PGE STUDIES: COSTS AND EFFECTIVENESS

W. Seltzer, The Population Council

1. The PGE technique

The Population Growth Estimation (PGE) technique has been employed in many countries either to estimate fertility, mortality, and natural increase where these key demographic variables are not reliably known (e.g., India, Liberia, Pakistan, Thailand, and Turkey) or to estimate the completeness of the civil registration system (e.g., Canada, Chile, Commonwealth West Indies, Tunisia, USSR, and the United States).

Nearly all of the work on the PGE technique has proceeded as if it were a technique unique to demography. Actually, an identical procedure has been used in a number of studies of response error in a variety of fields. Moreover, it is also closely related to the "capture-tag-recapture" technique used in estimating the abundance of animal populations. Note that in the capture-tag-recapture technique one estimates population size. In PGE studies we estimate live births and deaths -- the two components of natural increase -- while population size is usually measured by some form of direct enumeration.1/ Even more broadly, the PGE procedure is related to any situation where a statistic, known to be incomplete, is adjusted by a ratio reflecting the estimated completeness of the statistic.

As used in vital statistics estimation, the PGE technique involves (1) collection of reports of vital events by two quasi-independent data-gathering procedures; (2) case-by-case matching of the reports from these two systems; and (3) the preparation of an estimate of the number of events adjusted for omissions or of the relative completeness of either system on the basis of the obtained matched rates.2/

Specifically, $\hat{N}$, the PGE estimate of the total number of births or deaths occurring in a given area during a given time period, can be expressed as

$$\hat{N} = \frac{N_1 N_2}{M} \qquad [1a]$$

or

$$\hat{N} = M + U_1 + U_2 + \frac{U_1 U_2}{M}, \qquad [1b]$$

where

$N_1 = M + U_1 =$ total number of births or deaths in source 1;

$N_2 = M + U_2 =$ total number of source 2 reports;

$M =$ number of reports in each source identified (by some matching procedure) as referring to the same event, i.e., the number of matches;

$U_1 =$ number of source 1 reports identified as referring to events not reported in source 2; and

$U_2 =$ number of source 2 reports identified as referring to events not reported in source 1.

Equations 1a and 1b yield identical estimates of N. The latter equation was given by Chandrasekaran and Deming in their 1949 article [1] which recommended the use of this approach in countries lacking adequate vital registration data.

One may also form an estimate of the completeness of either source 1 or source 2. Using equation 1a, we have for the estimated completeness of source 1

$$\hat{W}_1 = \frac{N_1}{\hat{N}} = \frac{M}{N_2}. \qquad [2]$$

Equation 2 has been used in a number of countries, starting with Canada in 1931, to estimate the completeness of civil registration.

As presented in equations 1 and 2 these estimates refer to sources which attempt to gather all the events occurring in the area under study. For example, infants enumerated in a census may be searched for in a civil registration system and the completeness of the latter may be estimated using equation 2. However, one may also restrict the collection of reports to a probability sample in either or both systems. The only restriction to the use of sampling unique to PGE estimation is that the samples used in each source must be identical or one must be a subsample of the other. If sampling is employed, equations given earlier can be rewritten in terms of the sample estimates. The expected value of the estimate is not changed by the use of sampling, i.e., $E(\hat{N}) = E(\hat{n})$, but its variance is usually larger.3/

Furthermore, these equations are quite general as to the method of data collection to be used. In practice, because the PGE estimate assumes that the probability of an event reported by one source is independent of its being reported in the other source, one tries to employ two methods of data collection as dissimilar as possible. This frequently leads to the use of some type of registration approach in one source, i.e., an effort to obtain reports of events as they occur, and some form of survey approach in the other source. The survey approach collects reports of events either by asking about events retrospectively or, in multi-round surveys, by obtaining a partial count of events by accounting for changes in the household composition recorded in consecutive survey enumerations.[4]

This, in brief, is the PGE technique. Before trying to examine its strengths and limitations, let us look quickly at some of the possible alternatives.

2. Alternative approaches

Faced with the problem of using and interpreting those statistics of demographic change which, in Morgernstern's language, "simply accrue without any overall design or plan" [8] -- for example, most census or civil registration data -- demographers have tried three basic approaches: they made use of the available statistics, they tried to improve the methods of data collection, or they introduced new techniques for the analysis of data.

