

THE 1974 POST-ENUMERATION SURVEY OF LIBERIA - A NEW APPROACH
Eli S. Marks and John C. Rumford, U. S. Bureau of the Census

Fifteen days after the beginning of the 1974 census of population and housing in Liberia, a post-enumeration survey was conducted. Though the short interval of time between the census and evaluation survey is unusual, what made the Liberian effort different was the simple and relatively inexpensive survey and implementation system used.

In the past, post-enumeration surveys have proven difficult, time consuming, and expensive to carry out. Because of this, many countries, particularly the developing nations, have avoided post-enumeration surveys in census evaluation planning. This is unfortunate because no census, regardless of how carefully planned and executed, is perfect. While perfection cannot be obtained, completeness of enumeration can and should be estimated.

The primary purpose of a post-enumeration survey (PES), is to provide estimates of census coverage. All censuses suffer from two types of coverage errors. The first, and most frequent, is undercoverage, and the second is overcounting. Undercoverage errors are caused by such things as missing persons in enumerated living quarters, missing living quarters entirely or, in some cases, missing whole localities. Overcounting is almost always of much smaller magnitude and results from enumerators overlapping at area boundaries, reporting of persons by more than one household and enumeration of persons who should not have been enumerated, such as (names entered for) nonexistent individuals or persons who died before or were born after the census date.

The traditional PES used to estimate these errors involves reenumerating a sample of census enumeration areas (EA's) or parts of EA's several months after the census, using a specially selected, well-trained group of enumerators. These enumerators are usually provided with preenumeration intelligence derived from the census, and each is instructed to reconstruct the population in the evaluation area as of the census date. At the close of the reenumeration, a case-by-case matching with the census questionnaires is carried out, and a full field follow-up "reconciliation" is made of all non-matching persons.

The principle underlying this methodology is that the PES will be much better than the census and, therefore, estimates derived from this survey are a "standard" to which the census can be compared and eventually adjusted.

However, experience in the United States and elsewhere has suggested that PES estimates are not necessarily better than census results (U. S. Bureau of the Census, 1960, Marks and Waksberg, 1966, and Marks, 1973). Moreover, in the Liberian context, evidence existed that the techniques associated with traditional PES systems were inappropriate. This evidence was primarily the experience gained from four years of conducting the Liberian Fertility Survey, a

national, multi-round, household survey, that used case-by-case matching and other techniques common to traditional post-enumeration survey designs (Rumford, 1970). During this survey, it was repeatedly demonstrated that a single enumeration system, using experienced, well-trained enumerators and supervisors, failed to enumerate many persons. Moreover, it was found that, when enumerators were grouped and tested by age, experience, and education, the older, more experienced, and better educated enumerators missed about the same number of persons as did their younger and less experienced colleagues with average education (Rumford, 1972). Confronted with this evidence, it was decided to use a new PES approach in Liberia based on dual system estimation.

Dual System Estimation

There are four main methodological elements in this new approach as implemented in Liberia, which depart from the traditional PES method. The first is emphasis on independence between the census and PES; the second is the use of one-way matching to reduce "geographic out-of-scope" error; the third is providing for a very brief time interval between census and PES in order to minimize the problems of tracing migrants; and the fourth is the elimination of field verification of unmatched (Census or PES) enumerations. The third and fourth features were introduced in Liberia for reasons of cost and operational feasibility. The first two features are attempts to reduce the effects of two of the three major biases of dual system estimation.

As used in the current paper, "dual system estimation" involves (1) collecting data from a sample of the target population with two independent data collection systems (in census evaluation, the Census and the PES); (2) matching the reports of the two systems to determine which of the sample individuals were reported by both systems; (3) using the proportion matched of all cases reported in one system as an estimate of the completeness of reporting (the "coverage rate") in the other system. Thus, the proportion matched of all PES cases is used as the estimate of the completeness of the Census coverage.

