic Yearbook: 1970.* New York.

U.S. Bureau of the Census. 1977a. *Country Demographic Profiles - Honduras* by Glenda Finch. Washington, D.C.

_____. 1977b. *Country Demographic Profiles-Jamaica* by Marilyn K. Sharif. Washington, D.C.

_____. 1978. *Country Demographic Profiles - Thailand* by James F. Spitler. Washington, D.C.

_____. 1979a. *Country Demographic Profiles - Malaysia* by Glenda Finch. Washington, D.C.

_____. 1979b. *Country Demographic Profile - Mexico* by Patricia M. Rowe. Washington, D.C.

_____. 1980. *Country Demographic Profiles - Pakistan* by Frank B. Hobbs. Washington, D.C.

ACKNOWLEDGEMENTS

The authors with to express their appreciation to Frank B. Hobbs, Glenda Finch, and Patricia M. Rowe of the International Demographic Data Center for providing unpublished data for specific country census evaluations. Also, appreciation is extended to Samuel Baum, James Dinwiddie, and Eli Marks for their valuable comments and suggestions made on earlier drafts of this paper. The paper was prepared under a Resources Support Services Agreement with the Development Support Bureau, Agency for International Development.

Table 1. Availability of Individual Record Checks, by Type and Available Major Cross-Tabulations: Selected Countries and Years

Region, country, and year	Type	Available cross-tabulations
ASIA		
Malaysia (Peninsular)		
1970	Postenumeration survey	Total, by age, sex, and race.
Pakistan		
1961	Postenumeration survey	Totals, by urban and rural residence.
1972	Postenumeration survey	Totals, by sex, age and urban and rural residence.
Thailand		
1960	None	(X)
1970	Postenumeration survey	Total only.
LATIN AMERICA		
Honduras		
1961	Reenumeration, matching with birth register	Total, by type of method
1974	None	(X)
Jamaica		
1960	None	(X)
1970	Reenumeration of select areas	Published population figures incorporated figures for reenumerated areas.
Mexico		
1960	None	(X)
1970	None	(X)

X Not applicable

Table 2. Enumerated Census Population and Estimated Net Coverage Error, by Method for Both Sexes and Specified Ages: Selected Countries and Years
(Population in thousands; net coverage error in percent)

Region, country and year	Enumerated census population (both sexes)		Net coverage error (both sexes)					
			Individual record check		Aggregate comparison		Accepted	
	All ages	Under age 10	All ages	Under age 10	All ages	Under age 10	All ages	Under age 10
ASIA								
Malaysia (Peninsular)								
1970	8,810	2,728	-4.1	-3.6	(X)[1]	-6.4	-4.-	-6.4
Pakistan								
1961	42,978[2]	14,088[2]	+0.4	(NA)	-16.0	-9.2	-16.0	-9.2
1972	65,309[3]	20,548[3]	-6.3	-7.4	(X)[4]	-4.8	- 6.3	-4.8
Thailand								
1960	26,258	8,246[5]	(X)	(X)	- 4.0	-9.5	- 4.0	-9.5
1970	34,397	10,958[5]	-1.7	(NA)	- 6.6	-5.1	- 6.6	-5.1
LATIN AMERICA								
Honduras								
1961	1,885	666[5]	-5.3	(NA)	- 3.0	-4.3	- 6.0	-5.9
1974	2,657	910	(X)	(X)	-12.5	-13.9	-12.5	-13.9
Jamaica								
1960	1,610	489	(X)	(X)	- 0.9[6]	- 2.9[6]	- 0.9[6]	- 2.9[6]
1970	1,832[7]	598[7]	(X)	(X)	- 5.5[8]	- 4.2[8]	- 5.5[8]	- 4.2[8]
Mexico								
1960	34,923	11,130[5]	(X)	(X)	- 3.3[1]	- 9.6	- 3.3[1]	- 9.6
1970	48,225	15,891	(X)	(X)	- 2.4[1]	- 7.0	- 2.4[1]	- 7.0

NA Data not available.
X Not applicable.

[1]The population 10 years of age and over was not adjusted for underenumeration.

[2]Includes estimates and reported figures for tribal areas and non-Pakistanis (Pakistan, no date b, Chapter 4, table 10; Chapter 5, tables 13 and 14; and Chapter 9, tables 1 and 4 (sections I and II).

[3]Includes reported figures for the Federally Administered Tribal Areas, the Kohistan Area of Hazara District, and the Tribal Areas adjoining Hazara District.

[4]The population 20 years of age and over was not adjusted for underenumeration.

[5]Includes persons of unknown age proportionally distributed.

