## Discussion W. Parker Mauldin, The Population Council

Introduction: The studies reported on by the Lauriat-Jabine team in Pakistan, Thailand, and Turkey are large in scale, serious in effort, important in purpose to economic planners, health personnel, and other government officials. Each of them applies the Chandrasekaran-Deming formula and procedures which were presented to us in 1949, and then lay dormant for more than 15 years. Such procedures are relatively expensive in terms of time required to collect and analyze the data. expensive in terms of funds, and, more important, expensive in terms of skilled manpower required. Those involved with these studies have reluctantly concluded that the present state of our knowledge of surveys is such that reliance cannot be placed on vital events collected only by retrospective surveys. This is the major issue of this session: Can valid estimates of vital events be obtained from sample surveys? Such surveys are useful for many purposes and they are relatively inexpensive. Must we seek other ways to get current information about vital events in those areas where registration is sorely inadequate, in areas where it certainly will require many years to produce good registration data throughout the country.

Linder and Simmons note that a very great deal needs to be done before adequately proven survey methods are a part of our statistical arsenal. They offer some principles as guidelines for research and for action. They suggest that one line of attack is intensified effort to develop and utilize more incisive survey procedures. They remind us that many methods must be tried because the current state of the art requires this.

A. Linder-Simmons: During the past 2-3 years I have had an opportunity to discuss this topic with Linder-Simmons and their colleagues on several occasions and to collaborate with them on aspects of the Pakistan program. I am happy to associate myself with the tenor of the Linder-Simmons remarks and to say that I have no major disagreement with them. I might make a minor quibble with reference to their statement that family planning programs, if they are to be effective, are necessarily expensive -- they cost only a few cents per capita; and at times I think that isn't so very much! I think they overstate the reliance of national programs on demonstration programs. And where they suggest that a family planning program developmental area must be visualized as including populations of the order of magnitude of 1,000,000 persons, I would prefer to say one should think in terms of a population of six digits more often than seven -and thus we are only one digit apart! I would cite the studies in Taichung and Sun Dong Gu of Seoul, Korea, each with populations of 300,000 as very good examples of the type of research study that they have in mind. Also, I would want to question their statement that the universe of study will be the Family Planning Developmental Area rather than an entire nation (p.11). This

seems to me to be applicable to India to a greater extent than most other countries. But these are relatively minor points. I agree with them on their major propositions.

B. Horvitz: The Triangle Research Institute project reported by Horvitz is most promising and is very unusual. Their design is so clean and pure and shining and complicated that it reminds me of one of our spacecraft -- highly sophisticated, enormously impressive, and the thing might actually work! Their design includes households in which births and deaths are known to have occurred: some of the births were illegitimate and, as expected, a much larger proportion of these were not reported to the enumerator. It is an anomaly that a larger proportion of illegitimate than legitimate nonwhite rural births were reported; the sample sizes were rather small, I believe, and it is doubtful that this differential is statistically significant (the figures are 98.5 and 92.2). The design also included deaths from socially stigmatized causes such as those from tuberculosis, venereal disease, mental diseases, cirrhosis of the liver, congenital malformations, suicide and homicide.

Recall lapse: The most surprising finding reported by Horvitz is that "There was no indication that underreporting increased as the interval between the interview date and the birth date of the infant increased." He notes that this is contrary to the Indian National Sample Survey which finds that memory decay increases with time. Horvitz notes that the differential noted by the Indians could occur if the underreporting were confined to those births which did not survive the first year of life. This assumed an Indian infant mortality rate of about 180/1000 as compared with a rate of 27/1000 in the U.S. He leaves it to us to make the inference that this difference in infant mortality might indeed account for the memory decay found in India but not in his study. Data from the 7th round of the NSS indicate that if the index for the birth rate at the point of origin is considered to be 100, respondents report decreasing proportions of births each month, with the index being only 83 in the twelfth month. Similarly, the index for reported deaths decreases to 67 for the twelfth month. I believe that a good deal more than infant deaths are involved in the memory decay observed in India. The Indian NSS also reports another interesting phenomenon, namely that by increasing the reference period to two years instead of one year, the number of events reported for the last year increases appreciably.