The first of these alternatives -- the uncritical use of whatever statistics are available -- continues to be a source of mininformation to the policy maker and confusion to the social scientist. Certainly policy makers and social scientists will not stop their activities until "good" data become available. Statisticians should respond to data requests from such sources as constructively as possible, attempting to guide the uninformed user so as to avoid the pitfalls of flawed data. However, not all of the requests for poor data come from uninformed users. The practice of various international statistical offices of requesting from each country, and then publishing, a single estimate for a long list of fertility and mortality variables is a powerful force influencing both producers and consumers of demographic statistics to act as if all such estimates were of equal substantive value.

Improvements in data collection techniques and methods of data analysis may be interrelated. However, reliance has often been placed on either alone to do the job. Thus, efforts are often made to improve questionnaire wording, training, and supervision so as to improve demographic estimates in censuses or household surveys. (The enormity of the task of improving civil registration has usually discouraged any attempts to improve this source.) In addition, more radical approaches have been tried: dual collection, pregnancy history, chemical pregnancy tests, randomized response, Sirkin's multiple respondent approach, etc.

Approaches involving improved data analysis -- which I shall refer to collectively as demographic analysis -- attempt to adjust deficient data on the basis of assumptions either about the nature of the population being studied (e.g., stable population analysis), or about the regularity of reporting errors (e.g., Brass fertility estimates, Som's recall lapse adjustments, and the Grabill-Cho method of estimating fertility from census data on own children), or both (e.g., Brass childhood mortality estimates). Some of these estimates are described in some detail in U.N. Manual IV [12] and in the papers by Page, Cho, and Zachariah presented at this session.

3. Choosing between alternatives

It is clear that a number of different methods for measuring demographic change exist. A decision to use one particular method should be based on a rational review of the alternative methods. This review should be based on a close look at the actual data needs and the resources available to carry out the measurements, rather than on one or more abstract imperatives. Three factors are needed to keep the review from being merely a formal exercise buttressing our methodological prejudices. They are: (1) comparable experience in the use of alternative methods; (2) specification of the measurement problem and the available resources; and (3) specification of one or more standards by which the choice is to be made.

Full knowledge about previous experience is required if we are to avoid past mistakes and benefit from past successes. The need for specification is of critical importance if we are to go beyond our preconceived notions. In brief, then, the need is for appropriate specification of the measurement problem and relevant knowledge about the means for its solution.

At this point let me specify five criteria for assessing the adequacy of basic demographic estimates such as the population growth rate, the crude birth rate, and the crude death rate: (1) accuracy, (2) timeliness, (3) detail, (4) user confidence, and (5) the cost of producing the estimate.

Also, let me quickly add that we do not have the techniques or the experience to apply these criteria to the alternative estimation procedures and come up with an unambiguous answer as to which procedure is preferable in a specific case. However, it is helpful to review what we do know, or think we know, about the various procedures available to us in terms of these criteria. Such a review may also help to define the criteria somewhat.

4. Costs of production

One of the truisms about PGE estimates is that they are more expensive to produce than single-system estimates. Let us leave aside for a moment that cost in the abstract, without reference to value received (in this case some mix of accuracy, confidence, detail, and timeliness), has little meaning, and compare the budget of a PGE study with that of a household survey.

Unfortunately, the analysis of cost information about data collection and analysis is an undeveloped science. The problem is threefold: (a) lack of interest, since the size of budget and the type study are often decided independently of each other; (b) lack of information on study costs, particularly for studies conducted in the developing world; and (c) the complexity of any equation that attempts to describe fully the costs of various components of the data production process.[5]

As soon as one tries to list all the factors that can affect study costs one quickly becomes discouraged by the length and varied nature of such a list. After a largely fruitless search of the literature on this topic for a scheme, or a methodology, or a notation, or a something that would both simplify the cost picture and yet preserve those features needed to contrast the costs of a single system measurement effort with those of a PGE study, I gave up.[6] All of this is by way of introduction to the inelegance of table 1 which is an unimaginative listing of all the types of activities that go into a PGE and a single-system study. Despite its awkward notation, I think the cost picture revealed by the table is helpful. If nothing else, it may stimulate others to do better.[7]