Independence

It will be noted that the purpose of the PES is to provide an estimate of the completeness of the Census coverage. The PES estimate of Census coverage will not be biased by the erroneous omission of some sample cases or by the erroneous inclusions of some nonsample cases (e.g., due to the PES enumerator getting outside the boundaries of the sample segment), provided the PES coverage errors are independent of the Census coverage errors--i.e., provided the probability of a PES case being matched to the Census (found to have been enumerated in the Census) is the same for

the cases erroneously omitted from or erroneously included in the PES as it is for the cases correctly included. Note that independence implies nothing about causality. Obviously, whether a person is or is not reported in the PES cannot causally affect his probability of being reported in a census taken prior to the PES. However, the classes of persons with low probabilities of being reported in the PES could also (apart from any direct causal connection) have lower (or higher) probabilities of being reported in the Census than the classes of persons with high probabilities of being reported in the PES. There is, in fact, evidence from recent post-enumeration surveys in Paraguay and in Korea that persons who change their places of residence between the Census and the PES tend to have lower probabilities of enumeration in both the Census (taken before their change of residence) and the PES than persons who do not move between the Census and the PES.

Correlation between PES and Census errors may be "direct" or "indirect". Indirect correlation results from the fact that the probability of certain individuals or classes of individuals being enumerated is high or is low for both Census and PES.^{1/} Direct correlation involves a causal relationship between Census and PES errors--i.e., the fact of a person being enumerated or not being enumerated in the Census actually changes the probability of his being enumerated in the PES.

Since indirect correlation is an inherent feature of the population and the reporting methods, its control involves careful selection of the basic PES procedures. The newer techniques of taking a PES, used in Korea and Paraguay, were developed primarily for the purpose of eliminating a major source of indirect correlation.

The preservation of direct statistical independence is largely a matter of avoiding collusion between the two enumeration systems in the field. One of the steps that can be taken to do this is to postpone the selection of the sample areas until after the Census enumeration is complete. In Liberia, the PES sample was selected immediately prior to national census day. However, the sample enumeration areas were not identified to the field officers until the census enumeration was completed 5 to 10 days later. When the field officers were notified, they were instructed to immediately impound the completed census questionnaire workbooks for the EA's scheduled to be re-enumerated. These workbooks remained in custody throughout the PES enumeration.

In another effort to prevent direct correlation, the PES enumerators were recruited from County Inspectors and District Supervisors. These workers were selected because they were familiar with the general census enumeration procedures, but took no active part in the census enumeration at the EA level. The enumerators were briefed on the PES questionnaires but they were not given any additional training nor were they provided with any preenumeration intelligence. Moreover, they used listing sheets and EA maps that were duplicates of those provided to the census enumerators. A time limit of 72 hours was imposed for the enumeration (the same time target

prescribed for the original census enumeration) and the PES, like the Census, was conducted on a de jure basis.

After completing the enumeration, the PES questionnaire workbooks were transferred to a census regional officer. However, no review was made at regional headquarters. Instead, the original census questionnaire workbooks were transmitted to national headquarters for review and matching by a completely separate and specialized group. In spite of these elaborate precautions, and as a tribute to the ingenuity of Man, two of the thirty-two sample EA's were compromised and had to be eliminated.

It should be stressed that although it is not sufficient for unbiased PES estimates, technical and administrative independence is essential. Without it, the Liberian PES system cannot be used.

Other Biases

In addition to correlation bias, dual system estimates are subject to matching bias and out-of-scope (or erroneous inclusion) bias. Matching bias is the result of "erroneous matches" and "erroneous nonmatches". Erroneous matches will increase the number of PES cases considered to be enumerated in the Census and will, consequently, result in an overestimate of the completeness of Census coverage and erroneous nonmatches lead to an underestimate. The overall matching bias depends on the "net matching bias", which is the difference between the number of erroneous matches and the number of erroneous nonmatches.

"Out-of-scope" error is the result of improper inclusion of cases in the PES or the Census. Erroneous inclusions in the Census are duplicate enumerations and enumerations of persons who should not have been enumerated--e.g., persons who died before or were born after the Census date, diplomatic personnel or other persons excluded from the Census by definition and enumeration of fictitious persons, either deliberately (i.e., enumerator "curbstoning", the completion of Census entries without interviewing every household) or accidentally (e.g., entry of a dog or other household pet under the mistaken impression that it was a child). Erroneous inclusions in the PES include the PES enumeration of persons who should not have been enumerated in the Census (as described above) and also the enumeration of nonsample persons. Enumeration of nonsample persons results from boundary difficulties (i.e., the PES enumerator enumerating households actually located outside the sample segment) and from improper handling of "migrants" (persons who move between the Census and the PES). On the latter, "migrants" can be sampled either on the basis of where they were on the Census date or of where they are at the time of the PES interview. Thus, an out-of-scope error occurs in a PES which samples on the basis of residence on the Census date, if the PES enumerator includes someone who moved into the sample segment after the Census date. For PES samples based on location at the time of the PES interview, an out-of-

scope error occurs if the PES enumerator includes someone who moved out of the sample segment before the PES date.

