rotating samples instead of censuses

by Leslie Kish

In spite of this provocative title, I must admit that I am not necessarily advocating here replacing altogether the decennial censuses of population, housing, agriculture, and others, which have been adopted and approved during the past few decades in most countries of the world. These censuses have represented great progress in mankind's self-knowledge, progress greatly facilitated by the United Nations. We want further progress, not retrogression. During the same period, the techniques and the utilization of sample surveys have made even greater advances—which have been facilitated (in materials, data, techniques, and technicians) by the growth of censuses. Simultaneously, sampling techniques have done much to improve census procedures. Thus, the existence and progress of samples and of censuses is both symbiotic and competitive (Kish, 1979).

Perhaps in the future even the decennial censuses will be replaced by larger samples and by better registers. But those changes would now face in most countries technical, legal, administrative, and political obstacles. Competition, however, between censuses and samples is now active and is bound to become even more active in the near future. The use of administrative registers also enters into this competition, but they should be omitted from this discussion because registers are so diverse and national resources and situations are so different that it is too difficult to give a general yet brief treatment of them.

We are in an era of both coexistence and competition between better censuses and more and better survey samples, and it is a dynamic era with changing aspects. First, advances in census taking have encouraged demands for ever more information from censuses. These demands may lead to placing undue burdens on the censuses, increasing both the numbers and the difficulty of census questions. Such burdens may lead to lower overall quality as well as greater expense. Second, increased census burdens have led to sampling of the census questions. Complete censuses are becoming limited to a small core of "simple," basic questions, and most of the census questions are becoming confined to "partial censuses" (that is, samples of the entire census). Third, there are plans to shift some of the burdens of censuses to "minicensuses," which can be more frequent and timely, more versatile, and richer in content, as well as cheaper than complete censuses.

"Minicensuses" and "microcensuses" are terms that have been used recently to describe samples that differ from the partial censuses attached to complete censuses. For example, there are plans in the United Kingdom for a 1 percent yearly sample to yield basic census data. Minicensuses would typically be smaller than partial censuses and would also lack the financial and political support that censuses have. But they can have greater flexibility in timing, in staffing, and in methods than complete censuses with their rigid requirements. Minicensuses, however, would also differ from most sample surveys: they would be larger, they would collect more basic data, and they would be periodic.

Need for current, detailed, diverse data

Censuses yield great detail, but their data are neither current nor diverse enough to meet administrative demands of today's dynamic and mobile populations. Obsolescence of two to 12 years—seven years is the average—is common now for census data, and these data often seem too out of date for modern needs. Conducting quinquennial censuses would shorten delays only to a range of two to seven years, an average of 4.5 years, while doubling costs (though delays are being reduced, often they are even greater than those mentioned above [Waksberg, 1968]). The cost of quinquennial censuses may be lowered by taking a sample census, as in the 10 percent census of the United Kingdom in 1975. Annual censuses would provide more timely data but at great cost. To the pressure for timely data are added demands for more diverse and richer data on more variables.

Survey samples are used ever more widely to yield diverse, rich, and accurate data, and they can be made timely.

Dr. Leslie Kish is Professor and Research Scientist at the University of Michigan's Institute for Social Research. The foremost expert on sampling techniques, Dr. Kish has for 20 years been Director of the Sampling Program for Foreign Statisticians, also at the University of Michigan. The Program has graduates in 83 developing countries of the world. Dr. Kish is the author of Survey Sampling, published in 1965 by John Wiley and Sons.
and current. They have great flexibility—in timing, definition, collection, and other aspects—but they are typically too small (for reasons of economy) to yield the great detail needed by administrators, especially for local area estimates. Administrative registers can be (at best) current and detailed, but they seldom have the needed diversity and accuracy; and they are altogether inadequate in most situations.

Thus the triple demands for detailed, current, and complete (rich, diverse, accurate) data remain typically unsatisfied (Kish, 1979). Those demands are becoming increasingly urgent. To the curiosity of statisticians and demographers have been added the needs of social planners and of administrators, and the commands of political bodies for bases for dispensing funds and making other allocations in accord with detailed and current data.

What can one do with obsolete data? Commonly, they are used—old as they are—after useless disclaimers and with large biases against fast-growing places. Adjustments based on smooth mathematical-demographic models cannot keep up with the vagaries of irregular migrations.

New methods of postcensal estimation for small domains have greatly improved this situation (Purcell and Kish, 1979, 1980; Eriksen, 1973, 1974; Gonzalez and Hoza, 1978). But these estimates also need some source of new data; hence they are partly symbiotic and partly competitive with the periodic surveys proposed below. Their results can be further improved by using the estimation techniques.

Rotating sample censuses
First let us describe a basic rotating sample, and then we can justify, compare, and modify it. A rotating, nonoverlapping monthly sample of 1:120 would use an area sample, probably census enumeration areas (EAs), or perhaps sub-divisions of them for finer detail in small domains (administrative areas). Regular, balanced area samples would be designed for periodic summings. For example, 12 monthly samples would give yearly averages based on 12:120 = 10 percent samples.

Periodic samples can also be used for month-to-month comparisons. Though overlaps of the monthly samples would not be necessary, they could greatly reduce the variances of monthly changes. For example, a monthly sample of 1:100 could have overlaps of 1:6, yet have fresh samples of 5:6, and still contain the 1:120 rotating new samples. This method would yield modest reductions in the variances of monthly changes. For larger reductions one could have samples of 1:80 with 1/3 overlaps and 2/3 rotation; or samples of 1:60 with 1/2 overlaps and 1/2 rotation. (The reduction in variances is 1(1−R)/(1−PR) for simple estimates of changes and 1−(1−R)(1−PR) for weighted estimates, where P is the overlap and R the correlation between months [Kish, 1965:464].)

Increasing the overlaps increases the costs—and perhaps the complexity—of the samples. But at the same time it yields a valuable new service, which also may be greatly needed. If, on the other hand, a country already conducts large-scale, periodic surveys, these surveys may become the vehicle for the needed detailed minicensuses. In such cases the cost of the minicensus should be balanced against the combined costs of the periodic surveys plus the cost of a planned quinquennial census. How close the size of the periodic samples comes to the needs of the minicensus depends not only on the size of the samples but also on the size and needs of the population; a large periodic sample could be adequate for details in a small population but not in a large population.

Here we should add that it is not necessary for the minicensus to cover the entire country in ten years. For example, a monthly rotating sample of 1:600 would yield yearly samples of 12:600 = 2:100 each year and 20 percent in ten years. The sample could cover every fifth area (EA), or one-fifth of all areas, a coverage larger than that of most partial censuses in decennial censuses.

Though our basic model concerned monthly samples, it is possible that weekly samples, or, contrariwise, quarterly samples, would be more convenient. Instead of a monthly sample of 1:120, a weekly rotating sample of 1:520 would also cover the nation in ten years, and so would a quarterly sample of 1:40. Administration of data collections should dictate the choice, and I will say more about this later. Any of these kinds of samples could be used to obtain yearly and similar summaries. Of course, weekly or quarterly surveys would lend themselves to overlapping designs for improving monthly or other comparisons.

It is not necessary to have the same sample design over the entire population. In fact there may be good reasons for having different procedures and sampling rates for diverse provinces, for urban versus rural strata, or for rapidly growing sections and districts. This flexibility is much greater than for censuses and should be exploited. (Of course, any departure from uniformity that increases complexity bears heavy burdens of justification.) Such flexibility allows better estimates for critical, growing areas without introducing bias. On the contrary, the common practice now of having special censuses conducted between decennial censuses for places that claim unusual growth introduces bias against competitors that fail to make similar claims.

Collection, reference, and reporting periods
Though the periods of collection, reference, and reporting are often confused, it is clear that they can and should be separated for flexibility in designing periodic surveys. For example, the collection period could be the first week of a month, but the reference period could be the previous week for some data (employment), the previous month for some other data (rent), and the whole previous year for still (continued on page 12)
JAPAN

- The next census of Japan will be held 1 October 1980, and a report on the progress of preparations for that census was sent by Mr. Shotaro Yanagawa, Assistant Director of the Population Census Division of Japan’s Statistics Bureau. A Planning Committee, set up in December 1978, has seven subcommittees, each with a different area of responsibility for planning the census. One subcommittee, for example, determines and establishes enumeration districts; others plan the publicity campaign for the census and design tabulation methods and procedures. The overall Planning Committee, composed of leading officials of the Statistics Bureau, examines detailed reports of the subcommittees and frames plans for execution of the census.