First let me explain the notation used in table 1. Annual aggregate costs are indicated by upper-case "C's", while lower-case "c's" are used for unit costs. The first-level subscripts α, β, γ, and δ refer, respectively, to study activities associated with data collection, vital events processing, base population processing, and presentation of study results. The second-level subscripts refer to a particular phase within one of the four broad types of activities designated by the first-level subscripts. The prime symbol is used to distinguish costs associated with the second collection source in a PGE study from those associated with the first source. Similarly, p and p' refer to the proportion of vital event reports sent for field investigation in source 1 and source 2, respectively. The letter m refers to the number of clusters in the sample, n to the total number of persons in the sample, and $\bar{n}$ to the mean number of persons per cluster. Assuming a crude birth rate of about 50 per 1,000 and a crude death rate of about 20 per 1,000, that some events go unreported, and that some out-of-scope events are reported, the maximum number of events reported by each system is .07n (i.e., 50 per 1,000 plus 20 per 1,000).

The cost equations in table 1 give the total cost of each type of activity (i.e., C_α, C_β, C_γ, and C_δ) in terms of the sample size (n), or the sample design (n and m), and the appropriate unit costs. Cost equations of this form permit one to assess directly the efficiency of a given sample design in terms of sampling error. Unfortunately, the equations in table 1 make no explicit recognition of differences between studies attributable to the timeliness, detail, or the accuracy of the estimates. Differences between studies involving these factors may be reflected indirectly in the unit costs of various phases of the study, as well as in the choice between single or dual collection. From the rightmost column of table 1, one can see that cost differences are limited to data collection activities and certain phases of vital events processing. Clearly, the extent to which extra collection costs are associated with dual collection will depend on the costs of the two data gathering procedures used in the PGE study and the single-system collection procedure used as a standard. To facilitate cost comparisons we assume that the first source in the dual collection system is identical to the single system source and that the sample design remains constant. Thus, all the added collection costs are associated with the second source.

Theoretically the data collection costs of the second source C_α' can take on any value; but, in practice, C_α' is usually less than C_α so that the ratio $1 + C_\alpha'/C_\alpha$ is almost always less than 2. For example, the second source may be the civil registration system, so that collection costs for the second source need cover only the cost of office sampling and, possibly, the transcription of records. Alternatively, as in the Turkish Demographic Survey, the second source also serves as the supervisory control for the first source. At a minimum, one of the sources may use an already existing infrastructure of statistical administration (e.g., the same regional field offices). To my knowledge no PGE study yet undertaken has used two fully funded, new data collection sources.

Another major determinant of data collection costs, given that source 1 is a household survey, is the frequency of survey rounds. If it is assumed that the aggregate annual data collection costs of source 1 in table 1, C_α, refers to a one-round household survey and that r_s survey rounds are carried out annually in the single-system survey then the total annual cost for data collection activities in the single-system survey is $r_s C_\alpha$. Assuming r_p rounds are employed in the comparable survey conducted in a PGE study, then the data collection costs for this survey are

Table 1 - Comparison of Cost Components for a Single-system Study and a PGE Study

[Upper-case 'C' refers to an annual aggregate cost, lower-case 'c' to a unit cost; for assumptions used and details of notation, see footnotes and text.]

Type of activity[1/]	Single-system costs	PGE study costs	Ratio of PGE total costs to single-system costs
Data Collection -- total[2/]	$C_\alpha = m(c_{\alpha_1} + \bar{n}c_{\alpha_2})$	$C_\alpha + C'_\alpha$	$1 + (C'_\alpha / C_\alpha)$
(1) Related to number of clusters	mc_{α_1}	$m(c_{\alpha_1} + c'_{\alpha_1})$	$1 + (c'_{\alpha_1} / c_{\alpha_1})$
(2) Related to number of elements	$m\bar{n}c_{\alpha_2}$	$m\bar{n}(c_{\alpha_2} + c'_{\alpha_2})$	$1 + (c'_{\alpha_2} / c_{\alpha_2})$
Vital events processing -- total[3/]	$C_\beta = .07n(c_{\beta_4} + c_{\beta_5})$	7/	8/
(1) Prematching phase[4/]	0	$.07n(c_{\beta_1} + c'_{\beta_1})$	∞
(2) Matching phase	0	$.07n(c_{\beta_2} + c'_{\beta_2})$	∞
(3) Field follow-up[5/]	0	$.07n(pc_{\beta_3} + p'c'_{\beta_3})$	∞
(4) Pretabulation phase[6/]	$.07nc_{\beta_4}$	$.07nc_{\beta_4}$	1
(5) Tabulation phase	$.07nc_{\beta_5}$	$.07nc_{\beta_5}$	1
Base population processing -- total	$C_\gamma = n(c_{\gamma_1} + c_{\gamma_2})$	C_γ	1
(1) Pretabulation phase[6/]	nc_{γ_1}	nc_{γ_1}	1
(2) Tabulation phase	nc_{γ_2}	nc_{γ_2}	1
Presentation of results -- total	C_δ	C_δ	1
All activities -- total	$C = C_\alpha + C_\beta + C_\gamma + C_\delta$	$C = C_\alpha + C'_\alpha + C_\beta + C_\gamma + C_\delta$	9/