Persons in developing countries are not as well educated as those in the Horvitz sample; they are not as <u>time</u>- and <u>number</u>-conscious as are people in a developed economy. Perhaps the memory decay that has been observed in developing countries does not occur in ours. We shall look forward to a more detailed report on this at a later time.

Reference period: The Pakistan and Thailand surveys have adopted a quarterly survey procedure; in the Thai survey the reference period is for a quarter whereas in the Pakistan survey it is for a year. The purpose of inquiring about events during the past year in each quarterly survey is to avoid the boundary effect. If there is some telescoping of reporting events one gains by obtaining reports for a longer period than that for which data are required, provided you can later sort out the date of the event. Similarly, if respondents are not time-conscious, there would be some more or less random variation in reporting the date of a vital event, and if there is no overlap in the reference period one would tend to miss some of these events. The failure of the Thai survey to have an overlapping reference period tends to underestimate vital events, it seems to me. I don't have any data with which to quantify this observation, but as a matter of procedure it seems to me to be unfortunate.

We do not have definitive information about the optimum length of reference period, nor of the optimum frequency of surveys. I have argued with both our Pakistani and Thai colleagues that quarterly surveys are unnecessarily frequent; but they counter with the strong argument that enumerators will do a better job if they are in the field once a quarter collecting data on the same topic.

Open interval-pregnancy history: In a search for sensitive indicators of changes in level of fertility it has been suggested that use of the "open interval" -- the length of time since termination of last pregnancy -- would be rewarding. The idea is that if the distribution in months of time since termination of last pregnancy increases, women are postponing or reducing the number of births. Presumably this would occur first among higher parity women in a country where a family planning program is being launched. There are problems of obtaining sufficiently good dates of termination of last pregnancy, but this is an approach worth pursuing. There is also the possibility that well-trained interviewers could get good data on pregnancy history, with which one can reconstruct fertility rates for the women interviewed. I am personally pessimistic about the pregnancy history approach, but should like to encourage others to try it!

Randomized response technique: The TRI group has been experimenting with the randomized response technique suggested by Stanley L. Warner. This involves alternate wording of questions or statements, e.g., answer as true or false (1) I have never had an induced abortion, or (2) I have had an induced abortion. The questions or statements appear with a specified frequency, but the statement being answered is known only to the respondent. This approach is promising for sensitive subjects such as abortions, illegitimacy, crimes, use of drugs, cheating on exams, etc. A major disadvantage is that the sample size required increases sharply. I don't see any application of this technique for births and deaths on a large-scale survey, although if estimates of illegitimate births are needed, this approach might well be used.

<u>Registration in sample areas</u>: A number of persons have suggested that the least expensive and best single method of obtaining estimates of vital events in an area having poor registration data would be to establish a special registration system, or else "beef up" registration in a probability sample of areas. This is such an obvious technique that it is surprising that it hasn't been tried. One would have to supplement it with censuses, or head counts, but such counts need to be made only once in two years, I believe.

<u>The Chandrasekaran-Deming approach</u>: It involves collection of vital events by two independent systems, the comparison of those events on a case-by-case basis, and a classification of the events as:

- 1. Counted by both procedures, say, survey and registration
- 2. Counted by the survey procedure only
- 3. Counted by the registration procedure only

From this one can estimate the proportion of events missed by <u>both</u> procedures. They also suggest that if one can divide the population into homogeneous strata, the estimate can be improved.

Independence: Complete independence is difficult to achieve, but as Lauriat points out the studies cited apparently have done reasonably well on this score. I should like to note that the general effect of lack of independence is to underestimate the events that are being counted. If one of the investigators copies events from the other, or obtains them with the same biased procedure, e.g., talking with a few respondents rather than all those designated, then coverage is inadequate and some events are likely to be missed. Inadequate canvassing procedures are analogous to the Bureau of the Census Post Enumeration Survey in which roughly the same canvassing procedures are used as in the Census. The result is that the two systems are not wholly independent, and certainly there is an underestimate of non-white young adult males. Here the problem is not collaboration between two investigators but rather the difficulty of devising different procedures, each of which is adequate or better.