Three pilot surveys for the 1980 census have been held, and a fourth, scheduled for October and November, will be a dress rehearsal conducted in 21 prefectures. Among the concerns already tested in the pilot surveys: reliability of returns using questionnaires completed by householders; reading efficiency of the Optical Character Reader developed in Japan; and the feasibility of using three kinds of questionnaires with somewhat different items, for households in the same enumeration district.

The fourth pilot survey is to be expanded to cover about five times as many respondents as the third survey. The questionnaires in this final survey will be read through the Optical Mark Reader and tabulated as a test case. The questionnaire will contain 23 items.

TAIWAN

- The third complete population census of Taiwan since World War II is scheduled for 16 December 1980, marking a change in the census series for that country. Censuses were taken in 1956 and 1966, and the decision was made to change census taking to years ending in zero to conform to international practices and facilitate comparison of Taiwan’s population with that of other countries. Sample censuses of population and housing were conducted in 1970 and 1975.

News of next year’s census was provided by Mr. Tze-Hwa Fan of Taiwan’s Bureau of Statistics. He writes that a Census Planning Group, consisting of scholars, experts, and representatives of governmental agencies, was set up in July 1979 to launch various designs and formulate plans for the 1980 census.

A pretest is scheduled for one year before the census date to test various aspects of the census, including the census schedule and instructions to enumeration personnel. Taiwan plans to have both de facto and de jure enumeration: nationals of Taiwan residing within the boundaries of the Taiwan-Fukien area (including military and nonmilitary officers stationed abroad and their dependents) and aliens residing in Taiwan (not including foreign military and nonmilitary officers and their dependents) will be counted, as will the housing units of the above-mentioned persons.

Enumeration will consist of preliminary enumeration before the standard census time (the zero hour of 16 December 1980) and verification enumeration within five hours after the standard census time. A postenumeration survey is planned immediately after the census to detect and measure errors.

REPUBLIC OF KOREA

- Mr. Young Kwon Kim, Director of the Population Statistics Division of Korea’s National Bureau of Statistics, wrote with news of preparations for the 1980 census of Korea. He says that a questionnaire was designed early this year on the basis of information collected from about 250 users and potential users of census data. The questionnaire has been tested in two pretests and is to be finalized following another pretest at the end of this year under controlled conditions that will simulate the actual census.

Several new items of information are under consideration, including address of office or school, means of transportation, and handicaps and disabilities. “It is well known,” says Mr. Kim, “that it is difficult to obtain reasonably accurate responses to a question on the handicapped and disabled. But we are under heavy pressure to include this question from users of data, especially the Ministry of Health and Social Affairs.” A final decision will be made after the last pretest.

In the area of census mapping, Mr. Kim reports that base maps are being prepared mainly from aerial photographic maps for large cities and from topographic maps for other areas. Delineation of enumeration districts for rural areas will be completed by the end of this year. For urban areas, where changes are more frequent, the districts will be established by May 1980.

Twenty percent of the population will be given questionnaires with a larger number of items. In the 1975 census, the sample size for collection of additional information was only 5 percent. A major emphasis on publicity is planned for the 1980 census, according to Mr. Kim. Plans are to spend about 10 percent of the total census budget for publicity. The publicity program has already been formulated and materials are being prepared by different publicity media.

Korea plans to replace its IBM 370 computer with an IBM 3031 to expedite data processing in the census. “Studies are also being made,” says Mr. Kim, “to look into the possibility of using the CONCOR package program not only for editing but also for automatic correction at the later stage of editing.”

INDONESIA

- Indonesia plans to conduct its coming census of population in October 1980. Dr. Sam Suharto, Director of the Data Processing Center of Indonesia’s Central Bureau of Statistics (CBS), described some of the main preparations.

Mapping. Fieldwork on mapping the entire country will begin in September 1979. Maps will be made for each of about 180,000 enumeration areas (EAs). Each EA will then be divided into census blocks of about 50 households, and the census blocks will be delineated on the maps. A list will be produced of all census blocks within each EA by each village and within each district.

Formation of a master sampling frame. The list will be used as the base for the formation of a “master sampling frame,” a sample of about 15 percent of all the census blocks selected so that it can be used as a frame for all sample surveys carried out during the 1980s. The frame must represent the demographic, social, and economic characteristics of the population. It must also rep-
resent the geographical distribution of the population or households as well as their economic activities. The master sampling frame will be stored in the computer and will be updated from time to time.

Census enumeration. Indonesia will use two types of enumeration: a complete census enumeration and a sample census enumeration. Indonesia consists of about 13,000 islands, about 6,000 of which are inhabited, so it is necessary to plan everything far ahead of time, to minimize problems during the enumeration. The forms for the complete enumeration are of two types: Form L1 for the housing information and Form L2 for household members’ information. As many as half a million copies of Form L1 will be printed, one form for each census block, whereas about 35 million copies of Form L2 will be printed. The printing of these forms will start in September 1979; they must be ready for distribution by May 1980.

The sample census form is being tested in an August pilot census. It will then be used in October in a census rehearsal, which will provide a simulated census operation exactly one year before the actual census. One use of the results will be to test data for the development of the data processing system and programming. The form should be almost final for the census rehearsal, as only minor changes can be made before it goes to the printer in January 1980.

Enumerators in the sample census will act as trainers and supervisors for the enumerators in the complete census. The enumeration period will last for about one month with census day falling on 30 October 1980.

Data processing developments. Parallel to the planning of the census field operations, planning of other activities has also taken place, including the organization, infrastructural facilities, and data processing equipment. For the coming population census, CBS is building a 2000-square-meter building to house a new computer. By January 1980 CBS expects to install two ACOS 500 NEC computers, each with one megabyte of memory. Negotiations with the vendors on various technical aspects are currently taking place, and staff training in the use of the new computers is being carried out. The NEC computers will be in addition to the current ICL 1930S, which will still be used to support the routine workload.

The processing of the complete census initially will be done manually in the field. Village summaries will be sent to the central office for input to the computer. The processing of the sample census will all be done on the new computer. Input to the computer will be through key to diskettes data entry equipment. It is planned that the advance tabulation will be completed by mid-1981 and the final tabulation by mid-1982.

**TOKELAU**

- In the paper by Drs. Ko Groenewegen (“Censuses in the South Pacific, 1976–77,” *Asian and Pacific Census Forum*, Volume 5, Numbers 3 and 4), a footnote should have been added to draw attention to the census held in Tokelau on 25 October 1978. This census was not included in the paper because insufficient information was available at the time it was written. Since then, Drs. Groenewegen has received a copy of the census form and of the census tabulations from the Department of Statistics, New Zealand, but instruction manuals and other census documents have not yet been provided.

Drs. Groenewegen reports that the census questionnaires consisted of a dwelling schedule and a personal schedule. The former asked for name of village, name of occupier or person in charge, total number of occupants on census night (by sex), and materials of which the dwelling was built, especially material of outer walls and roof. The personal schedule asked for name, sex, and age (in years and completed months) of the person. Individuals also had to report whether they were married, never married, widowed, or divorced and what “race” they belonged to (e.g., Tokelauan, Samoan, Ellice). Questions on country of birth, religion, and occupation were also included on the questionnaire. Females 15 years old or older were asked to report the number of children born alive to them and the age when their first child was born.

The enumeration was a de facto count. It is unlikely that any sampling activity was included, because the population is far too small to make sampling efficient. Data processing was carried out manually by the New Zealand Department of Statistics in 1977, but the resulting provisional tables were not published in a separate report.

The three atolls of Tokelau—Atafu, Fakaofo, and Nukunonu—together had 244 dwellings and a population of 1,575 persons, of which 747 were males and 828 females. The total shows a decrease of 24 persons from the 1972 census figure. The population is young: 46 percent are under the age of 15 and 62 percent are under 25. The shortage of males is very heavily concentrated in the 25–45 age range, indicating the influence of out-migration. Of all males aged 15 and over, 85 percent are gainfully employed; 44 percent of these are classified as plantation workers. Only 10 percent of females 15 and over are reported as gainfully employed.
CENSUS QUESTIONS ON FERTILITY AND CHILD MORTALITY: PROBLEMS WITH QUESTIONNAIRE DESIGN

by Judith Banister

The wording of census questions and the order in which they are asked are generally thought to have strong effects on responses to the questions. Yet we seldom encounter a clear example that demonstrates how changes in wording or order of census questions have indeed elicited very different answers from respondents. I encountered one such example recently while analyzing data on children ever born and children surviving from five consecutive censuses of Western Samoa. This particular instance identifies problems that may arise when designing census questions and suggests ways in which we might be able to detect and correct such problems. As Asian and Pacific countries approach their 1980 round of censuses, they should give careful attention to avoiding these potential pitfalls in questionnaire design.

