II. THE MAJOR METHODOLOGICAL ISSUES

Application of the various methods of measuring the effects of family planning programmes on fertility involved a number of critical methodological issues. Although it was not possible for the Expert Group to resolve those issues, it dealt with a variety of them and emphasized that awareness of the issues by those who attempted to use the various methods was essential if the limitations of the methods were to be fully appreciated. Moreover, an understanding of these methodological issues should lead to a better assessment of the results obtained with the different methods.

A. Data requirement problems

The Expert Group considered problems associated with the selection, use and interpretation of data to be among the more difficult of those attending measurement of the impact of family planning programmes on fertility. The principal difficulties related to the types of data needed, their sources and quality. The type of data needed depended upon the particular method to be applied and, in that respect, some methods were more demanding than others. Essentially, statistics on the composition of a population, fertility data of reliable quality and quantitative information on the scope of the programme were indispensable. Certain methods also required data on mortality and on biological factors. Frequently, a method that might be considered the most appropriate could not be applied because of lack of relevant statistics.

The data were drawn from four principal sources: population censuses; vital statistics registration systems; demographic or other special surveys; and family planning programmes (service statistics).

The opinion was that there was no substitute for reliable data based on population censuses. Census data, along with post-enumeration surveys and special surveys, could provide the essential data on the demographic, economic and social characteristics of a population and on various other non-programme factors affecting fertility, such as health and nutrition. The Expert Group recommended that post-census surveys, in which more detailed social and economic statistics were collected along with data on fertility, should be carried out as a standard procedure. A substantial sample enumeration at the time of the census would also provide the requisite statistics on a number of factors not associated with the family planning programme.

It was advantageous to collect the required statistics through sample surveys, not only because the surveys could be designed to meet specific data needs but because they could provide fertility data specific for other variables considered important. Surveys could also produce data on characteristics of acceptors, and they could be conducted more often than censuses, which provided data only at substantial intervals.

It was recognized that systems of vital registration were generally poor in most developing countries and it would therefore often be necessary to rely upon population census and sample survey data as substitutes for vital registration statistics.

The family planning service statistics system provided data on acceptors and their characteristics, and on other programme factors, including contraceptives supplied, use of specific methods and so on. Those statistics were a by-product of the records usually required for delivery of service by a family planning programme, and as they were not, in most cases, designed or intended to serve evaluation purposes, the statistics that they produced were frequently inadequate for scientific evaluation. That short-coming in the quality of service statistics was frequently a major source of error in evaluation efforts, and the system should be improved. However, for a majority of developing countries, demographic and other related data also were generally of inferior quality. Indeed, the inadequate quality of the service statistics was a reflection of the absence in such countries of a history of good data collection systems. Certainly, improvement of data on the activities of family planning programmes should be a primary goal. It would be appropriate gradually and incrementally to build a system for collecting adequate data for evaluation.

The amount of information routinely recorded at family planning clinics for each client should be only that necessary for the delivery of services, which should be the primary emphasis at the clinic level rather than data collection. For research purposes, it was not essential to have service statistics for each client; a sample would frequently suffice. At any rate, the service statistics system should not be regarded as the sole source of detailed data on acceptors (i.e., their characteristics, their contraceptive practices and their continuation of contraceptive use).

Careful thought should be given to the statistical data required from the family planning programme and the means of collecting it. The statistics needed for studying the impact on fertility of a family planning programme might be obtained from samples of family planning clinics or service units or through the use of periodic surveys of women in the childbearing ages. In that way, the burden for the collection of additional or special types of data of a service statistics nature could be placed on the staff of the evaluation unit of the programme.

There was not consensus among members of the Expert Group with respect to the role of family planning programmes in providing certain essential data. In opposition to the view expressed above was the opinion that service statistics represented a third major data source in countries with family planning programmes and that efforts should therefore be devoted to obtaining and improving the quality of such data. However, service records on use of conventional contraceptives, for example, were probably not needed. Techniques should be devised for obtaining such data directly from the sources supplying the conventional methods.

One solution to the problem of data on the use of conventional contraceptives was to make provisions for the collection of such data when a programme was established. Bench-mark surveys were an example of what could be done. Knowledge-Attitude-Practice (KAP) surveys yielded useful information, particularly at the beginning of the programme and at its early stages.