1/ Includes stated activity, plus proportional share of costs of supervision and overheads. The cost analysis presented here makes no explicit recognition of expenditures incurred to increase the accuracy, timeliness, or detail of the estimates.

2/ It is assumed that collection costs are related to the number of clusters (m) and their mean size ($\bar{n}$), rather than to the level of vital events. For simplicity, all collection costs are treated as if they were solely linear functions of m and $\bar{n}$.

3/ It is assumed that vital event processing costs are related to the number of birth and death reports obtained (i.e., a maximum of .07n).

4/ Editing and other processing necessary to make the documents ready to matching. The matching is assumed to be done manually.

5/ As a first approximation field follow-up costs are assumed to be related to the proportion of reports from each source (p or p') sent for follow-up. Follow-up costs will usually be described more accurately by $mc_{\alpha_1} + .07n(p + p')c_{\alpha_2}$.

6/ Editing, coding, and punching required for purposes of tabulation.

7/ $C_\beta = .07n\ (c_{\beta_1} + c'_{\beta_1} + c_{\beta_2} + c'_{\beta_2} + pc_{\beta_3} + p'c'_{\beta_3} + c_{\beta_4} + c_{\beta_5})$.

8/ $$\text{Ratio} = 1 + \frac{c_{\beta_1} + c'_{\beta_1} + c_{\beta_2} + c'_{\beta_2} + pc_{\beta_3} + p'c'_{\beta_3}}{c_{\beta_4} + c_{\beta_5}} .$$

9/ $$\text{Ratio} = \frac{m(c_{\alpha_1} + c'_{\alpha_1}) + n(c_{\alpha_2} + c'_{\alpha_2}) + .07n(c_{\beta_1} + c'_{\beta_1} + c_{\beta_2} + c'_{\beta_2} + pc_{\beta_3} + p'c'_{\beta_3}) + n(c_{\gamma_1} + c_{\gamma_2}) + c_\delta}{mc_{\alpha_1} + nc_{\alpha_2} + .07n(c_{\beta_4} + c_{\beta_5}) + n(c_{\gamma_1} + c_{\gamma_2}) + C_\delta} .$$

$r_p C_\alpha$.[8/] It follows that ratio of total costs for a PGE study to that for a single-system survey is approximately

$$\frac{r_p}{r_s} + \frac{C'_\alpha}{r_s C_\alpha} \qquad [3]$$

or, if C_α is assumed to equal C'_α,

$$\frac{r_p + 1}{r_s} \quad . \qquad [4]$$

In other words, if more than three survey rounds are contemplated per year (i.e., $r_s > 3$) it is possible that meaningful savings in data collection costs can be achieved by reducing the number of survey rounds and employing an appropriate dual collection procedure.

The expense associated with matching and field follow-up is unique to a PGE study. Unless the unit costs of these operations are very high, the relatively small numbers of cases involved, .07n and (p + p´).07n, suggest that the aggregate costs of these operations are only moderate compared to the total cost of a multi-round survey. Nevertheless, there is an urgent need for additional data on the costs of matching and field follow-up.

To indicate some idea of the range of costs involved in demographic field studies in the developing world let me cite two figures. The annual cost of the Pakistan PGE study came to $6.50 per household. This estimate is based on aggregate cost data covering all aspects of the study and all sources of funding. In the Pakistan study data collection continued for four years and the sample involved some 20,000 households so that the impact of necessary overheads on the annual average is not large. The second figure comes from another country in the developing world where the cost of a 27,000 household multi-round demographic survey extending over two years came to $16 per household per year. This costs estimate does not make any provision for the costs of tabulation or the presentation of results. Another major difference between these two cost figures is that personnel costs were quite low in Pakistan relative to those in the country in which the multi-round survey was conducted.

