FAMILY PLANNING EXPERIMENTS: A REVIEW OF DESIGN

W. Parker Mauldin and John A. Ross, The Population Council

The last few years have seen a number of field experiments of relatively large size, all dealing with highly sensitive matters, and carried out mainly in underdeveloped areas. They have been inspired primarily as guides to larger family planning programs, but they attract interest for other reasons as well:

-- They have produced a large body of cross-cultural data on the sociology of fertility and its reduction.

-- Their before and after sample surveys give insights into changes in knowledge about, attitudes toward, and practice of family planning; and also give information about demographic changes, both secular and planned, past and present.

-- They have illuminated diffusion dynamics, especially with reference to adoption of the new intrauterine device (IUD).

-- They have proven the feasibility of such field experimentation.

This paper focuses upon another aspect, their character as experimental designs.

The India-Harvard-Ludhiana Population Study: This project was a pioneer effort to apply the methods of epidemiology to the population problem. The project was designed to determine whether the population of rural India would practice a simple method of contraception sufficiently to make a significant change in the rate of population increase. In order to evaluate the effectiveness of the program, researchers needed dependable figures on population, births, deaths, and migration. In the course of securing reliable base-line figures, deaths were examined for cause, births for factors favoring survival, and migration both to aid in getting base-line figures and to learn the reasons for movement. The underlying purpose of the program was a quantitative evaluation of the factors affecting fertility and population movement.

The study was done in 16 villages in a rural area in Ludhiana District, Punjab State, in northwest India. The 16 villages had a population of about 16,000 in January 1960. The action program was run in three stages: an exploratory stage in which the contraceptive method to be offered subsequently was selected; a pilot study in which the acceptance of the method was tested; and "the definitive test" designed to measure the effect on the population, and particularly on the fertility rate. During the exploratory phase the foam tablet appeared to be the most acceptable simple method of contraception, and this single method was offered in the action program which continued for a period of four years.

The over-all program was designed along the lines of a laboratory experiment: the exploratory study was to determine the most acceptable of five simple contraceptive methods, the pilot study made a preliminary test of the acceptance and effectiveness of the method under field conditions, and in the definitive study a test population was measured against both a working control population (to isolate the effect of the contraceptive advice and materials) and a "blank" control population (no project workers except for the collection of demographic data from village officials). In order to test the effectiveness of the contraceptive program (and the medical influence in the working control villages), accurate and continuing figures on births, deaths, and the rate of natural increase were required in the three types of population -- test, working control, and "blank" control.

In each village, an area survey was made at the outset -- maps of the villages showing all residences -- followed by a census by residence. In addition to such demographic data, the staff monthly collected information on attitude, knowledge, and current practice of family limitation. And during the course of the study there was an annual census of all villages, except those in the "blank" control where the census was taken only at the end of the study, in 1960.

The results of this carefully designed study were disappointing. After two and a half years of a sustained program, 17% of couples established themselves as contraceptive users. Furthermore, acceptance and use rates followed a characteristic pattern: although 80% reported in the preliminary survey that they wished to learn a contraceptive method, only about 45% of the couples used some method of family planning at some time in the study period. Perhaps 30% at any given point was the largest participation in the program, and that figure then fell to 17%. Observed efficiency among regular users of foam tablets was 63%.

The experimental design was excellent; but in our view it was over-elaborate for the program input. I think it is not hindsight to say that dependence on a single method was unnecessarily restrictive, and in this case apparently doomed the experiment to program failure, however successfully the design was

adhered to and however accurately the data were collected. A more fundamental criticism is that the informational and educational aspects of the program were limited in scope, in theory, and in application. Frequent home visits where there was face-to-face contact were wrongly regarded as frequent "intimate" contacts. Each worker was left to his or her own devices as to what was said at each visit; there was no systematic plan of what messages were to be imparted at each visit, no systematic recording of what transpired at such meetings other than the collection of demographic data. Long before the experiment had run its course it was clear that the impact of the program on the fertility rates would be minimal.

The Singur Study: Singur is in West Bengal, not far north of Calcutta, and is the site of the demonstration and research center of the All-India Institute of Hygiene and Public Health. This study is an outgrowth of the Harvard-Ludhiana study and was designed to offer several simple methods of family planning, using local field workers rather than high-level workers from outside the area, and to provide more systematic information about family planning.