Infant mortality estimates for Western Samoa

Western Samoa is an island country in the southwestern Pacific with a population of about 150,000 people. It conducts a census every five years. The censuses before 1962 were conducted under the administration of New Zealand, but, beginning with the 1966 census, Western Samoa has conducted its own censuses as an independent country.

All these censuses have asked questions on the number of children ever born to adult women and the number of children surviving. Responses to these questions can be used to assess levels and trends of infant and child mortality, using the Brass child mortality estimation procedure or any of its recent variants (Brass et al., 1968). One such variant estimates the infant mortality rate over time directly from census data on proportions dead of children ever born to women in different age groups. Figure 1 shows the resulting infant mortality estimates for Western Samoa based on the last five censuses.

If all five censuses had obtained complete and accurate reports from female respondents regarding the number of live births they had had and the number of their children still surviving, we would expect that all the infant mortality estimates from different censuses for the same date would be similar, though not exactly the same because of imperfections in the infant mortality estimation technique used here. For example, infant mortality estimates for the period around 1954–56 are available from five different censuses of Western Samoa. Instead of clustering around the same infant mortality figure, the census data give us very different infant mortality estimates for the mid-1950s: an infant mortality rate of 88.6 for the year 1954.3 (1956 census data); a rate of 81.1 for 1955.3 (1961 census data); a rate of 98.6 for 1955.2 (1966 census data); a rate of only 21.8 for 1954.0 (1971 census data); and a rate of 55.3 for 1956.2 (1976 census data).

The first three censuses are in fairly close agreement on estimates of infant mortality for the 1930s and 1940s. Thereafter they diverge, with the 1966 census producing the highest series of infant mortality estimates for the 1950s, and the 1956 and 1961 censuses agreeing on a somewhat lower range of estimates for that decade. All three of these censuses suggest that infant mortality declined during the 1950s in comparison with the 1940s. Initially, some confidence can be placed in the results from these three censuses, because they are in general agreement in their range of infant mortality estimates for each time period. Particularly, points that are clearly out of line should be rejected—for example, the estimate of an infant mortality rate of 149.9 for the year 1951.7 obtained from 1966 census data.

The 1976 census produced a low series of infant mortality estimates which do not overlap with the results of the three early censuses at all. This census suggests that Western Samoa's infant mortality rate was already rather low during the 1950s, declined slowly during the 1960s, and then leveled off at about 35 infant deaths per thousand live births during the 1970s. The results of this census, however, are suspect for the whole series, because the infant mortality estimates are clearly too low for the 1950s and early 1960s in comparison with the earlier censuses. We cannot trust the most recent infant mortality estimates from this census if the earlier ones are significantly too low.

Figure 1 Infant mortality estimates from child survivorship data: Western Samoa

SOURCE: These infant mortality estimates are based on the data in Table 2. From each census, the infant mortality estimate for the most recent date prior to the census is based on the proportion dead of children ever born to women age 20–24 at the time of the census. Each infant mortality estimate for earlier years is based on the reported proportion dead of children ever born to women of an older age group, such as women 35–39 or women 70–74 at census time. This infant mortality estimation technique is described in Fenney (in press).
The 1971 census data produced very low infant mortality estimates—rates in the range that only developed countries have been able to attain. It is certain that Western Samoa did not actually achieve an infant mortality rate as low as 14—20 per thousand live births during the 1960s. The lack of agreement between 1971 census results and those from the other censuses makes it obvious that something went wrong with responses to the questions in the 1971 and 1976 censuses on children ever born and children surviving.

What happened to produce such conflicting infant mortality estimates? From each census, these estimates are derived from the reported proportion dead of children ever born to women in each adult age group. The proportion dead of children ever born is in turn calculated from responses to the question on how many children a woman has ever borne and from other answers on the number of children surviving or number of children dead. Therefore, we must search for the sources of error either in the reported children ever born in different censuses, or in the reported children surviving, or both in combination.

Census data on fertility and child mortality

Table 1 lists the reported children ever born to women in five-year age groups from all five recent censuses. Reading horizontally, one sees the number of children (parity) that women in the same age group have reported at different dates. The diagonal lines trace the same cohort of women, who become five years older at each successive census. (A cohort is a group of people born in the same year or same group of years.) If a particular group of women reports at one census that they have borne fewer children than the number that the same cohort said they had already borne five years earlier, it is reasonable to suspect underreporting of children ever born at the more recent census. A very slight decline in the reported children ever born of a cohort of women over time might be accurate, if there is massive emigration of high-parity women or immigration of low-parity women, or if adult females with high parity are much more likely to die young than females of low parity. But any significant decline in the reported parity of a cohort of women is likely to be an error.

The 1966 census produced the highest numbers of children ever born reported in any census. Older women respondents in 1966 gave remarkably complete reports of the total number of live-born children they had had. All the cohorts who had completed their fertility—that is, women age 40—44 and older—reported having borne 7.3—7.8 children in their lives. This result indicates that fertility in Western Samoa had been high and constant for most of this century. Yet these same cohorts of women, even if they had completed their childbearing by 1956, reported lower parity in 1956 and 1961 than they reported in 1966. I conclude that reporting of children ever born was most accurate and complete for the 1966 census, and that fertility was just as high prior to the 1956 and 1961 censuses but that respondents of many ages slightly underreported their children ever born in those censuses. The 1971 and 1976 censuses had clear underreporting of children ever born by all the cohorts of women over age 44, because those same cohorts had already reported achieving significantly higher parity by 1966. I suspect that the younger cohorts of women in 1971 and 1976 also underreported parity. Therefore, the best reporting of children ever born seems to have been achieved in the 1966 census. Slight underreporting of parity in the other censuses may cause underestimation of infant mortality, because those children whose births are not reported are more likely to be those who died. If children surviving are more fully reported than children ever born, the proportions dead are underestimated and, consequently, so is the infant mortality rate.

Figure 2 shows the census questions on children ever born and children surviving from all five censuses. The wording of the questions has changed from census to census. The 1966 wording of the parity question—"total children ever born"—elicited the most complete fertility response from women of all ages. Other phraseology, such as "number of children born alive," "total number of children born," or "total number of children living/dead," got slightly less complete fertility responses in the Western Samoan context.

Table 2 shows the percentage dead of children ever born to women respondents, as calculated from children ever born and children surviving questions on all five censuses. Again reading horizontally, one can see the child mortality experience of the same age group of women at each successive five-year date. If the infant mortality situation has been steadily improving over time, we would expect that women in a particular age group would report lower proportions of their children dead at each successive census. Indeed, for some age groups and some censuses, this does happen. For instance, women in the 30—34 age groups reported slowly declining proportions of their children dead from 1966 to 1961 to 1966, as expected.

The diagonal lines in Table 2 trace the reported experience of each cohort of women. We would expect that women who had completed their childbearing would report slightly greater proportions of their children dead at each successive five-year census. A significant decline in the reported proportions dead is in error. For example, women who were 40—44 in 1956 had already lost 20.7 percent of their children. By 1961, they reported that 21.0 percent of their children had died, and by 1966, 24.5 percent. But in 1971 this same cohort of women reported that only 6.2 percent of their children had died, and in 1976 that only 14.3 percent were dead. If 24.5 percent of these women's children had already died by 1966, how can it be that only 14.3
Table 2 Percentage of children ever born to women respondents: censuses of Western Samoa, 1956–76