Handling of Migrants

Persons who move into or out of a sample segment between the census date and the time of the PES enumeration represent a particularly difficult problem. Prior to 1970, all PES sampling was based on the person's location on the census date. However, this tended to produce a correlation bias because (1) migrants tend to have a larger census omission rate than nonmigrants and (2) persons who move away from an area prior to the PES are likely to be missed by the PES. Persons who move away tend to be omitted from the PES because (a) where a whole household moves, neighbors may be able to furnish only very incomplete information about the individual household members; and (b) if an individual moves out of a household, the remaining members of the household may be very vague about the date that he left.^{2/} The reasons for poor census enumeration are not immediately obvious. Part of the difficulty may be with the fact that census enumeration is dragged out over a long time period and, therefore, the canvass for many of the migrants occurs after the migration. However, the higher omission rates for migrants occur also in de facto censuses in areas where 80% or more of the census enumeration is actually completed on the census date. It may be hypothesized that the causal mechanism may be the fact that many of the "migrants" (particularly in a de facto Census) have only tenuous connections with the household where they were staying or living on the census date and are, therefore, not mentioned when the enumerator asks for "all the people staying (or living) in this household".

There are two ways of reducing the correlation bias due to migrants. One of these has been introduced only recently into PES work--in the PES of the 1970 Census of Korea and the PES of the 1972 Census of Paraguay. This method involves sampling migrants on the basis of where they are at the time of the PES interview. In Paraguay, two samples of approximately equal size were used in the 1972 PES, one asking about people in the sample segment on the census date and the other asking about people in the sample segment at the time of the PES interview. The second sample gave nearly 4 times as many migrants (people who moved into sample segments) as the number of migrants (people who moved out of sample segments) given by the first sample, although the number of nonmigrants was about the same (only 7% difference) for the two samples.

While basing the sampling of migrants on their residence at the time of the PES will improve the reporting of migrants and thus reduce the correlation bias due to migrants, it considerably increases the difficulties (and the errors) in matching the migrants to the Census. That is, in order to search for a migrant in the Census files, it is necessary to have his address at the time of the Census, with the kind of precision and detail that permits an accurate deter-

mination of the enumeration area (EA) in which he should have been enumerated. While people will usually know their former addresses, they frequently cannot furnish the kind of address information needed to determine the EA.

The other method of reducing correlation bias due to the migrants is to reduce the number of migrants. This can be done by reducing the time lag between the Census and the PES. This is the method adopted for Liberia, where the time lag between the Liberian census date and the PES enumeration was set at fifteen days. This period was long enough to allow the census enumerators to canvass and clear their areas and yet, was short enough to limit opportunities for migrations, births, and deaths to occur. The PES asked whether anyone listed on the questionnaire was born after the national census day and eliminated such persons from the tabulations. No such accommodation was considered necessary for deaths, since the rarity of this event in a ten to fifteen day period would make errors from this source negligible. However, it should be pointed out that, if the period between the Census and the PES is longer, a specific question on deaths may be needed. It should also be pointed out that the short time period increases the danger of overlapping the two enumeration systems in the field if strict precautions are not taken.

Use of One-Way Matching

A second problem is control of the geographic out-of-scope errors, which occur at the boundaries of sample evaluation areas. These errors can be particularly troublesome in heavily populated urban centers where no easily identified natural or man-made boundaries exist. In these areas, it is relatively easy for a census or PES enumerator to erroneously extend his canvassing area beyond the limits of the designated sample area. Because the Liberian system is based on maintaining strict independence between the Census and the PES, it is not desirable to use information obtained during the actual Census enumeration to help locate the EA boundaries. To minimize the effect of boundary errors, the Liberian PES used "one-way matching", with a provision for searching for cases in adjoining EA's where the possibility of boundary errors existed.