[6]The population 5 years of age and over was not adjusted for underenumeration.

[7]Includes persons of unknown sex and age proportionally distributed, but excludes the institutionalized population.

[8]Excludes institutionalized population for which no adjustments for coverage error were made.

Note: All figures are subject to sampling and/or response variance. A plus (+) sign denotes net overenumeration; a negative (-) sign denotes net underenumeration.

Sources:

Malaysia (Peninsular) - Population as reported in Department of Statistics, 1975, tables 4.4 and 5.1; record checks as reported in Department of Statistics, 1973, table 6; and aggregate comparison and accepted coverage error from U.S. Bureau of the Census, 1979, unpublished data.

Pakistan - Population from U.S. Bureau of the Census, 1980, unpublished data; record checks for 1961 as reported in Pakistan, no date a, p.I-15, and for 1972 based on a weighted average of urban and rural estimates as reported in Pakistan, 1974, tables II, V, and VIII; aggregate comparison from U.S. Bureau of the Census, 1980, unpublished data; and accepted coverage error as reported in U.S. Bureau of the Census, 1980, p. 2.

Thailand - Population from U.S. Bureau of the Census, 1978, unpublished data; record check as reported in Arnold and Phananiramai, 1975, table 13; and aggregate comparison and accepted coverage error as reported in U.S. Bureau of the Census, 1978, p. 1.

Honduras - Population for 1961 and 1974 as reported in United Nations, 1971, table 6 and Honduras, 1977, table 6, respectively; record checks as reported in Honduras, 1962, table 1; aggregate comparison from U.S. Bureau of the Census, 1977, unpublished data; and accepted coverage error as reported in U.S. Bureau of the Census, 1977a, p.1.

Jamaica - Population for 1960 as reported in United Nations, 1970, table 6, and for 1970 from U.S. Bureau of the Census, 1977, unpublished data; aggregate comparison and accepted coverage error as reported in U.S. Bureau of the Census, 1977b, p. 1.

Mexico - Population as reported in U.S. Bureau of the Census, 1979b, tables A-1 and A-2; aggregate comparison and accepted coverage error as reported in U.S. Bureau of the Census, 1979b, p.25.

Table 3. Enumerated Census Population (Both Sexes), Estimated Sex Ratio and Net Coverage Error, by Age and Method: Peninsular Malaysia, 1970

Age	Enumerated census population (in thousands)	Estimated sex ratio (Male per 100 females)				Estimated net coverage error for both sexes (Percent)			
		Census	PES	Aggregate comparison	Combined	Census	PES	Aggregate comparison	Combined
All ages	8,810	101	102	102	102	(X)	- 4.1	(X)	- 4.7
0 to 4 years	1,370	104	104	104	104	(X)	- 3.9	-8.0	- 8.0
5 to 9 years	1,358	104	104	103	103	(X)	- 3.2	-4.8	- 4.8
10 to 14 years	1,198	103	103	102	102	(X)	- 3.6		- 2.9
15 to 19 years	977	98	100	101	101	(X)	- 4.9		- 5.7
20 to 24 years	745	97	100	100	100	(X)	- 5.2		- 5.2
25 to 29 years	550	99	100	100	100	(X)	- 4.3		- 4.3
30 to 34 years	534	99	102	99	99	(X)	- 3.9		+ 0.2
35 to 39 years	420	95	97	99	99	(X)	- 2.8		- 7.6
40 to 44 years	374	100	101	99	99	(X)	- 2.4	-4.0[1]	- 1.2
45 to 49 years	310	97	98	100	100	(X)	- 2.9		- 4.3
50 to 54 years	276	103	102	104	104	(X)	- 3.4		- 2.5
55 to 59 years	223	110	110	107	107	(X)	- 4.4		- 5.5
60 to 64 years	195	109	109	111	111	(X)	- 4.9		+ 3.1
65 to 69 years	121	123	121	116	116	(X)	- 3.7		-14.5
70 to 74 years	83	106	102	106	106	(X)	- 5.5	0.0	0.0
75 years and over	76	89	92	89	89	(X)	-20.5		0.0

X Not applicable

[1]Based on an acceptance of the estimated total underenumeration for ages 10 to 69 years obtained by the PES.

Note: All figures are subject to sampling and/or response variance. A plus (+) sign denote net overenumeration; a minus (-) sign denotes net underenumeration.

Source: Population as reported in Department of Statistics, 1975, tables 4.4 and 5.1; PES coverage error as reported in Department of Statistics, 1973, table 6; aggregate comparison and combined coverage error from U.S. Bureau of the Census, 1979, unpublished data.