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15–19</td>
<td>10.64</td>
<td>7.08</td>
<td>8.41</td>
<td>1.17</td>
<td>3.78</td>
</tr>
<tr>
<td>20–24</td>
<td>10.84</td>
<td>8.90</td>
<td>10.58</td>
<td>2.41</td>
<td>4.14</td>
</tr>
<tr>
<td>25–29</td>
<td>13.29</td>
<td>9.87</td>
<td>9.97</td>
<td>2.18</td>
<td>4.02</td>
</tr>
<tr>
<td>30–34</td>
<td>15.75</td>
<td>12.57</td>
<td>11.09</td>
<td>2.17</td>
<td>5.43</td>
</tr>
<tr>
<td>35–39</td>
<td>17.50</td>
<td>14.80</td>
<td>13.97</td>
<td>2.37</td>
<td>6.60</td>
</tr>
<tr>
<td>40–44</td>
<td>20.72</td>
<td>16.77</td>
<td>17.51</td>
<td>3.32</td>
<td>8.07</td>
</tr>
<tr>
<td>45–49</td>
<td>23.16</td>
<td>21.02</td>
<td>27.53</td>
<td>3.66</td>
<td>8.58</td>
</tr>
<tr>
<td>50–54</td>
<td>25.76</td>
<td>21.80</td>
<td>24.51</td>
<td>5.06</td>
<td>11.19</td>
</tr>
<tr>
<td>55–59</td>
<td>28.43</td>
<td>25.52</td>
<td>26.27</td>
<td>6.24</td>
<td>13.47</td>
</tr>
<tr>
<td>60–64</td>
<td>34.48</td>
<td>30.37</td>
<td>28.54</td>
<td>5.73</td>
<td>14.27</td>
</tr>
<tr>
<td>65–69</td>
<td>37.27</td>
<td>31.74</td>
<td>32.20</td>
<td>5.53</td>
<td>17.31</td>
</tr>
<tr>
<td>70–74</td>
<td>41.81</td>
<td>37.67</td>
<td>39.92</td>
<td>7.08</td>
<td>18.30</td>
</tr>
<tr>
<td>75 and over</td>
<td>47.64</td>
<td>41.97</td>
<td>42.57</td>
<td>8.28</td>
<td>21.03</td>
</tr>
</tbody>
</table>


percent were dead by 1976 and 6.2 percent by 1971? The systematic tracing of the reports of each cohort of women suggests that the reported proportions of children ever born are systematically underreported for the 1971 and 1976 censuses. It is clear from such analysis that no infant mortality estimates derived from 1971 census data on proportion dead of children ever born can be believed. The 1976 census also produces suspiciously low estimates of the proportions of children dead for each age group of women, and therefore little confidence can be placed in the infant mortality estimates derived therefrom.

Changes in census questions

What happened? A look at Figure 2 shows how the wording of census questions on children surviving has changed over time. The first three censuses asked simple questions—"total number of children living/dead," "total number of children still living," and "total number still living"—which brought fairly accurate responses. But the 1971 and 1976 censuses introduced more complicated questions which seem to have confused the respondents and resulted in an overestimation of children surviving.

In the 1971 census, enumerators first asked women "number of children born alive," then "number of children still alive and living with mother," and finally "number of children still alive and not living with mother." This distinction was made because in Samoan society the phenomenon of adoption is very widespread. Many children move in with adoptive parents while maintaining ties with their natural parents. The intent of the more complicated 1971 questions on children still living was to get more complete reporting of each woman's living children, even if they were no longer living with their natural mother. Instead, women were apparently confused by the two questions on surviving children, with the result that adopted children were evidently reported by both their natural mothers and their

1956 census

<table>
<thead>
<tr>
<th>FOR FEMALES 15 YEARS OR OVER</th>
<th>Age when first child born</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total No. of Children Living</td>
<td>Age when first child born</td>
</tr>
<tr>
<td>Total No. of Children Dead</td>
<td>Age when first child born</td>
</tr>
<tr>
<td>Total No. of Children Ever Born</td>
<td>Age when first child born</td>
</tr>
</tbody>
</table>

1961 census

<table>
<thead>
<tr>
<th>a) Total Number of Children Born</th>
<th>b) Total Number of Children Ever Born</th>
<th>c) Age when First Child Born</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Children Living</td>
<td>Total Number of Children Dead</td>
<td>Total Number of Children Ever Born</td>
</tr>
</tbody>
</table>

1966 census

<table>
<thead>
<tr>
<th>a) Total Number of Children Ever Born</th>
<th>b) Total Number of Children Living</th>
<th>c) Age when First Child Born</th>
</tr>
</thead>
</table>

1971 census

<table>
<thead>
<tr>
<th>a) Number of children ever born alive</th>
<th>b) Number of children ever born alive during the twelve months preceding the census</th>
</tr>
</thead>
<tbody>
<tr>
<td>total number of children born alive</td>
<td>total number of children born alive during the twelve months preceding the census</td>
</tr>
</tbody>
</table>

1976 census

<table>
<thead>
<tr>
<th>a) Number of children born alive (inclusive adopted children)</th>
<th>b) Number of children still living (inclusive adopted children)</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of children born alive</td>
<td>number of children still living</td>
</tr>
</tbody>
</table>

Figure 2 Census questions on fertility and child mortality: Western Samoa, 1956–76
adoptive mothers. The net result of the three questions on children born and children still alive was that the total number of children ever born may have been slightly underestimated while the total number of children surviving was seriously overestimated. The difference between these numbers, which ought to equal the number of children dead, was therefore far too small. A census pretest of these new questions might have resolved some of the difficulties. If the confusing nature of the questions had been detected before the 1971 census was taken, the new questions could have been discarded and the previous simpler ones retained.

The 1976 census included a different series of complex questions on children ever born, children surviving, and adopted children. Female respondents were first asked "number of adopted children living." This question is ambiguous. Does it mean those children from other natural parents whom this respondent has adopted, or does it mean those children born of the respondent who have been adopted by other families? The next question asked for "number of children born alive" qualified by the phrase "excluded adopted children." This question could produce underreporting of the respondent's total live births, because she could easily assume that the question meant "Tell me the number of children you have borne, excluding those who have since been adopted by other families." The next question, "number of children still living (excluded adopted children)," somehow elicited an overresponse, possibly because respondents were not clear which adopted children to exclude.

In 1971 and 1976, two different attempts were made to ask about adopted children in conjunction with Western Samoan census questions on children ever born and children surviving. These attempts were not successful in eliciting accurate answers concerning fertility and child survival. The earlier censuses, those of 1956, 1961, and 1966, asked simpler and more standard questions on children ever born and children surviving, and they seem to have produced reasonable responses. It appears that the women of Western Samoa know what is meant by "total children ever born" and "total number still living," the straightforward inquiries from the 1966 census. These are the answers that are needed for infant and child mortality analysis. Western Samoan respondents can clearly distinguish between their own natural-born children and adopted children if they are asked the right questions. For the coming 1981 census of Western Samoa, census planners might consider returning to the clear wording of the 1966 census questions on parity and child survival, with the addition of the very useful fertility question on births during the preceding twelve months. A pretest should be conducted to determine whether the questions are getting accurate responses.

Conclusion

Several useful lessons can be learned from the experience with census fertility and child mortality questions in Western Samoa. First, the accuracy of responses to census questions can be extremely sensitive to the exact wording and the order of the questions. The design of a census interview schedule is a crucial determinant of the final usefulness of census data.

Second, if a particular census question is eliciting a complete and reasonable response for one census, the wording of the question should not be changed for the next census without an adequate reason. Change for its own sake is very risky when designing a census. Besides, continuity in census questions is useful in itself. Responses to the same questions at several points in time give us comparable longitudinal data. Changing the way a question is asked makes it less likely that the responses are comparable, so that trends over time cannot be detected or measured with confidence. It is ironic that the reason for changing census questions is usually an anticipated improvement in completeness and accuracy of responses, and yet the change can unexpectedly cause deterioration in the quality of the collected data.

Third, when a change in the wording or order of census questions is contemplated, it is wise to pretest at least this portion of the questionnaire, or the whole questionnaire if possible. Pretesting in this case does not mean simply a trial run of census field operations; it should be a trial run of the whole sequence of census operations, including final analysis of the collected data. Such pretesting should begin early enough to allow time for the adjustment of faulty census questions and further pretesting on the adjusted questions.

Finally, small countries as well as large ones should try to develop enough expertise and experience within their own governments to design and carry out their own censuses. In the case of Western Samoa, some lack of continuity and variation in quality from census to census has been due partly to the strong role played by a changing group of foreign advisers in the design and conduct of the censuses. Now, Western Samoa has in its Department of Statistics a number of experts trained in demography, statistical techniques, and the skills of census design and analysis. It is important that developing countries such as Western Samoa first train experts for census work and then honestly assess the degree of expertise available in their own country. They should rely on outside advisers only to the extent that they are really needed. In this way, national census organizations should gradually be able to take over the conduct of their own censuses and improve the quality of their censuses over time as they gain experience.

REFERENCES


CENSUS SAMPLING WORKING GROUP
by Leslie Kish and Jay-Soo Park

A Census Sampling Working Group held 24 January–16 February 1979 at the East-West Population Institute brought together 18 participants from 16 countries for discussions of sampling plans connected with the 1980 round of censuses of population and housing. The participants were statisticians, primarily from national, central statistical offices, where they are involved in sampling operations for the national censuses.

Specific goals of the Working Group were:
- to learn about the different uses of sampling in connection with censuses;
- to explore the advantages and disadvantages of alternative sample designs; and
- to outline proposed sampling designs for each of the participating countries.