One-way matching was used in Liberia primarily because of its relative simplicity and economy. The technique is most effectively employed in circumstances where one of the two data sources has records covering the entire population. With one-way matching, it is only necessary to determine the exact matching status of each report for one of the two record systems (the one with data for a sample only). As implemented in the Liberian Post Enumeration Survey, the PES records were compared to the census records and each person was categorized as "found" (or "matched") or "not found" in the census listings.

After the initial matching within the sample EA, the unmatched PES cases were searched for in adjacent EA's. For the PES cases that were matched in an adjacent EA, an attempt was made to deter-

mine whether the error was made by the Census or the PES. However, the determination of whether the error was made by the Census or the PES was not very successful. In most cases, an acceptable determination of the correct EA boundaries requires a further field visit and it was not felt that the expense of such a visit was warranted since the number of cases involved was small.

A major purpose of a field reconciliation is to determine whether any of the PES cases, matched or unmatched, should not have been included either because the person was not supposed to be enumerated in the Census or because the person was a nonsample case who should have been enumerated outside the sample EA boundaries. Actually, since the only purpose of the PES is to determine the estimate of the proportion of the population enumerated in the Census, the bias due to a few cases erroneously included in the PES sample is minor, provided the same treatment is given to unmatched as to matched cases. Of course, there will be a bias if the cases near an EA boundary do not have exactly the same Census omission rate as those in the interior of an EA. In general, cases near an EA boundary will have a somewhat higher census omission rate and, also, a somewhat higher census duplication rate. However, both the difference and the number of cases involved will be small (unless the census maps and procedures are totally inadequate) and the bias will be trivial. The important thing is to extend the search for PES cases near an EA boundary to the adjacent EA's, so that no PES case will be called unmatched because the census or PES enumerator (or both) made a mistake in locating the EA boundary. It would, of course, be desirable to search all boundary cases, matched or unmatched, in the adjacent EA's to check on duplicate census enumerations. However, such duplication is rare in practice and the bias is minor of not searching adjacent EA's for cases matched within the sample EA.

PES RESULTS

The PES estimates of census completeness by age and sex appear in Table 1. Overall, the 1974 Liberian Census appears to have achieved about 90% coverage. The figures for particular age-sex cells are subject to fairly high variances (and, possibly, to biases in the reporting of age). However, the overall pattern represents more than variance and shows interesting differences from the underenumeration pattern typical of more industrialized countries. Thus, there is little or no overall difference between males and females in completeness of coverage for the 1974 Liberian Census. On the other hand, the United States censuses show better coverage of women than men largely due to the poorer enumeration of men in the age range 15 to 45, which also seems to hold for Liberia. For Liberia, the higher coverage of females 14 to 45 is balanced by the lower coverage for females under 10 and 45+ which does not obtain for the U. S. In fact, U. S. coverage of females under 10 and also of females 65 and over tends to be better than the coverage of males in the corresponding age

groups. In the U. S. 1970 Census, the coverage of males dropped for ages 20-24 and remained low up to age 45 but no such decline appears for U. S. females ages 20 to 44. In Liberia the coverage of both sexes seems to drop off at ages 15 to 19 but then improves for ages 20 to 24 and continues to improve for males up to age 60 and for females up to age 45.

Some Defects of the Liberian PES

A key feature of the Liberian PES system is completing the PES field work within a short interval after the Census date. This is also a key weakness of the system.

Having a very short interval between the Census and the PES eliminates many of the very serious problems associated with substantial migration between the Census date and the PES date. It makes it possible to keep the PES costs down to relatively low levels and also to utilize in the PES a good part of the field and administrative structure set up on a temporary basis for the Census. On the other hand, it imposes certain requirements which may be serious handicaps. One of these is the difficulty of maintaining independence and administrative control between the Census and PES. As noted above, special instructions were issued that a PES enumerator not be told the location of his assignment until all of the Census enumeration for the evaluation area was complete and the Census schedules for the EA were in the hands of the regional supervisor. In spite of this precaution, there was a coincidence between Census and PES reports for two of the thirty-two sample EA's which could not possibly have occurred without someone altering either the Census returns to accord with the PES results or vice versa.