5. User confidence

In our real world of uncertainties and mistakes, user-decisions on confidence involve processes that are far from being either rational or accurate. In fact, there are many instances where the most "rational" procedures for determining the confidence to be placed in a particular estimate are not the most accurate.

Certainly the establishment of a confidence interval around a crude birth rate estimate in order to reflect the uncertainties introduced by sampling is an objective and rational procedure. However, most demographers would consider that the 1962 crude birth rate for rural India lay well outside the range of 32.6 to 36.6 per 1,000, even though this is the 2σ confidence interval of this estimate from the Indian National Sample Survey [9]. The point here is the simple one that the probabilistically determined consequences of random errors are not the only factors which should affect the confidence we place in any demographic estimate.

Though an ideal procedure for ascertaining confidence is well beyond us, we might try to approximate some measure of this concept in terms of (a) the likely accuracy of a statistic given the various types of errors it may be subject to, and (b) the likelihood of these errors occurring. I realize that even this somewhat loose and limited goal will be hard to achieve. Quite often it is the figure with no error statement attached on which the user places his greatest confidence. However, even if the policy makers are slow to heed technically sound assessments of data quality our fellow scientists should not be.

6. Detail and timeliness

I shall touch only briefly on the detail and the timeliness criteria as the paper by Louwes [5], deals extensively with these two factors. Details may refer either to the types of variables covered (e.g., crude rates, characteristic-specific rates, life table rates), or to the extent to which estimates are made for various analytical or geographic subgroups. The effect of increased detail on the cost equations of table 1 varies somewhat with the type of detail under consideration. For example, added geographical detail generally will necessitate an increase in the number of clusters (m) and hence data collection costs, as well as increasing tabulation and data presentation costs.[9/] On the other hand, increased analytical detail (e.g., obtaining estimates of fertility) often will have only a marginal impact on collection costs. However, beyond a certain point increased analytical detail can also affect the cost and quality of the collection operation.

While timeliness is usually thought of in terms of speed of production, frequency and regularity are also factors involved in the concept of timeliness. In general, there is a reciprocal relationship between speed and detail as well as speed and accuracy. On the other hand, frequency of data collection, and to a lesser extent regularity, tend to be directly related to accuracy. Similarly,

Table 2 - Mean and Range of Approximate Intraclass Correlation Coefficients for Crude Birth and Death Rates, by Type of Cluster, for Six Specified Studies: 1950-66.

[For full qualifications, see sources cited. Values of δ are approximate and are rounded to 3 places.]

Type of cluster	Number of domains 1/	Mean population per cluster 2/	Crude birth rate: Mean δ 3/	Crude birth rate: Range of δ: Low	Crude birth rate: Range of δ: High	Crude death rate: Mean δ 3/	Crude death rate: Range of δ: Low	Crude death rate: Range of δ: High
All Types	46	572	+.002	-.001	+.008	+.003	-.001	+.013
Region and country								
Africa	33	333	+.002	-.001	+.008	+.003	-.001	+.013
Cameroon, 1960-65	23	356	+.001	-.001	+.005	+.002	-.001	+.010
Chad, 1964	7	300	+.001	-.001	+.005	+.005	+.000	+.013
Nigeria, 1965-66	3	235	+.005	+.004	+.008	+.005	+.002	+.012
Asia	13	1,179	+.002	+.000	+.005	+.002	-.001	+.006
India, 1950-52	4	455	+.002	+.000	+.004	+.004	+.000	+.006
Pakistan, 1964-65	2 4/	5,000	+.002	+.001	+.002	+.001	+.001	+.002
Turkey, 1965-66	7	501	+.003	+.001	+.005	+.001	-.001	+.002
Type of residence								
Urban	6	498	+.002	+.001	+.005	+.001	-.001	+.005
Rural	38	351	+.002	-.001	+.008	+.003	-.001	+.013
Mixed	2	5,000	+.002	+.001	+.002	+.001	+.001	+.002
Cluster size 5/								
Under 300	11	275	+.001	-.001	+.008	+.005	+.001	+.013
300-349	19	323	+.002	-.001	+.005	+.003	-.001	+.009
350-649	12	457	+.003	+.000	+.005	+.001	-.001	+.010
650 and over	4	2,917	+.002	+.001	+.002	+.002	+.001	+.005

1/ A domain is a group of clusters for which the intraclass correlation coefficient is separately available. Domains often correspond to sample strata.