The Expert Group recommended that data should be collected by a census or a sample survey at least once every five years. Where sample surveys were carried out, provision should be made for the collection of detailed areal and socio-economic data needed for family planning programme evaluation. The samples of surveys designed for evaluative purposes should be of adequate size to provide data both on an areal basis for a country and on significant subgroups in the population.

It was observed that in various respects, censuses, vital registration statistics and service statistics were all limited. Surveys must be designed to provide the types of data not usually included in those more limited sources. Although one could question whether there existed an over-emphasis on surveys, some sample surveys were indispensable if the data necessary for evaluative purposes were to be obtained. A well-designed and carefully planned periodic survey programme could obviate the need for many hastily designed surveys and might even reduce the total number of surveys in some countries.

Other aspects of data requirement problems considered by the Expert Group were the types of problems encountered in the collection and classification of data. A great deal of error in evaluation efforts could be traced to the types of errors that arise in data collection and classification.

The evaluator was required to adjust for two categories of errors. First, there were non-random errors and biases that either affected all units similarly or were distributed in such a way as to have differential effects. Those errors included, *inter alia*, errors of non-response, recall lapse, and digit preferences in age reporting. An excellent example of the latter type of error commonly encountered was reported in the case study of Karnataka State, India.[1]

Well-known techniques were available for identifying and adjusting for such errors. For example, a number of techniques were available for correcting faulty age distribution due to the misreporting of ages. The choice of the proper method of adjustment, however, could be a major problem. As indicated in the Indian case study, the application of two different procedures for smoothing the age distribution yielded different evaluation results.

In general, the identification of the magnitude and the direction of non-random errors posed a particularly difficult problem; even for experienced statisticians, who were fully cognizant of the ramifications of such problems, solutions were not always easy. Non-random errors might affect all types of variables. Because of their nature and the difficulties in observing and measuring them, the Expert Group felt that such errors were potentially the most hazardous encountered in the application of the various methods of evaluation.

The second category of errors encountered in data collection and classification that constituted serious impediments to obtaining valid results from evaluation efforts were those termed "random". Such errors could originate either from sampling procedures or from other sources. The random errors that arose from sampling procedures could be dealt with in a satisfactory manner, as standard statistical theory provided a means for handling such errors. But adjustments of the random errors that resulted from chance factors not easily identifiable and that frequently had a large cumulative effect could be made only by the most complex procedures and frequently with limited results. Methods of evaluation that involved matching and regressional analysis were particularly susceptible to the influence of random errors.

It was considered that the extent to which it was possible to deal satisfactorily with the various types of error bore directly on the interpretation of the results of evaluation studies. Reports of such studies should, therefore, include a description of the techniques employed to assess measurement errors and other types of errors as well as the types of adjustments made. If that policy were followed, analysts would invite greater confidence in the results obtained in studies of the impact of family planning programmes on fertility.

B. Minimum data needs

The Expert Group recognized that the problem of data requirements was ever present in attempts to measure the impact of family planning programmes on fertility. Indeed, as previously mentioned, the

[1] In the section of that paper which deals with data problems, see figure entitled "Raw and smoothed population distributions of Karnataka State, based on 1971 census"

evaluator's choice of method might frequently be dictated not by the purposes of the evaluation or the validity of results that might be derived from the method but by the type and amount of data at hand. Given a wide variety of high-quality demographic data and a full complement of reliable family planning service statistics, it would be possible to apply in one evaluation effort any or all of the methods currently used for that purpose. Moreover, researchers should be able to sharpen the measures and to develop new, and possibly more reliable, methods of evaluation. But such an ideal situation never existed, and evaluators were compelled rather to work with scarce resources, making adjustments and assumptions where data were inferior or altogether lacking.

The Expert Group undertook, therefore, to list the minimum statistical data required for each of the eight methods of evaluation that were being considered during the meeting. Table 1 summarizes the types of data that were thought to be essential for application of each of the eight methods of evaluation. The minimal data requirements shown in table 1 were detailed as described below.