Thirty-two lecture sessions and lively discussions were supplemented by individual study of relevant literature. Participants also prepared and presented country reports and sampling design outlines.

Uses of samples in censuses

Early in the Working Group the several possible uses of sampling methods in connection with censuses were classified and described. The different types of samples connected with censuses are outlined below.

1. Sample enumerations as substitutes for complete censuses (an article on this topic begins on page 1 of this Census Forum)
   a. Provide richer, more diverse, detailed, deeper data
   b. Reduce costs of collection and tabulation
   c. Lessen aggregate social burden on respondents

2. Samples added to complete censuses to evaluate and to improve them
   a. Content evaluation of the quality of data
   b. Coverage checks; dual coverage studies
   c. Pilot studies of methods and questions prior to census
   d. Quality control of individual enumerators, coders, processors

3. Samples from census records, microfilms, tapes
   a. Permit early (preliminary) tabulation and release of data
   b. Reduce cost of complex analyses of relations
   c. Provide public use of tapes for further, deeper analyses

4. Census as auxiliary data for samples
   a. Data for selections: measures of size, stratiﬁers, maps of enumeration districts
   b. Data for improved estimation with ratios, regressions
   c. Samples added to censuses to serve as bases for continuing surveys

5. Joint uses of several sources
   a. Synthetic estimates for local areas and small domains
   b. Rotating (moving) monthly samples of 1/120 (or weekly of 1/520)

Alternative sample designs

Designs for sample enumerations and for evaluation of content and of coverage are important issues in census sampling. The choice among possible alternatives should fit the specific needs and the resources of each country. Three factors most strongly influence the choice among alternative sampling plans: the nature and training of the enumerators and the supervisors for both the basic, complete census and for the special samples; the nature, the mapping, and the sizes of enumeration areas (EAs); and the type and timing of the listing procedure for households and persons in the EAs.

Three alternative designs appear most practical, particularly for sample enumerations. The first entails a sample of 1, 5, or 20 percent of EAs, including all households and persons within them. The disadvantages of this design are the increased variiances and complex computations caused by design effects.

The second design uses a sample of households selected from a complete listing of all households in all EAs. The disadvantages of this design are double or triple entries into sample households and the care needed to avoid large selection biases.

Finally, there is the sample drawn by listing, selecting, and enumerating the basic census and the sample households simultaneously with one entry. The disadvantage of this design is the possibility of serious selection biases.

The group also discussed using ratio estimation to adjust the sample enumeration to the complete census data, thereby reducing the bias of selected households in the second and third methods and the sampling errors of the first.
Sampling problems and expected uses

Two sessions of the Working Group were devoted to questions, answers, and discussions of actual, practical problems of survey sampling experienced by the participants. Questions about survey sampling not connected with censuses were also raised.

Participants outlined existing plans for sampling in connection with the 1980 round of censuses and reported on proposed uses of sampling in their own countries. The potential use of eight major sampling techniques appears quite widespread.

CENSUS SAMPLING WORKING GROUP
24 January-16 February 1979

COORDINATORS: Dr. Leslie Kish, Professor and Research Scientist, Institute for Social Research, University of Michigan; Dr. Jay Soo Park, Advisor on Sampling and Agricultural Statistics, Central Bureau of Statistics, Indonesia; and Dr. Griffith Feeney, Research Associate, EWPI

PARTICIPANTS

Afghanistan: Mr. Mohammad N. Naser, General Director of Design and Analysis Department, Central Statistics Office

Bangladesh: Mr. Siddiqui Rahman, Deputy Director, Population Census Wing, Bangladesh Bureau of Statistics; and Dr. Kazi Saleh Ahmed, In-charge, Data Processing Division, National Foundation for Research on Human Resource Development

Hong Kong: Mr. Kai Cheong Chan, Senior Evaluation Officer, Family Planning Association

India: Mr. Ailt Kumar Biswas, Assistant Director, Census Operations (Technical), Office of the Registrar General; and Mr. Akhlaq Ahmad, Assistant Director, Census Operations (Technical), Office of the Director of Census Operations

Indonesia: Mr. Hardjoperwono Kartono, Deputy Director of Bureau II (Basic Statistics), Census Preparation Committee, Central Bureau of Statistics

Republic of Korea: Mr. Moon Bang Kye, Chief, Sampling Section, Survey Management Division, National Bureau of Statistics

Malaysia: Mrs. Sutharavanathana Paranjothy, Statistician, Department of Statistics

Mexico: Mrs. Estrella Luevano-Moya, Researcher, General Direction of Design and Implementation, National Systems of Information

Nepal: Mr. Radha Krishna Ghartiy Chhetry, Statistician (Programmer), Central Bureau of Statistics

Philippines: Mr. Nelia Marquez, Chief Census Statistical Coordinator, National Census and Households Surveys Department, National Census and Statistics Office

Republic of Singapore: Mr. Cheong-Shek Wong, Acting Deputy Chief Statistician, Department of Statistics

Sri Lanka: Mr. Eugene Joshua Chandrapaj, Assistant Director, Department of Census and Statistics

Taiwan: Mr. Tun-Yih Lu, Deputy Director, Bureau of Statistics

Thailand: Mrs. Maneerat Pilpat, Statistician, Statistical Techniques Division, National Statistical Office

United States: Mr. Charles Jones, Chief, Statistical Methods, U.S. Bureau of the Census; and Dr. Michael J. Levin, Fellow, EWPI

Western Samoa: Mr. Salale Salale, Statistician, Department of Statistics

This activity was part of the EWPI Project on Development and Application of Techniques for Censuses, Surveys, and Vital Registration.

CENSUS EDITING WORKSHOP

by Michael J. Levin

The Census Editing Workshop, held at the East-West Population Institute 12-16 March 1979, brought together specialists in data processing from 19 Asian and Pacific countries to discuss common problems in editing census data. In conjunction with the Workshop, a six-week Workshop in February and March brought programmers to the Institute to learn the CONCOR (CONSistency and COREction) editing package. The Workshop provided a forum for subject-matter personnel and programmers to discuss the role of editing in census data processing, to become familiar with the latest software packaged programs for editing survey and census data, and to consider the most appropriate editing procedures for the 1980 round of censuses.

Approaches to the practicality of field and office manual editing were found to vary according to population size, literacy, and labor costs influencing the completeness of the data. Many of these same factors also determine each country's proportions of manual and computer editing. Usually some manual editing is necessary to establish criteria for computer editing; computer editing can be done in a number of established programming languages, or a computer editing software package can be used.

Editing packages

Most of the Workshop discussions centered on computer editing packages, with the proposed COBOL CONCOR editing package used as the primary example. One of the strengths of CONCOR is its imputation feature: either a previously determined, stored value may be assigned when an illegal value is encountered in the data, or a "hot-deck" array may be used to assign a value based on previous values of the data and dependent on related variables.

The participants noted some of the problems with the present CONCOR version, including the extensive error listings, which are somewhat impractical for large data sets. CONCOR is presently designed for use on an IBM 370 system only, but versions for other systems are expected soon.

Other computer packages for editing census data were also presented. UNEDIT was developed by the United Nations for projects needing an easy-to-use, portable package to edit small data sets on computers with limited capacity. It can be run on mini-computers equipped with the RPG-II compiler. The translation of UNEDIT to a COBOL version is in progress.

An advantage of UNEDIT is the ease with which users can learn the system capabilities and specification requirements. Users code edit specifications and input data specifications on pre-formatted forms. UNEDIT can accept different types of input, allowing population and housing information to be edited at the same time. The disadvantages of UNEDIT include its use of RPG-II language, and its not permitting imputation through "hot-deck" allocation.

The CANEDIT system, also known as GESH (Generalized Edit System using Hot Deck Approach), is large, sophisticated, and easy to use. The advantage of this software program is its masterful edit performance. The imputations performed by CANEDIT check for passage of all the edit tests to be done on the variable being imputed. An analysis of the variables in error is made to determine the fewest number of variables that must be imputed to resolve all the errors. The major disadvantage is that the package is not portable since it is linked to Canada's data-based system. It requires a large computer capacity and is very costly.
Imputation

Imputation, or allocation, is a method used to replace unknown or inconsistent information with a "best guess" alternative. Although the intent is to improve the overall quality of the data, the procedure actually is used to eliminate unknowns so that results in comparable tables are compatible.