The study covered a group of 8 experimental villages with about 7,500 population, and 15 control villages with about 13,000 population. Both male and female field workers were used, usually a husband-wife team. Preliminary contacts were made with village leaders, elders, and medical practitioners. Community meetings followed, separately for men and women, with a maximum attendance of 200 for women and 250 for men. The major part of the action program consisted of small group meetings with an average attendance of four to six, lasting an hour or more, and individual meetings between field workers and married men or women at their homes approximately once every two months. At a typical group meeting the field worker gave a prepared talk covering (1) introduction to the subject as part of the activity of the Singur Health Centre, (2) economic and health reasons for family planning, (3) physiology of human reproduction including especially the cycle of ovulation and menstruation as a basis for the use of rhythm, and (4) methods of family planning, initially rhythm, coitus interruptus and foam tablets -- at a later stage the condom was used, then vasectomies, and still later the IUD's. After the prepared talk there was free discussion within the group. The teaching was reinforced with visual aids: flip charts, flannel boards, and flash cards, each telling a story related to family planning. Wall calendars were distributed as a reinforcer and reminder of the program, the first with a portrait of Tagore, the Indian poet and Nobel prize winner, and a quotation from himthat condemned having more children than could be well cared for.

The results from this study showed that fertility rates fell more rapidly in the experimental than in the control area; more important, this was perhaps the first program in a rural, underdeveloped area to demonstrate that fertility rates could be brought down using conventional methods of contraception. Periodic surveys demonstrated that although couples were learning about methods of family planning, in the early months they often were misinformed about physiology of reproduction, about rhythm, and about the mechanism of action of foam tablets. Information collected in this manner was necessary for the continuing revision of the educational program.

The Koyang Experiment: In 1962, near Seoul, Korea, 14 villages became the target of two birth control programs. One program was actually nationwide in scope, but was studied by investigators only in a group of seven villages called the Kimpo area. On the far side of a large river, in the Koyang area, seven other villages were exempted from the national program and received instead a more intensive local program. Altogether, some 3,200 married women in the childbearing ages were affected. Before, during, and after measures were obtained on a sample of 1,200 women. Preliminary results were quite clear that both programs succeeded, with the local, intensive program being more effective than the less intense nationwide program in raising levels of contraceptive use, in reducing pregnancy rates, and in spreading favorable attitudes toward family planning.

Dacca, East Pakistan: This project illustrates the complexity and difficulties that are encountered in a sizeable field experiment. A principal aim was to compare the effectiveness of four treatments for spreading the practice of family planning. The four treatments were executed in four different areas of the city, each involving visits by trained field workers to the home. The four treatments differed according to who was contacted:

1. Both men and women
2. Men only
3. Women only
4. No one (the control group was served by a pre-existing and quite low key government program)

The hypothesis was that the treatments would be effective in the order listed.

Each treatment was brought to bear on one of four government "housing colonies." These were selected from a larger group of colonies and were matched on a number of demographic characteristics. Each colony contained 150 or more apartments, and all colonies were filled with clerical or lower scale government workers. The four colonies were well insulated from each other spatially, and residents of

each colony worked mainly in the same governmental agency.

A before-survey of demographic characteristics, fertility, and family planning attitudes, knowledge, and practices was conducted in each of the three areas receiving stimuli. It was recognized that in theory the questioning involved in the pretest might change behavior, but the design did not permit a measurement of this effect.

Several features of the experiment were calculated to permit generalization to the sort of groups that would become future targets. First, residents of the four colonies were not representative of Dacca's population. Instead, they constituted a homogeneous type which would receive high priority in family planning efforts. Second, the treatments applied called for cost and personnel requirements that were realistic for general application. Third, the records systems used greatly improved the action aspects and were realistic for wider application.

Strenuous efforts were made to keep field procedures uniform over all treatments. Workers of about equal qualifications were obtained and underwent careful, uniform training and similar supervision. In addition:

1. Home visitors were rotated among the three stimulus colonies.

2. A clinic was set up in each of the four areas to dispense contraceptives and information; the four staffs were rotated quarterly.

3. All action procedures were checked constantly to keep them uniform, i.e., initial home visits, follow-up visits, group discussions, large group meetings, etc.

4. Clinic and field records were kept to monitor all inputs to the colonies, including some stimuli not part of the family planning program. All this required supervision and painstaking maintenance so that, insofar as possible, the only difference between the treatments would be the sex channel used: men, women, both, and neither.

The clinics afforded a measurement of the results. Attendance at the clinics and the flow of contraceptive supplies were two quickly available measures. Diffusion could be gauged by the number of persons from outside the colony coming to the clinic for information and service. The after-survey gave additional measures of program effectiveness.