Standardization approach

The standardization approach required: (*a*) crude birth rates of the population at two points in time; (*b*) age-sex-marital status distributions of the population at two points in time; (*c*) general fertility rates at two points in time; (*d*) marital age-specific fertility rates of the population at one point in time; (*e*) if possible, data on (*a*), (*b*), (*c*) and (*d*) separately for rural and urban areas.

Trend analysis

The trend analysis approach required one of the following: (*a*) crude birth rates; (*b*) general fertility rates; (*c*) marital age-specific fertility rates and proportions married; (*d*) gross reproduction rates; (*e*) distribution of women by number of children born (parity); and (*f*) age distribution of the population during the period of the programme and before the start of the programme.

Experimental designs

The experimental-design approach required the following data for both the experimental and the control groups: (*a*) socio-economic variables (e.g., income, educational level, labour force participation, religion, ethnic characteristics, residence); (*b*) demographic variables (e.g., sex, age, marital status, parity); (*c*) biological variables (e.g., foetal mortality, amenorrhoea); (*d*) general fertility rates and rates for both groups at the beginning and at the end of the observation period.

Couple-years of protection

The couple-years-of-protection method required the following data:

(*a*) Programme variables by method: (i) for IUDs,

TABLE 1. ESSENTIAL DATA NEEDED FOR EACH EVALUATION METHOD

Data needed	*Evaluation methods*							
	Standardization	*Trend analysis*	*Experimental designs*	*Couple-years of protection*[a]	*Component projection*	*Analysis of reproductive process*	*Regression analysis: Areal*	*Regression analysis: Micro*
Fertility								
Crude birth rate	X[b]	X[b c]	N	N	N	N	X[c]	X[c]
General fertility rate	X[b]	X[b c]	X	X	N	X	X[c]	X[c]
Marital fertility	X[b]	X[b c]	N	[d]	X	N	X[c]	X[c]
Gross reproduction rate	N	X[b c]	N	N	N	N	X[c]	X[c]
Parity	N	X[b c]	N	N	N	N	X	X[c]
Programme variables								
Number of acceptors, by method	N	N	N	X	X	X	N	(X)
Continuation rates, by method	N	N	N	X	X	X	N	(X)
Number of users	N	N	N	(X)	(X)	(X)	N	N
Measures of programme input	N	N	X	N	N	N	X	X
Non-programme variables								
Demographic variables	X	N	X	N	[d]	X	X	X
Socio-economic variables	[d]	N	X	N	[d]	N	X	N
Cultural variables	N	N	X	N	[d]	N	N	N
Mortality rates	N	N	N	N	X	X	[d]	N
Biological variables	N	N	X	N	N	X	N	N
Other	N	N	N	N	N	X	N	(X)

Note: X = minimum requirements;
(X) = estimates from other data;
N = no application for data in method.

[a] Modified to give births averted.
[b] Data needed for two or more points in time; any one fertility measure would suffice.
[c] Any fertility measure would suffice.
[d] Optimum data.

number of users or number of acceptors and continuation rates for two points in time; (ii) for pills, number of users or number of acceptors and continuation rates for two points in time, or number of cycles distributed and estimate of wastage; (iii) for injectables, number of injections; (iv) for sterilization, number by age and sex; (v) for conventional contraceptives, number of units distributed and estimate of frequency of use and wastage;

(*b*) Fertility measures: fertility rates or age-specific fertility rates if users were known by age.

Component projection approach

The component projection approach required the following data: (*a*) programme variables by method as listed above for CYP but by age of user; (*b*) pre-programme age specific marital fertility rates of acceptors; and (*c*) age/sex-specific mortality rates.

Analysis of reproductive process

Potter method

The following data were needed for the Potter method[2] of analysis of the reproductive process. In the case of the IUD: $B = I/D$, where $I = F(R - A - PW)$ and D = average birth interval per birth required in the absence of programme contraception (i.e., the reciprocal of age-specific pre-acceptance fertility rate classified by never-use and use, or combined never-use and use of contraception during the pre-acceptance period of from three to five. In the formula to obtain I:

F = proportion of couples not sterile at time of acceptance deduced from L. Henry's estimates of age-specific sterility proportions (no special data required);