2/ Mean of average cluster size reported for each domain in original source. Cluster size shown is that prior to additional within-cluster sampling, if any.

3/ Mean of unrounded intraclass correlation coefficients for specified number of domains.

4/ Each province is treated as a domain, with intraclass correlation coefficient based on the average survey and registration values for 1964 and 1965.

5/ Reported mean population per cluster of each domain.

Sources:

(a) Cameroon:

Scott, Christopher, "Vital Rate Surveys in Tropical Africa," in The Population of Tropical Africa, edited by J. Caldwell and C. Okonjo, London, 1968, Chapter 15, table 1, pages 164-165.

(b) Chad and Nigeria:

Scott, Christopher and J.B. Coker, "Sample Design in Space and in Time for Vital Rate Surveys in Africa," paper presented at the International Union for the Scientific Study of Population, London, 1969, tables 1 and 2. For Nigeria, estimates are based on artificially constructed clusters of 50 consecutive household questionnaires completed by the same interviewer.

(c) India, Pakistan, and Turkey:

Intraclass correlation coefficients calculated from published variance estimates from the Mysore Population Study, the PGE Experiment in Pakistan, and the Turkish Demographic Survey, using Scott and Coker's binomial approximation.

the more detailed statistics one has available from a study the more confident one can usually be about assessing its quality.

7. Accuracy

Accuracy as used here is a synonym for data quality and is measured in terms of the difference between an estimate and the value one is trying to estimate. Defined in this way the accuracy of an estimate is affected by both random and nonrandom errors, whether arising in the collection, processing, estimation, or presentation process.

The special sources of error unique to the PGE technique are:

1. lack of independence between the two collection procedures which, except in rare circumstances, can lead to an underestimate of the number of events;

2. use of matching criteria which fail to distinguish between reports referring to different vital events, leading to erroneous matches and an underestimate of the number of events; and

3. use of data in the matching process containing reporting or recording errors so that reports referring to the same event are not linked, resulting in erroneous nonmatches and an overestimate of the number of events.

In addition, the existence of out-of-scope reports in one or both sources or the use of a deficient estimate of the base population can lead to an upward bias of the PGE vital rate estimates. However, these two sources of error also affect most types of single-system estimates. All of these sources of error in the PGE estimate are discussed in much greater depth in Seltzer and Adlakha [11], Marks [6], and Marks _et al_. [7].

The principal advantage of the PGE technique is that the PGE estimate is largely unaffected by the errors and the uncertainties encountered in the collection phase of many single-system surveys and registration systems. Whether a single system is used to provide a vital rate estimate directly or is used as a source of data for demographic analysis, the amount of information available about the population being studied is limited to that obtained from the single source. Dual collection and matching by its very nature provides more information than is available from a single source. While the amount of information from a single source may be stretched by the use of suitable assumptions, the accuracy of the estimates made are then subject to both data errors and errors arising from the failure to meet the assumptions made.

In addition to the sources of error listed above, most PGE estimates are also subject to sampling variability.10/ Estimates of sampling error from the Pakistan PGE experiment indicate coefficients of variation between 4 and 9 percent for the annual crude birth rate estimate, with values approximately twice this size for the crude death rate. Because PGE studies often involve some type of registration of vital events, clustering is often more pronounced than is traditionally employed in survey sampling. Clearly, we pay a price for this clustering as the estimates of intraclass correlation coefficients presented in table 2 indicate.

Even with a δ as small as those shown in table 2, very large clusters will have a major impact on sampling variance; but sampling error is only one component of our accuracy criteria. Indeed, one might hypothesize that, in general, the smaller the cluster for a fixed budget and fixed total sample size, the larger would be nonsampling error. In other words, one supposes that a given supervisory effort is spread more thinly when the sample is based on a large number of small clusters than when an equal-size sample is composed of a smaller number of larger clusters.

I know of no data available to test this hypothesis directly and even if the relationship is established in one instance, there is no guaranty that any observed relationship between cluster size and nonsampling error will remain constant from study to study. Nevertheless, the design implications are important enough that some effort to test this hypothesis should be made.