The fact that this contamination affected only 6% of the Liberian PES sample reflects the great efforts which were made to preserve independence while adhering to an extremely tight time schedule. With adequate planning and supervision, five days is sufficient time for completing a Census in 80% or more of the EA's of almost any country. Unfortunately, there are delays and slip-ups that will affect a minority of EA's in almost any country. An enumerator may have misunderstood his map (with or without the "help" of ambiguous boundaries) and failed to enumerate a whole section of his EA; or an entire EA or group of EA's may not have been assigned for enumeration; and these errors might not be detected until preliminary count figures are announced and complaints from local areas start pouring (or dribbling) in.

A short time interval between Census and PES imposes other restraints. The sample areas must be selected well before the Census, possibly before the work of setting up the Census EA's has been completed. (Note: There are always areas where some of the initial EA's are too big and must be split up and other areas where combinations of some small EA's are desirable). Once selected, the sample EA's must be kept confidential until the Census work in those EA's has been completed.

Also, the short time interval usually means that the PES schedules and procedures cannot be tested and revised under actual census conditions. This is likely to be particularly harmful with respect to the matching procedures. Most countries have some experience with other Census and PES operations. If not, it is possible to draw on experiences of other countries with similar conditions (e.g., in collecting and coding data on occupation and industry) or to dispense with or simplify the operation. Most developing countries (and not a few statistically advanced countries) have no directly applicable experience with modern matching methods. The matching operation is indispensable to a PES and simplification can pose very serious dangers of matching errors which raise questions about the validity of the entire PES. There were, in fact, serious defects in the design of the Liberian PES matching. In trying to keep the PES procedure simple and the questionnaire short, some information very useful for matching was omitted, notably relationships and alternative names. Both of these items are fairly easy to obtain and add only trivially to the total interview time. "Relationship" will be useful for matching in almost any culture (particularly those where most of the names are common ones and it is necessary to distinguish the Ali bin Muhammed who is Muhammed bin Muhammed's younger brother from the one who is his son). The need for alternative given names is culturally limited; and the fact that they may not be needed in the United States does not mean that they will not be needed in Liberia.

One problem of the Liberian PES system is particularly acute for de facto censuses. Providing for a short time interval between PES and Census is not satisfactory for dealing with the "floating population"--i.e., persons with no usual residence any place. In de jure censuses, such persons are to be enumerated on a de facto basis--i.e., where they happen to be staying on the census date. Even with a PES taken two weeks after the census date (as in Liberia), the probability is small of finding persons with no usual place of residence where they were on the census date.

Conclusion

The Liberian system is no panacea. When the final 1974 census results become available, all the available methods of demographic analysis will be required to refine and improve the coverage estimates derived from dual system estimation. Comparisons with previous census and survey results will be necessary to fully evaluate the complete census coverage picture.

Moreover, as indicated above, the system itself has several deficiencies. Nevertheless, with some minor modifications, such as the provision of additional matching information or a small extension in the time period between Census and PES (e.g., 30 days instead of 15), the Liberian PES system is applicable to and effective for countries which do a de jure Census with a very short Census enumeration period and also to countries with a de facto census where the matter

of temporary sojourners in a location is minor (usually countries which are primarily rural and nonindustrial).

Where a country does a de facto census and has a substantial number of temporary "sojourners" (or persons with no usual place of residence) or where a short time interval between PES and Census is not feasible, one should consider the use of a PES sample based on where the individual is at the time of the PES. This requires determining for migrants the location on the Census date and doing the matching search at that location. While this kind of matching can be difficult and expensive, methods for simplifying the procedures and improving the results are being explored and may provide an answer in the near future.

We can be certain that there will continue to be problems of evaluating Census coverage, no matter which methods are used, and whether they involve a PES or other dual system estimation or do not. However, the need for methods of more adequate census evaluation and correction is painfully evident for many developing countries where recent censuses have been so inconsistent with all the known demographic facts as to be completely unacceptable.

No one method of census evaluation and correction has a complete answer. However, the post-enumeration survey is a powerful technique which should not be ignored or avoided. It is hoped that the simplicity and economy of the dual estimation system used in Liberia will help remove the post-enumeration survey from the luxury item list in census evaluation and place this important tool within reach of all census planners. It is at least a starting point from which census planners can go on to more sophisticated and more powerful (but not necessarily, more expensive) tools.