R = mean time programme contraception was used among couples not sterile at time of acceptance (deduced from the two parameters of the modified negative exponential distribution, which required data for an ordinary life-table continuation function as well as mortality data to calculate an age-specific annual risk of marital dissolution by death of a spouse and as judged necessary, data on divorce to calculate the annual rate of divorce; with the risks of discontinuation, marital dissolution by death of a spouse, and of divorce treated as independent within age classes);

A = mean overlap of contraception and post-partum anovulation (which required data on mean intervals from birth to acceptance of the method and on estimates of mean length of post-partum anovulation);

P = estimates by age class of the proportion of IUD terminations attributable to accidental pregnancy (which required data on reasons for discontinuation that included accidental pregnancy);

W = an estimate of the mean number of fecundable months required per conception in the absence of IUD, by age (usually a rough deduction using Perrin-Sheps model[3] from the values of F and anything known about level of pregnancy wastage and post-partum anovulation).

Wolfers method

The data needed for the Wolfers[4] method of reproductive analysis were thought to be about the same as those required for the Potter method, except that the estimates of potential fertility were obtained directly from questions related to birth intervals. Both cases involved the calculation of a life table following month-by-month post-acceptance experience. The table could be constructed for as long a duration as the data would permit. The results of the life-table calculations gave the proportions of acceptors who were still using a method and the number of births that were being averted. Separate tables could be calculated for different subgroups of the population (e.g., age, socio-economic group, type of contraceptive method used). In respect of the Wolfers method, corrections for duration of contraceptive use, post-partum amenorrhoea, accidental pregnancy and proportions who were becoming sterile were applied monthly.

Regression analysis

Application for areas

The application of regression analysis for areas required the following data for each area: (*a*) one current fertility measure other than parity (e.g., crude birth rate); (*b*) some measure of programme inputs (e.g., number of clinics, number of family planning personnel) and/or other programme variables that would summarize conditions prior to the date to which the measure of fertility related; (*c*) measures of non-programme variables having a direct or indirect effect on fertility (e.g., percentage of women in reproductive ages married, urban-rural residence and socio-economic variables).

Application for micro data

Regression techniques for micro data required: (*a*) one current fertility measure; (*b*) some programme variables (e.g., acceptors by method, continuation rates by method, a measure of programme input); and

---

[2] For a detailed exposition of this method, see R. G. Potter, "Estimatimg births averted in a family planning program".

[3] Edward B. Perrin and Mindel C. Sheps, "Human reproduction: a stochastic process", *Biometrics*, vol. 20 (1964), pp. 28–45.

[4] For a detailed exposition of this method, see D. Wolfers, loc. cit. Further, for a proposed synthesis of the Potter and the Wolfers methods, see David Wolfers, "Births averted", in C. Chandrasekaran and Albert I. Hermalin, eds., *Measuring the Effect of Family Planning Programs on Fertility* (Liège, International Union for the Scientific Study of Population for the Development Centre of the Organisation for Economic Co-operation and Development, 1975), pp. 163–214.

(*c*) some measures of non-programme variables (e.g., age of acceptors or other demographic characteristics, urban-rural residence, educational level); biological variables, such as foetal mortality, amenorrhoea and menopause status of women.

In the opinion of the Expert Group, it could not be overstressed that, in all cases, it was important to have accurate statistics and that data of at least moderately high quality were essential. The requirements specified by the Expert Group were far from optimal and, accordingly, some analysts might wish to impose higher standards. For example, for the estimation of couple-years of protection, statistics often were not available on users of conventional contraceptives; and, therefore, a crude estimate could be made from the number of pieces distributed to clients if it were possible to derive reasonably good estimates of frequency of use and of wastage.

However, the data suggested by the Expert Group represented minimum requirements; and, ideally, a greater variety of statistical information would be useful. In addition, it should not be assumed that the module of data needs presented in the present report was the only one that would serve; ultimately, that factor would depend upon the evaluation model applied. Further, it was of importance that although data on some of the variables might be needed for evaluation at one time, on another occasion a different set of data might be required.