I do not doubt that an intensive, well-run, single-visit retrospective survey can come up with high-quality demographic estimates. The problem is how can we rely on its accuracy in any given instance? The evidence is not encouraging. One study found a median 33 percent undercount of the number of births reported in one-time retrospective surveys relative to that of comparable dual collection estimates [10].

Finally, in attempting to improve the accuracy of any collection procedure we are caught in the dilemma of how much effort to spend per household to better the quality of data collection versus how many households should be sampled; that is, the choice of allocating limited resources to reduce nonsampling or sampling errors. Basically, dual collection provides a highly effective means of spending more per household so as to concentrate on the reduction of nonsampling errors.

8. Summary: effectiveness

My objective in this paper has been to compare the relative effectiveness of the PGE technique with that of some alternative procedures for obtaining estimates of basic demographic variables. Based on present knowledge, it is not possible to construct utility functions to this end that are both meaningful and rigorous.11/ Nevertheless, by specifying five factors (i.e., accuracy, timeliness, detail, user confidence, and cost) that might ultimately compose such a utility function and by examining alternative estimates in light of these factors I believe the goal of rigor has been advanced somewhat. At the same time, the introduction of experience from actual studies has kept the discussion from wandering too far from reality.

Unless the purposes for which demographic estimates are prepared are also adequately specified the concept of effectiveness has limited meaning no matter how rigorously this concept is defined. The question of proposed uses of demographic data has not been dealt with explicitly in this paper. However, the range of possible uses for demographic estimates is broad enough to guarantee that no technique can be termed, "universally most effective."

Given this general limitation, the findings of this paper can be summarized in terms of the five criteria of effectiveness as follows:

1. Accuracy -- The PGE technique is as good as the best of the alternatives; nevertheless the precision of our measurement techniques is such that small year-to-year changes in fertility and mortality can not be measured accurately in countries without an effective civil registration system.

2. Detail -- The PGE technique is as good as the best of the alternatives and, except in the case of historical data, provides more extensive detail than demographic analysis.

3. Timeliness -- The PGE technique will usually provide estimates more slowly than a one-time retrospective survey, more quickly than the 6 to 7 year lag between the mid-decade reference point of an intercensal growth rate and the date that such a growth rate becomes available subsequent to the census, and at about the same time that a good-sized multi-round household survey produces comparable estimates.

4. User confidence -- The informed user will find the PGE technique far ahead of the other available alternatives with respect to the degree of confidence that can be placed in the estimates. The fact that the PGE technique provides a built-in self-evaluation device -- through dual collection and matching -- does not guarantee that PGE estimates will be correct, or that the user will realize that any given estimates are quite incorrect. There is with the PGE technique, however, a much greater likelihood of realizing something has gone wrong, if it has, as well as of producing estimates that are moderately robust to the variations in the quality of data collection. This is an important point because sharp variations in the quality of field work are frequently encountered where data collection experience is limited or where field conditions are particularly difficult.

5. Costs -- In terms of cash outlay the PGE technique is usually, though not necessarily, more expensive than alternative approaches using comparable-sized samples. Whether the additional cost is justified depends upon the uses to which the estimates will be put. However, the difference in cost between a multi-round survey and a PGE study is generally not that much, so that whenever a multi-round demographic survey is contemplated, very serious consideration should be given to conducting a PGE study.

In order that we proceed beyond the tentative formulations of this paper we will need additional data on the costs, the accuracy, and the uses of various types of demographic estimates. Therefore, I would like to close with the request that statisticians concerned with demographic measurement increasingly turn their attention to identifying the costs, accuracy, and ultimate uses of the estimates they produce. In this request I am merely echoing some of the recommendations made two years ago by Ross Eckler [4] in his Presidential Address to this association.

FOOTNOTES

1/ However, see [3].

2/ Estimating equations for use in a three-source PGE study have been given by Deming and Keyfitz [3] and, independently, for a k-source study by Das Gupta [2]. Because it will usually be more efficient to improve the quality and independence of two collection systems than to attempt to use a third collection procedure, this paper is confined to an examination of PGE studies using only two sources.

3/ If the completeness of a source using no sampling is very poor (say, less than 50 percent) it may be desirable from a variance viewpoint to use a source with sampling and higher completeness.