(continued on page 13)

CENSUS EDITING WORKSHOP
12–16 March 1979

COORDINATORS: Dr. Michael Levin, Research Fellow, EWPI; and Mrs. Amuri Wangee, Director, Population Survey Division, National Statistical Office, Thailand

PARTICIPANTS

Afghanistan: Mr. Rafiqdoin Malikzay, Head, Data Processing, Central Statistics Office

Bangladesh: Mr. Abdur Baten, Senior Systems Analyst, Bangladesh Bureau of Statistics; and Mr. Hamidur Rabanni Bhuiyan, Programmer, Data Processing Unit, Bangladesh Bureau of Statistics*;

Guam: Mrs. Yolanda Guarin, Computer System Analyst, Bureau of Planning*;

Hong Kong: Mr. Pun-ki Yu, Statistician, Census and Statistics Department;

India: Mr. R. L. Puri, Assistant Director, Office of the Registrar General; and Mr. Himakar, Assistant Director, Office of the Registrar General*;

Indonesia: Mrs. Anuwina Hadipramana, Computer Programmer, Central Bureau of Statistics*;

Japan: Mr. Katsuo Mizuguchi, Senior Programmer, Bureau of Statistics*;

Republic of Korea: Mr. Kyeong Joong Kim, Director, Standard Division, National Bureau of Statistics; and Mr. Kil Soo Kim, Senior Programmer, National Bureau of Statistics*;

Malaysia: Mr. Seng Wah Ong, Acting Deputy Chief Statistician, Department of Statistics; and Mr. Mohamad Fatoni bin Ali, Chief Programmer, Department of Statistics*;

Nepal: Mr. Bhaktal Basadur Shrestha, Data Conversion Officer, National Computer Center; and Mr. Hari Gopal Shrestha, Programmer, National Computer Center*;

New Zealand: Mr. Michael Anthony Moore, Executive Officer, Population Statistics Section, Department of Statistics; and Mr. David Jamiesson, Senior Programmer, EDP Division, Department of Statistics*;

Pakistan: Mr. Syed Afzal Ahmad, Director, Statistics Division, Government of Pakistan, Islamabad; and Mr. Mohammad Ayub, Chief, Operations, Statistics Division, Government of Pakistan, Karachi*;

Philippines: Ms. Elizabeth Go, Census Statistics Coordinator, National Census and Statistics Office; and Mr. Agustin P. Perez, Census Data Processing Chief, National Census and Statistics Office*;

Republic of Singapore: Mr. Min Chin Ling, Programmer/System Analyst, Ministry of Finance, Computer Services Department; and Mr. Hoon Kee Ang, Principal Programmer, Computer Services Department*;

Sri Lanka: Ms. Sivagnaneswary Sivahubramaniam, Statistical Officer, Department of Census and Statistics; and Mr. Nakhalage Don Gunasiri, Programmer, Department of Census and Statistics*;

Taiwan: Mr. Shu-Jen Chu, Section Chief of Application and Programming, EDP Center, Director-General of Budget, Accounting and Statistics*;

Thailand: Mr. Angsumal Sunalal, Senior Program Analyst, National Statistical Office*;

Trust Territory of the Pacific Islands: Mr. Gabriel M. Gamow, Statistics Specialist, Office of Planning and Statistics*;

United Nations: Mr. Toshio Shigematsu, Senior Statistician, United Nations Demographic and Social Statistics Branch;

United States: Mr. Howard Brunson, Consultant*; Mr. William O’Leary, 1980 World Census Staff, International Statistical Programs Center, United States Bureau of the Census; Dr. H. Paul Greenough, Agency for International Development, U.S. Department of State; Mr. Steven Snyderman, International Statistical Programs Center, United States Bureau of the Census*; Ms. Carol E. Allison, Computer Programmer, Population Laboratories, School of Public Health, University of North Carolina

This activity was part of the EWPI project on Development and Application of Techniques for Censuses, Surveys, and Vital Registration, Dr. Griffith Feeny, Project Coordinator.

* Also participated in Census Editing Working Group, 14 February–30 March 1979.

Participants in the Census Editing Workshop were (first row, left to right) Nakandilage Don Gunasiri, Gabriel Gamow, Agustin Perez, Shu-Jen Chu, Hari Gopal Shrestha, Bhakta B. Shrestha, Michael Levin, Hamidur R. Bhuiyan, and Toshih Shigematsu; (second row) R.L. Puri, David Jamiesson, Min Chin Ling, Angsumal Sunalal, Abdul Baten, Mohammad Ayub, Syed A. Ahmad, Anuwina Hadipramana, Sivaganeswary Sivahubramaniam, Yolanda Guarin, Anuwina Wangree, Carol Allison, and Elizabeth Go; (third row) Michael Moore, Pun-ki Yu, Paul Greenough, Mohammad F. Bin Ali, William O’Leary, Rafedidin Malikzay, Seng Wah Ong, Hoon Kee Ang, Himakar, Katsuo Mizuguchi, Steven Snyderman, Kil Soo Kim, EWPI Program Officer Janet Mason, and Kyeong J. Kim.
Leslie Kish on sampling (continued from page 2)

others (births, deaths, income). Reference periods are fixed separately for each variable; the collection period may vary between teams to permit efficient collection.

The reporting period is something else again, and there may be several of these periods from different summations of the same data. Let us focus on the reference periods and consider weekly samples. They can be added into yearly or quarterly or monthly reports—provided the months are made artificial integral months of four or five weeks.

Even more flexibility may be had by increasing the reference periods. Rotating weekly samples of 1:520 can yield data for four weeks each (or more) and increase the data base, especially for items that are rare and not positively correlated (births and deaths in the family).

Sampling the time dimension

Here we come to an aspect where traditional wisdom and statistical logic meet head-on. In concepts of sampling, the time dimension has been neglected, whereas sampling the spatial spread of populations has made great strides. Yet populations vary from year to year and from week to week, just as they vary among regions or among counties (Kish, 1965: section 12.5D).

Some of the variation is cyclical: seasonal, like agricultural activity, vacations, and some industries; or daily during the week, like weekends; or diurnal, like hours of work. But many variations are irregular or haphazard and cannot be foreseen or accounted for in models or procedures.

Probability methods of sampling are designed to counter and cover uncertainties in spatial variation, but the temporal sources of variation are not treated so well. Arbitrary selection of “typical” periods is most common: April 1 for the de jure day of reference for the U.S. Census; a “representative” week for monthly or quarterly surveys of the labor force; a “convenient” year for decennial censuses. This strategy may work somehow against “known” (more or less) and regular sources of variation, but not against haphazard sources or events. They are ignored or treated with explanations of the ad hoc kind—which one would hardly dare to use for spatial irregularities.

In some situations entire periods are covered in the reference period—annual income, for example, or annual crops for agricultural surveys. This approach attempts to cover the variations (seasonal, weekly, diurnal, haphazard) during the period, though this wide span may overreach the method’s capabilities. In other words, the period may be too long.

Even year-long coverage of the reference period would require annual surveys for complete coverage of temporal variation. And for many variables monthly, weekly, or even daily coverage is needed. Since frequent censuses are impossible, we must use either arbitrary, judgmental selection of “typical” or “representative” reference periods or probability selection of reference periods from the entire reporting period.

To illustrate, consider again monthly samples of 1:120, each a national sample. They are balanced to provide 12:120 = 1:10 samples when summed over each year and also to provide quarterly samples of 3:120 = 1:40. These yearly and quarterly samples represent summations for reporting purposes of reference periods tied to collection periods.

For national samples we thus have monthly, quarterly, and yearly samples. But for small domains and especially for local areas, the monthly and even the quarterly samples will probably be too small and not well balanced. For very small domains even the annual samples of 1:10 may be insufficient, and summations over five or ten years may be needed. Such averaging over several years is unusual if annual variations are possible, but to statisticians this averaging should be more acceptable than the current practice of using obsolete data.

Essentially the temporal and spatial dimensions define a two-dimensional matrix of variation. The individual cells (monthly data for local area) have small and unbalanced samples. But both the spatial and temporal margins represent samples balanced against the other dimension: that is, monthly national samples, and yearly (or decennial) samples for small areas. An early (though surely not the earliest) application of this design was for the measurement of traffic flows (Kish, 1961).

Continuing versus sporadic operations

The design described above can be modified to fit diverse practical collection procedures. For continuous field collection of simple data, from any responsible adult in households, the best collection period may be weeks (with extra time and procedures for absentees) rather than months (or days). Connected with weekly collection, we may have weeks as reference periods for some data (though days or months for others). A continuing sample of 1:520 would cover the country in ten years.