Acknowledgements

The Liberian Post-Enumeration Survey was conducted by the staff of the Ministry of Planning and Economic Affairs, Republic of Liberia. The funds for survey implementation were provided by the Government of Liberia and the United Nations Fund for Population Activities. Technical support was given by the United Nations' Office of Technical Cooperation, the African Census Program of the United Nations Economic Commission for Africa and the United States Bureau of the Census.

The opinions expressed and the techniques demonstrated in this paper are those of the authors and are not necessarily endorsed by any of the cited organizations or individuals within these organizations.

Footnotes

1. There can, in theory, be indirect negative correlation, in which individuals with a high probability of being reported in the Census have a low probability of being picked up in a PES and vice versa, but no actual instances of this have been noted to date.
2. Precise dating of "events" such as births, deaths and migrations is a major problem of vital

statistics measurement. It was, for example, the major reason for the development of techniques for estimating fertility which use "children ever born" (regardless of date) rather than children born during the past year.

References

1. Ahmed N., and K. Krotki, 1963. Simultaneous Estimations of the Population Growth: The Pakistan Experiment. Pakistan Development Review, 3, 37-65.
2. Chandrasekhar, C., & W. E. Deming, 1949. On a Method of Estimating Birth & Death Rates & the Extent of Registration. JASA, 44, 101-115.
3. Malawi National Statistical Office, 1973. Malawi Population Change Survey: Feb. 1970 - Jan. 1972. Zomba, Malawi: Govt. Printer
4. Marks, E. S., 1973. The Role of Dual System Estimation in Census Evaluation. Unpublished paper presented at the first meeting of the International Assn. of Survey Statisticians, Vienna, Austria, Aug. 1973. Scheduled to appear as Chap. 10 in Development in the Dual System Estimation of Population Size & Growth, edited by K. Krotki, U. of Alberta Press, 1976.
5. Marks, E. S., W. Seltzer & K. Krotki, 1974. Population Growth Estimation: A Handbook of Vital Statistics Measurement. N. Y., The Population Council.
6. Marks, E. S., & J. Waksberg, 1966. Evaluation of Coverage in the 1960 Census of Population Through Case-by-Case Checking. Proceedings of the Social Statistics Section, 1966, 62-70, American Statistical Association, Washington, D. C.
7. Rumford, J. C., 1970. Use of the Chandrasekhar-Deming Technique in the Liberian Fertility Study. Public Health Reports, 85, 965-973.
8. Rumford, J. C., 1972. Factors Influencing the Casefinding of Migrations in the Liberian Fertility Survey, Demography, 9, 431-441.
9. Rumford, J. C., Y. Heperkan & N. Fincancioğlu, 1968. The principles & Preliminary Results of the Turkish Demographic Survey. Public Health Reports, 83, 573-582.
10. Seltzer, W. & A. Adlakha, 1969. On the Effect of Errors in the Application of the Chandrasekhar-Deming Technique. Paper delivered at The Population Council Seminar on Problems of Population Growth Estimation, Jan. 1969. Reprinted in Laboratory for Population Statistics Reprint Series, No. 14, April 1974.
11. U. S. Bureau of the Census, 1960. The Post-Enumeration Survey. 1950, Technical Paper No. 4, Washington, D. C.

Table 1: Estimated Completeness of the Census by Age and Sex; Liberia, 1974

Age Groups	Both Sexes %	Male %	Female %
All ages	89.0	89.2	88.8
Under 5 years	86.1	88.0	84.0
5 - 9 years	88.3	88.8	87.8
10 - 14 years	89.0	88.7	89.4
15 - 19 years	84.6	83.8	85.4
20 - 24 years	88.2	86.7	89.2
25 - 29 years	90.2	88.8	91.3
30 - 34 years	90.2	89.1	91.0
35 - 39 years	90.9	89.8	91.8
40 - 44 years	92.8	92.3	93.5
45 - 49 years	91.4	92.4	90.2
50 - 54 years	92.7	93.8	91.0
55 - 59 years	93.9	95.7	91.5
60 - 64 years	89.7	92.5	85.4
65 years & over	92.3	93.1	91.1

Standard errors are as follows:

Both sexes, all ages - 1.5%
 Male, all ages - 2.2%
 Female, all ages - 2.4%