The Taichung Study: One of the most ambitious and successful field experiments in the social sciences ever conducted occurred during 1963 in the city of Taichung, Taiwan. The population of 300,000 was allocated to a twelve-cell design and exposed to nine months of stimuli to encourage family planning. Four different types of programs were instituted:

1. In all areas posters were displayed and mass meetings held in various public halls.

2. In some areas in addition to the above, letters and pamphlets were sent to married couples with at least two living children and to newlyweds.

3. As an additional program input in some areas field workers also visited every married woman aged 20-39 in order to inform her about family planning services; small group meetings were also held.

4. Finally, husbands were also visited in some areas.

These treatments were mixed in different proportions in three more or less similarly constituted sectors of the city. In the first sector, one-half of the neighborhoods received home visits; in the second sector, one-third; and in the third sector, one-fifth. "Mail" and "nothing" treatments were distributed equally in the remaining neighborhoods in each sector. For political reasons, the use of mass media was kept to a minimum and the posters and letters made no reference to the IUD. The target population was about 36,000 married couples, with the wife being between ages 20-39. This study is notable for its size, its multiple design features, the inclusion of IUD's as a method of contraception, and the fact that the study was undertaken in an area of relatively high literacy and, for an underdeveloped country, among a population with moderate levels of living.

From its start in February, 1963 to April 1, 1964, 5,454 couples accepted family planning guidance in the program. Nearly 80% chose the IUD; less than 2% the oral pills; and about 19% traditional methods, mostly condoms. The proportion choosing IUD's was least among those exposed to home visits, presumably because field workers gave explanations about various methods whereas diffusion of information by word-of-mouth was concentrated almost entirely on the IUD. The home-visit treatment yielded the highest returns, but additional visits to husbands did not yield much. The mail campaign was ineffective. There was not much difference between the light and medium density sectors, but the heavy sectors attracted appreciably more contraceptors than did medium and light sectors. This study is continuing in modified form and the proportion of users continues to rise.

Other Studies: This is not an attempt to list the various important family planning pilot or demonstration projects that have been undertaken. If we were to compile such a list we would include the large Sungdong Gu study in Seoul, Korea, modeled along the lines of the Taichung study; the Barbados experiment which combines an IUD program with an attempt to eliminate cervical cancer; various studies in Latin America designed to reduce the high incidence of abortion; the highly successful Photharam project among a population of about 70,000 people in rural Thailand; the Bogue studies in Chicago and the rural South; the comparative study of oral tablets and IUD's now underway in Puerto Rico; the studies in Taiwan and Korea of acceptance and continued use of oral tablets at different cost levels; and the comparative studies of what is called the hospital postpartum program -- a program to determine to what extent family planning can be spread through large maternity hospitals. This study involves about 20 hospitals in about 15 countries, and includes several of the largest "baby factories" in the world, those having 35,000 to 50,000 deliveries a year.

These experiments have been designed to determine (1) whether a given contraceptive method will be accepted and used and have an effect on fertility rates, and/or (2) whether a given treatment brings about a significant change in fertility. There are intermediate objectives,such as an increase in knowledge about physiology of human reproduction, contraceptive methods, location of clinics, etc., but the ultimate objective is to affect fertility.

Program inputs are almost infinitely variable, and are a composite of: (1) contraceptive methods, (2) informational content, (3) type of media, (4) frequency of stimulus, (5) intensity of stimulus, etc.

Some of the experiments have been designed to measure whether or not a given program has an effect, and typically differences between a control and the experimental population are used as measures of effectiveness. In other experiments two or more matched population groups are selected, different programs are introduced, and differences in the responses and behavior of the various populations are assumed to measure the relative effectiveness of the programs; in such cases control populations are sometimes, but not always, used.

It is difficult to identify and describe many of the stimuli in a manner that they can be replicated. A program may include posters, occasional radio "spots," group meetings, home visits, and information plus service at health clinics. One can describe and replicate the posters, the radio "spots," but not the group meetings or home visits. One can substitute another group meeting, another home visit; one can obtain information about how the group meeting was conducted, but it cannot be reproduced with precision. Indeed, records are not kept in sufficient detail to tell what went on at group meetings, although the general purpose and approach are known. In practice, generalization is also limited by poor specification of the services offered; for example, where clinics are located, what are their procedures, how are clients received and treated, etc. Even if this were known, an important source of variation would remain obscure. This is the set of unique personalities administering the program and supervising contacts with the populace. One could design an experiment which would largely remove this as a variable, but it would be time consuming and expensive, and of questionable worth.