4/ Unfortunately, not all events can be identified by examining changes from survey round to round in the list of persons enumerated in the household. Particularly in countries where infant mortality is high, migration rates are high, or women spend long periods of time at their parents home after childbirth, reports about a considerable number of events can only be obtained retrospectively.

5/ The classical approach of allocating collection costs between those associated with the number of clusters and those associated with the number of elements does not really help to assess designs which have major differences subsequent to the data collection phase.

6/ The major exception to this bleak picture is a paper by Louwes [5]. While Louwes' paper deals with agricultural surveys in the European common market countries and is thus not directly relevant to the problem at hand, it does suggest a number of promising leads, one or two of which are used in this paper.

7/ The basic monthly salary of the survey interviewer, or its hourly equivalent, seems to be a very promising standard unit from which comprehensive cost function can be built, thus permitting the kind of cost comparison suggested in this section.

8/ This is almost certainly an overestimate of the costs of a multi-round survey in that it assumes that data collection activities involve only recurring costs. However, the effect of this overestimate on the cost comparison of a PGE study with a single-system survey may be at least partially offset by assuming $C_\alpha = C'_\alpha$ as in equation 4.

9/ Of course, if sampling is not involved -- as is the case with a census or a national civil registration system -- no additional data collection costs are associated with increased geographical detail.

10/ PGE studies and single-system estimates based on retrospective survey questions are not alone subject to sampling errors. It is often unrecognized that many forms of demographic analysis are also subject to the effect of sampling errors. For example, the parameters used to enter stable population tables may be subject to sampling variability, implying a range of possible stable population estimates.

11/ In this formulation, utility is considered to be a joint function of the uses to which the estimates will be put and the effectiveness (in terms of the five criteria described in the paper) of the collection and estimation effort.

REFERENCES

[1] Chandrasekaran, C. and Deming, W.E., "On a Method of Estimating Birth and Death Rates and the Extent of Registration", Journal of the American Statistical Association, Vol. 44, No. 245, 1949, pp. 101-115.

[2] Das Gupta, P., "On the Estimation of the Total Number of Events and of the Probabilities of Detecting an Event from Information Supplied by Several Agencies", Calcutta Statistical Association Bulletin, Vol. 13, No. 49 and 50, 1964. pp. 89-100.

[3] Deming, W.E. and Keyfitz, N., "Theory of Surveys to Estimate Total Population" in Proceedings of the World Population Conference, Belgrade,... 1965 -- Volume III, United Nations, New York: 1967, pp. 141-144.

[4] Eckler, A. Ross, "Statisticians and Shoemakers ('Who is Worse Shod than the Shoemaker's Wife' from Heywood's Proverbs)", Journal of the American Statistical Association, Vol. 65, No. 329, 1970, pp. 9-21.

[5] Louwes, S.L., "Cost Allocations in Agricultural Surveys", Review of the International Statistical Institute, Vol. 35, No. 3, 1967, pp. 264-290.

[6] Marks, Eli S., "Problems in Estimating Population Growth in Developing Areas", paper presented at the seminar on Statistical Problems in Population Research, The East-West Center, Honolulu, August, 1971.

[7] Marks, Eli S., Krotki, K., and Seltzer, W., Population Growth Estimation: A Handbook of Vital Statistics Measurement, The Population Council, New York, forthcoming.

[8] Morgernstern, Oskar, On the Accuracy of Economic Observations, Second Edition, Completely Revised, Princeton University Press, Princeton, 1963 (Princeton Paperback Edition, 1965).

[9] Ramabhadran, V.K., "Experience and Problems in the Creation of the Sample Registration System in India", paper presented at the seminar on Statistical Problems in Population Research, The East-West Center, Honolulu, August, 1971.

[10] Seltzer, William, "Some Results from Asian Population Growth Studies", Population Studies, Vol. 23, No. 3, 1969, pp. 395-406, table 3.

[11] Seltzer, W. and Adlakha, Arjun, "On the Effect of Errors in the Application of the Chandrasekaran-Deming Technique", paper presented at the seminar on Population Growth Estimation, The Population Council, New York, January, 1969.

[12] United Nations, Methods of Estimating Basic Demographic Measures from Incomplete Data, Manual IV, Population Studies No. 42, U.N. Department of Economic and Social Affairs, New York: 1967.