<u>Matching</u>: The C-D formula necessarily assumes that matching is carried out perfectly; not necessarily quickly or easily, but perfectly. One cannot escape this assumption, nor can one escape the consequences of a mismatch. If one sets the matching criteria so that uncertain cases are counted as separate events, then one is "sure" that all events counted as matches, are real matches. But the doubtful cases, some of which probably are matches, are treated as separate events, thus counting them twice, and thereby inflating the estimate. As a matter of fact, one counts them as slightly more than two cases -one for the survey, one for the registration, and a fraction of this for the 4th cell -- those missed by both. This is true because NrNe/C has been inflated in the numerator, and deflated in the denominator.

On the other hand, if one adopts loose criteria, counting all doubtful cases as matches, thus insuring that each non-match is indeed a separate event -- by so doing one counts some separate events as the same event, and thereby decreases the estimate of total events.

The problems of matching are difficult indeed. In Pakistan, for example, babies frequently aren't given a <u>name</u> for weeks; there is a reluctance to speak, to use, the name anyway; in some societies where males are highly prized, a respondent may report the sex as female in order to ward off the evil eye -- but at a subsequent interview, possibly with the husband rather than the wife, the respondent may correctly report sex. The typical Pakistani village does not have house numbers or street names. As a consequence specific addresses are non-existent. In order to overcome this deficiency, the population growth estimation study numbered each house -- the enumerator carried with him metal placques and nailed one of these on the front entrance of each dwelling unit. There were also difficulties of transliteration of names into English.

I could go on in this vein for some time, but surely the point is clear -- matching is difficult; there is no theory of how to compare different records and determine what is a match; but collection of data by two different systems discloses with "certainty", if I may be permitted that term, that each system has missed some event; indeed, in studies undertaken to date it is abundantly clear that each system has missed appreciable numbers of events. Thus matching is a necessity.

We have devoted too little attention to these problems; and I hope that we can seek ways to add to our knowledge in this field. Let me advance one or two suggestions. What we need is an objective criterion of what is a real match. I believe it would be feasible in a methodological study for the registrar to take a footprint of the infant, and attach it to his report to headquarters. Similarly, the enumerator could take a footprint of the infant and forward it to headquarters. One could then carry out the matching procedure in the conventional fashion <u>after which</u> the foot prints could be read independently and match conclusively.

There are problems in this approach, of course (1) Mothers would object -- but give them a gift of cloth, a chance at a prize, a small inexpensive painting, a special certificate, tobacco -- even money! I'm sure they would cooperate, and at a price that is not too expensive.

## (2) <u>Some babies would be missed</u> -- those who die early; those who are seriously ill; those who move out of the area. But the residual group would be large and would contribute useful information. Indeed, one could take fingerprints of one or both the parents in some of these cases, and that would be helpful in the matching process also.

(3) You may also say that taking footprints is a tricky business, and that in the distant village of Pakistan where piped water and electricity are unknown, taking a clear footprint is virtually impossible. I know some of the problems; I have read some of the literature and talked with some of the experts. Perhaps I can dispose of this argument by saying that I know of experts who would be delighted to teach enumerators and registrars how to take footprints, in exchange for a trip to a distant land.

I do not mean to dwell overly long on the possibility of using footprints as a positive means of identification. The president of the ASA, Prof. Fred Stephan, was intrigued by this idea, and suggested one might also consider voice patterns, using a tape recorder to aid in the match. We know very little about voice patterns, and I could not resist asking Prof. Stephan whether it might be necessary to equip each investigator with a sterile pin, to stick the infant in order to insure a uniform "word", or its equivalent!

I do not mean to suggest a specific solution to this important problem -- footprints of infants, voice patterns of babies, fingerprints of parents, photographs of the family, a tattoo mark that would last for a few weeks only, etc. -- but I want to emphasize the very great need, the lack of an adequate statistical theory, and the opportunity to contribute to a fascinating and important problem.