Criteria: The measures by which stimuli are judged to be effective or not are mainly of three types: (1) acceptances, as indicated by IUD insertions, sterilizations, the practice of rhythm, and the taking of supplies of ordinary contraceptives; continued use is equally important, though less well reported; (2) fertility levels and change; and (3) KAP surveys designed to give information about knowledge, attitudes and practices of family planning. Information about acceptances within the program is readily available from service statistics, but special follow-up studies are necessary to give information about continued use. Locating initial acceptors who do not return for supplies or service is often very difficult. Sample studies of the population included in the program area gives information about persons who receive their information and services from outside the program -- from friends, private physicians, etc. -- and also can give one information about "dropouts" from the regular program. Very few of the developing countries have good vital statistics, and the typical survey procedure does not produce consistently valid estimates of birth and death rates. Several of the studies have used annual surveys or censuses supplemented by reports throughout the experiment from midwives and other knowledgeable people in the area. Some of the experiments have attempted to collect information from two independent systems, one a census and the other a modified registration procedure. Some thought is now being given to use of the "open interval," or length of time since termination of last pregnancy, as an early indicator of change in fertility -- if the distribution of months since termination of last pregnancy for women in the reproductive ages shifts so that the average interval is longer than before, particularly for high-parity women, this would indicate lower fertility rates.

Experimental Groups: Units assigned to the experimental groups have varied widely, but they have always been areas, not individuals. They have included villages, housing colonies, neighborhoods, and counties. The ideal of creating experimental

groups by random assignment has been realized partially, even impressively, given the magnitude and difficulty of operations involved. The principle of randomization has been followed in order to equalize the experimental material acted upon. For example, within each sector of Taichung and Sungdong Gu each of the thousands of neighborhoods was randomly assigned to treatments. In Dacca, each of the housing colonies was randomly assigned to one of the information channels.

Contamination: There are many sources of contamination in experimental programs. One might worry about a change in secular trend due to more widespread education, to a rising level of living, to more news articles about what other countries and groups are doing about population. One might reason that a before-survey which asks questions about family planning, and in the process almost necessarily gives some information about it -- one might reason that this is a contaminating influence. Some of the experiments have been designed to take advantage of natural channels of communication and to encourage diffusion of information.

Experiments have given some thought to such matters, but have not been greatly worried about most of them. In the first instance, it has proven to be rather difficult to change attitudes and behavior, particularly behavior so fundamental as having or not having children, of spacing or not spacing births. The problem has been to find a composite program that will have any effect at all! Further, the program specialists have been looking for things that make a considerable difference, not those that make a small difference and are hard to measure. Even so, attempts have been made to choose experimental and control populations that are geographically separated and which do not normally have much communication with one another. Care has been exercised to insure that the initial stimulus, the program stimulus is applied only to the population in the area specified; but no attempt has been made to decrease diffusion of information across experimental boundaries.

There is considerable evidence that diffusion of information, particularly about the IUD, is rapid and large in magnitude. It is likely that such diffusion works in the direction of reducing difference between the control group and the experimental group, or between the experimental groups with the least and with the most effective programs. Consider, for example, the differential effectiveness of a program that includes home visits, group meetings, well-located clinics with courteous personnel, posters, radio "spots," and all the rest. Call this the "everything" group. One may wish to determine whether a less intensive campaign is effective, say, a campaign that includes a few clinics, radio "spots" occasionally, and posters -- but no home visits, no group discussions. One may be concerned about the contamination of information given by people in the "everything" area to people in the "clinic" area. If the results of the experiment show that there is no difference between the two groups, and that the number adopting family planning is low, one's concern is not about lack of differences, but about lack of an effective program. If the differences are large, and in the expected direction, one is safe in concluding that the "everything" program is more effective than the "clinic" program. If the effects of both programs are large, and the differences are small, one would be considerably worried about contamination. The administrator would have to decide whether to invest funds for the "everything" program over a larger area; and if funds are tight he probably would try the less intensive approach, watch the results carefully, and plan to include more program input if initial results are disappointing.

Conclusions:

We are impressed with the size and complexity of these undertakings. The early studies seem to us to have worried too much about the measuring instruments, too much about contamination, too much about the possible influence of mild stimuli, and to have been too little concerned with major program inputs designed to inform substantial numbers, to answer their questions, doubts, fears, to offer a service economically, courteously, conveniently, privately. The design features of some of these studies have been very good indeed; most of them have been only fair. Some very sophisticated studies are needed to answer questions about the long-range effects of new contraceptive methods, about differences in effectiveness and side effects of the more promising devices and materials. But where the interest is to provide program information to an administrator who must make decisions about the number of field workers per unit of population, how they will go about their work, etc., it seems to us that the experimental programs we have referred to have given us substantial information. They could give us more information; but the aim is to provide useful information early, rather than more detailed information much later.