Such a continuing operation would probably depend on full-time professional data collectors (enumerators and interviewers), although a rotating staff of part-time workers is conceivable. This scheme would be in stark contrast to the one-time effort now used for decennial censuses. Census enumerators may be specially hired, or designated from a class (e.g., teachers) with or without pay, or “volunteered” by some agency; in any case their effort is ad hoc and sporadic and their training is brief. Even if most enumerators were conscientious and capable, and even if the enumeration were relatively simple, the problem remains that 100,000 enumerators need 10,000 supervisors. And finding qualified supervisors is probably more difficult than finding enumerators.

Perhaps impartial, scientific, complete inquiries would replace decennial censuses with large rotating samples in most countries. But tradition, public acceptance, and political and legal acceptance continue to uphold the role of complete decennial censuses. Before quinquennial or annual censuses are adopted, however, the idea of devoting such expenditures instead to rotating large samples deserves attention.

New technologies: survey sampling and computers

Demands for more timely, detailed, and relevant data are being propelled by the needs of modern societies and pulled by the advent of new technologies. Data that are two to 12 years old cannot satisfy the needs of administrators and planners of societies that are ever more mobile and ever more interconnected.

It would be wrong to call census taking an old technique; though it has old roots, it is still new in many countries, and in many more places it still needs to be introduced. And it would be worse to condemn it for being old. Yet it may be true that survey sampling, because it is a newer technique, has an uphill battle for recognition against the accepted methods of census taking.
Though sampling has older roots, most of its applications have been made within the past generation, within the last three decades. During that time it has been introduced into most countries into many new fields and with new techniques. The World Fertility Survey, and its successes with high quality and standardized results over the globe, symbolizes the success of survey sampling.

Survey organizations now exist in many (probably most) countries that can take on a wide variety of tasks on relatively short notice. Many of them conduct periodic surveys, and most have electronic computer facilities. It is possible for these days to have computer programs turn out the results of periodic surveys within a week or a month after their completion. Add to the spread of survey sampling and of modern computing a third new technology: new methods of estimation for local areas and small domains (Purcell and Kish, 1979, 1980).

Hence the difference in obsolescence between periodic surveys and censuses is striking: seven years versus one-tenth of a year! Meanwhile, the cost of obsolescence is increasing with greater mobility, dynamic activity, and changes in modern societies and with greater demands for even newer and richer data. The equilibrium seems to me clearly tilting toward current data and toward sampling.

It is unlikely that decennial censuses will soon be discontinued in favor of rotating samples (minicensuses), and it would be inadvisable to do so. But it is possible and advisable to begin those samples soon in order to provide current and detailed estimates for a great variety of relevant data, because estimates better than those now available are much needed. Financing for these new surveys can come from the budgets of quinquennial censuses, now being planned in several countries, and from the budgets of other rotating samples, either planned or actually operating. If the expense for the entire country is judged too great, the rotating samples may be first tried in some domains. These could be provinces and metropolitan areas in greatest need of current data for local areas.

Suppose that rotating samples will be operating for two decades in, let us say, ten countries. Those experiences would provide 20 occasions for comparing estimates from decennial censuses with those from competing rotating samples. The comparisons would be difficult but perhaps feasible, at least for some variables. In the postcensal years the rotating samples could greatly outperform those based on the census; but they may also do better than censuses in the decennial years if the censuses continue to be burdened with all the problems they now have. Unfortunately, many censuses still have grave problems of execution and of acceptance. In all fairness, censuses should become strengthened and improved during these next decades because they still continue their important traditional functions. If the rotating samples will seem conclusively better than the decennial censuses, then the grave question of discarding those traditional bulwarks could at last be raised. Their departure, like that of other great superannuated inventions, would be regarded with regret and nostalgia. And it should not be speeded prematurely.

REFERENCES


Levin: Census Editing Workshop (continued from page 11)

Four methods of imputing invalid or unknown data were discussed at the Workshop: manual correction, “cold-deck,” “hot-deck,” and the Fellegi-Holt method. Manual imputation may be made by selecting appropriate values when invalid data are encountered. In the “cold-deck” approach, a set of values is established, and the imputation substitutes selected values for unknown values in the input data. The “hot-deck” method replaces unknown values with data obtained from other similar respondents, questionnaires, or records: cells in arrays of information are initialized with “cold-deck” values which are then selected for use when an unknown is encountered or changed on the basis of known information. The Fellegi-Holt method selects imputed values from previous valid information given for respondents, questionnaires, or records when unknowns are encountered.

Because of time and cost restraints of manual editing for censuses, most of the countries represented at the Workshop have or plan to implement some form of computer imputation. The decision on how much imputation will be undertaken—all invalid values or only selected values—will depend largely on the editing specifications of the individual countries, based on time, human and technical resources, and the financial basis of the country’s census operations. In general, the quality of the collected data will determine the amount of imputation needed.

Imputation and the logic of the editing steps were discussed at length, using POPSTAN as an example; POPSTAN is a mythic country created by the International Statistical Programs Center, U.S. Bureau of the Census, as a training document on organization, planning, and operation of a population and housing census. Several basic premises for efficient editing became apparent from the discussions:

- Subject-matter people and programmers should work together in developing editing specifications.
- Sequencing of edits should be carefully considered so that the most efficient flow of editing can be maintained.
- Range checks should be done at the beginning of the editing process to ensure validity in subsequent values.
- In setting editing specifications, simple checks should be done first. The more difficult checks (including interrecord checks) should be done only when necessary.

These rules were followed as pairs of subject-matter personnel and programmers worked together to test simple range and consistency checks using CONCOR.
Censuses in China

Following is a translation of the letter reproduced below from the State Statistical Bureau, Beijing:

To the Managing Editor of the Asian and Pacific Census Forum: The copy of the Asian and Pacific Census Forum, which you sent to Mr. Chen Xian, Director of China’s State Statistical Bureau, has been received. Thank you. There are two points relating to China’s censuses on page 3 of Volume 5, Number 4, which need correction:

1. The 1953 population census was not the only census. We conducted another census in 1964.
2. The next population census is tentatively scheduled for 1980 or 1981, depending on the progress of major decisions in the preparatory work.

I hope you have an opportunity to make these corrections.

(Seal of the State Statistical Bureau, People’s Republic of China)

国家统计局

Identifying special groups for follow-up surveys

I have a question for the readers’ section of the Forum. We at the Australian Bureau of Statistics are interested in the use of a census to produce a name/address list of particular populations, for example, single-parent families or qualified persons not using their qualifications. The lists would be used for detailed follow-up surveys. I am interested in any documentation readers may have on such things as approach used to selected groups, publicity, success rate in identification of groups, and results of surveys.

Brian Doyle, Director
Development, Evaluation, and User Services Section
Population Census Branch
Australian Bureau of Statistics

Tracing respondents without telephones

As I read the article by Deborah Freedman et al. in the latest Census Forum (“Maintaining Response Rates in Longitudinal Studies: The Care and Nurture of Respondents,” May 1979), my constantly recurring thought was that in nearly all of the Asian countries for which the Forum is intended, the use of the telephone to keep in contact with respondents is not warranted since most people don’t have phones. Thus, tracing people with other techniques, such as contact with relatives and friends, would be much more expensive since interviewers would have to be restricted to making the contacts directly in the field.

Perhaps readers have suggestions for other approaches to maintaining contact in developing countries.

Louise Williams
Demography Division, Office of Population
U.S. Agency for International Development

Age redistribution: another method

I have noted with interest in the February 1979 copy of the Census Forum the article by Griffith Feeney on a technique for correcting age distribution for heaping on multiples of five. My particular interest in this article arises from the fact that, in common with a multitude of actuaries, I have been using a method derived by the British actuary, George King, some 80 or so years ago, which, as far as I can see, seems to duplicate the same function as you require in respect of the census data you have quoted.

George King’s method also begins with the assumption of heaping in multiples of five, but then goes on to derive a spreading technique based on the further assumption of third degree polynomial fitting over the five-year points, these five-year points being regarded as pivotal values. This process of osculatory interpolation is in fact similar to, although rather more primitive than, the cubic spline fitting which is so popular nowadays. Once the corrected pivotal values are obtained, the interpolated values at single ages can then be derived by one of the many interpolation formulae available.

A note on King’s method is enclosed herewith, and I should be grateful to receive comments on my submission, particularly if I have misread any part of the intention of Feeney’s note.

Arnold H. Gould
Government Statistician
Tonga

Correction for “Heaping” of Census Numbers at Five-Year Age Points

Let \( x, x+5, x+10, \ldots \) be the five-year age points at which heaping is evident (the “pivotal” ages). Let \( dx \) be the true (continued on page 16)
There are so many new publications to tell readers about this time that only the editor's dicta will keep this column within reasonable bounds! I had planned to comment on the new U.S. Bureau of the Census report, *World Population 1977*, in the last issue of the Forum, but space did not permit it. Since I regard this volume as an important reference work, let me say a few words about it now.

*World Population 1977* was prepared by the Population Division of the Bureau of the Census, under the general direction of Samuel Baum, Assistant Chief for International Demographic Statistics. Earlier versions of the book presented demographic statistics for 1973 and 1975. The new edition provides comprehensive, annotated data for 1977 for all the countries and regions of the world. The completeness and accuracy of the demographic statistics vary from country to country as the report itself indicates, and the material is already somewhat out of date, as new data are arriving at the International Demographic Data Center daily. Still, it remains one of the most complete sets of demographic data available in one volume, and it is more current than the U.N.'s *Demographic Yearbook*.

The report contains selected demographic statistics for the 200 countries of the world that have a population of at least 5,000 people. Some of the demographic figures have been derived by applying techniques for making estimates from incomplete data. The "benchmark data" in the monograph represent the most recent national data upon which reliable demographic estimates can be based, in most cases the latest population census. Information on vital rates varies considerably from region to region and is generally harder to get than total population counts. Because of this variability, and the resulting inaccuracies, the benchmark data cannot supply a "worldwide demographic picture at the same point in time for any of the measures presented."

The report does, however, include estimates of such demographic data as the total population projected to midyear 1977 and the birth rate, death rate, and growth rate to 1976. These estimates will enable users of *World Population 1977* to have recent figures for all countries for cross-national comparisons and will allow them to aggregate demographic data on a worldwide basis.

The source notes included at the end of each section are of great importance for the interpretation of the data on each country. They show the sources of information as well as the demographic techniques used to calculate rates. Wherever the data are not reliable enough to supply a single estimate, a range of figures is given. Maps for each region precede the text, and a set of illustrative maps on world population, fertility, mortality, and year of latest census is included in the pocket at the back of the volume.

This report will be used by government agencies, foreign missions, universities, private organizations, and the general public. Anyone involved in policy making, research, speech writing, marketing, or compiling country reports can turn to *World Population 1977* for demographic statistics. It may be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, for U.S.$6.25. The stock number is S/N 003-034-01583-8, and the Library of Congress Card Number is 78-602. A remittance from foreign countries must be by international money order or by a draft on a U.S. bank.

**United Nations publishes first monitoring report**


The report appears in two volumes: the first volume comprises an introductory overview of population trends and policies and a report on population trends; the second is a report on population policies. Volume I was prepared by the Population Division of the Department of Economic and Social Affairs with inputs from a number of other U.N. agencies; the volume on population policies was prepared by the Population Division on the basis of information gathered from the Third Inquiry among Governments on Population and Development, as well as from other sources. Both volumes contain numerous tables to support the text.

Because the 1977 round of monitoring was the first to be held, the scope of population trends has been interpreted widely in order to cover such basic demographic variables as population growth, fertility, mortality, migration, urbanization, and the main structural aspects, including age, sex, labor force participation, and dependency. Also included in the report are the interrelations of population and food and population and education.

Although the U.N. Population Division has published a number of assessments of population trends in the past 20 years (for example, the publication reviewed below), this is the first time that a report has covered such a wide range of simultaneous studies of world population trends. It would have been nice if the report had gotten into print faster, but nevertheless, it is an important United Nations document for all those working in the broad field of population studies. At its nineteenth session, held in January 1977, the Population Commission reviewed a draft of *World Population Trends and Policies* and concluded that the two volumes "contain a wealth of information and analyses which should be made available to Governments, demographers and planners. The finalized versions of those two studies, together with their annexes, should be given wide circulation."

Those wishing to purchase the volumes may do so from the U.N. Sales Section, New York, New York 10017, or from the U.N. Sales Section, Geneva, Switzerland. The sales numbers and prices are as follows: Volume I, *Population Trends*, sales number E.78.XIII.3, US$16.00; Volume II, *Population Policies*, sales number E.78.XIII.4, US$9.00.
Latest United Nations report on world population


World population has just passed through the most dynamic quarter century in its history. Consequently it enters an end-of-century period marked by enormous uncertainty with regard to long-term prospects. Vital trends by region have been departing from known precedents in such historically novel ways that judgments about the decades ahead have become uniquely risky ventures into the unknown (p. 1).

Released early in 1979 as number 63 in the Population Studies series published by the United Nations Department of International Economic and Social Affairs, this slim volume reflects the latest statistical analyses of population data collected through the United Nations Statistical Office. It gives a brief overview of the present world population situation and discusses population prospects for the next century. Other topics addressed include the social and economic aspects of population growth and governmental policies on population.

A number of useful tables appear in the text and a ten-page annex table covers the chief demographic characteristics of major regions and of individual countries within those regions. Figures are given for 1975 and estimated for the year 2000. It is too bad that the author or authors of this paper are not identified, because someone should be commended for the brevity and clarity of the prose. The report is available from any local United Nations document distributor or by writing to Unipub, Box 433, Murray Hill Station, New York, N.Y. 10016. The sales number is E.78.XIII.8, and the price is US$6.00.

New ESCAP report on social indicators

The latest in the Asian Population Studies series is number 43, *Measuring the Social Impact of Population Trends and Development: Report of an Expert Group* (Bangkok, 1979), published by the U.N. Economic and Social Commission for Asia and the Pacific (ESCAP) in cooperation with the Population Center Foundation of the Philippines. This is the report of a meeting held in Manila in October 1978 that brought together 17 experts from 11 ESCAP countries for a week-long exchange of views and experiences in the area of social indicators. The report, which contains recommendations for the future development of social indicators, is available from ESCAP, U.N. Building, Rajdamnern Avenue, Bangkok 2, Thailand.

Letters to the Forum (continued from page 14)

population at age x. Assume (1) that the “heaping” at age x derives from errors in the range (x−2) to (x+2) and from none outside that range; and (2) over the range (x−7) to (x+7) ux is a polynomial function of the third degree.

Then, from (1), if \( w_x = \frac{1}{2} u_{x-1} \), where \( u_x \) are the observed values, \( w_x \) contains no heaping error, so that

\[ w_x = \frac{1}{2} u_{x-1} \]

In terms of the usual finite difference functions,

\[
\begin{align*}
    u_{x-1} &= u_x - 7\Delta u_x + 28\Delta^2 u_x - 84\Delta^3 u_x \\
    u_{x-2} &= u_x - 6\Delta u_x + 21\Delta^2 u_x - 56\Delta^3 u_x \\
    u_{x+7} &= u_x + 7\Delta u_x + 21\Delta^2 u_x + 35\Delta^3 u_x
\end{align*}
\]

by virtue of (2)

\[ \Delta^3 u_x = 0 \text{ for } n>3 \]

\[ \Delta^2 u_x = 0 \text{ for } n>2 \]

\[ \Delta u_x = 0 \text{ for } n=1 \]

\[ u_x = u_x \]

If we denote the five-year finite difference operator as \( \Delta \) then

\[
\begin{align*}
    \Delta u_{x-5} &= 25\Delta u_x - 75\Delta^2 u_x + 200\Delta^3 u_x \\
    \Delta u_{x+5} &= 25\Delta u_x + 50\Delta^2 u_x + 75\Delta^3 u_x
\end{align*}
\]

so that \( \Delta^2 u_{x-5} = 125\Delta^2 u_x - 125\Delta^3 u_x \)

\[ \Delta^2 u_{x+5} = 125\Delta^2 u_x + 125\Delta^3 u_x \]

(3)

From the pivotal values \( u_x \) calculated from the above, the intermediate values \( u_{x+1} \) can be calculated by any of the standard interpolation formulas. It is to be observed that this method does not require any assumption of linearity at any point and assumption (1) seems entirely reasonable. Further, assumption (2), in postulating a third degree fit, derives third degree functions locally at each individual pivotal age and allows the function to vary from pivotal point to pivotal point. Thus the final life table derived is a smoothed aggregate of third-degree curves—the so-called oscillatory interpolation.

Griffith Feeney replies:

King’s correction procedure, which I had not previously been acquainted with, assumes that the true age of persons incorrectly reporting age x lies within two years of x. My procedure assumes that the true age lies within four years of x. I tried the two-year assumption, which has the appealing simplicity of redistributing persons whose ages are multiples of five to disjoint age groups, but the results suggested that this assumption was untenable for Indonesia. Hence the shift to the four-year assumption and the complication of iteration.

Is King’s method generalizable to the four-year assumption? If so, it would be interesting to apply both methods and compare results.

I am grateful to Mr. Gould for calling attention to King’s work.