C. Potential fertility

The Expert Group considered the estimation of potential fertility to be one of the most difficult problems encountered in the application of several of the methods. Potential fertility was a basic component of six of the eight techniques distinguished in the main background paper (ESA/P/AC.7/1) prepared by the Secretariat for the Expert Group meeting. The only exceptions were standardization and perhaps regression analysis. In regression analysis, however, potential fertility might be viewed as the fertility predicted by the regression equation when the independent variables which represented programme input were set at zero. In trend analysis, the projected birth or fertility rate was a measure of potential births. The projected (potential) births less the observed births yielded the number of births prevented. In an experimental design or matching study, the fertility of the control group defined potential births. The fertility of the control group minus that of the experimental group defined programme impact. In the CYP approach, couple-years of protection multiplied by an assumed rate of potential fertility yielded births averted. The same basic paradigm figured in the component projection approach. In the analyses of the reproductive process, a highly corrected mean couple-years of useful protection per acceptor was divided by a potential birth interval to give births averted per acceptor. In the case of simulation models, fertility rates with and without family planning were obtained. The latter measure of fertility was defined as the level of potential fertility. The simulated fertility rates with and without family planning were based on hypotheses about the value of various biological and family planning parameters. The derived fertility was then differenced as an estimate of the impact of a family planning programme.

Since either the measure of potential fertility or the observed fertility rate was being differenced or, alternatively, women-years of protection were being multiplied by a potential fertility rate, the resulting estimate of the impact of a family planning programme was extremely sensitive to the estimate of potential fertility.

The Expert Group noted that a distinction existed between total family planning programme impact and net programme impact. Total programme impact measured the effects of contraceptive use and other measures of fertility control regardless of whether the users adopted a method through the programme or adopted one outside the programme. Net programme impact took into account both the substitution effect (i.e., the extent to which the observed fertility behaviour would have occurred in the absence of the programme) and the catalytic effects of the programme (i.e., the extent to which the programme had stimulated fertility control that had at that time been provided from non-programme sources).[5]

Although potential fertility was non-observable, probable ranges of potential fertility might be estimated. The Expert Group took note of the statement in the background paper: ". . . estimating potential fertility consists of determining a particular fertility level that did not materialize. It is not possible to know with certainty what such a fertility level would have been and this problem cannot be solved. The purpose of the procedure is thus to compute reasonable estimates" (ESA/P/AC.7/1, section B).

Even if one dealt with total family planning impact and even if one treated potential fertility as natural fertility, the measurement of the potential fertility of contraceptive users was not straightforward. Users were to some extent selected for higher than average fecundity and were non-random with respect to physiological status, as there was less chance that they were pregnant or secondarily sterile than a married woman drawn at random from the same age class. If interest was focused on net programme impact, the Expert Group felt that there were a number of additional issues: what alternative contraceptive methods existed that the programme clients might have used if there was no programme; how much of contraception adopted by women outside the programme was attributable to the catalytic effect of the programme.

The basic significance that estimates of potential fertility had in most measures of family planning im-

[5] See Ronald Freedman and Bernard Berelson, "The record of family planning programs", *Studies in Family Planning*, vol. 7, No. 1 (January 1976), pp. 1–40.

pact and the difficult problems encountered when an attempt was made to estimate potential fertility had been previously recognized in a critique of the births-averted concept.[6] The Expert Group believed that progress would be facilitated if the concept of potential fertility, the estimation techniques and the data requirements were related to the classification of approaches presented in the main background paper (ESA/P/AC.7/1, section B). At the empirical level, progress could be achieved if provision for collecting data relevant to estimating potential fertility were built into study designs. Current designs tended to focus on acceptance and continuation measures of clients' behaviour in a family planning programme and to neglect the problem of estimating potential fertility, especially the problems of definition and measurement.

Problems of definition

Use of the term "potential fertility" to describe the level of fertility that might have characterized a population if its past reproductive pattern and the conditions that affected it had remained unchanged was not an entirely satisfactory means of identifying the concept. And although "expected fertility" was thought to be a reasonably clear term, since it had a well-defined meaning and was clearly understood to imply what would be expected under certain conditions, its meaning in demography was well established as applying to values derived in the standardization method. An alternative term might be "hypothetical fertility", which would imply the unrealistic nature of the measure. Moreover, the term "hypothetical" would distinguish the concept from that of "expected" fertility. It would be unfortunate if potential fertility were to be equated with expected fertility.

The Expert Group cautioned that potential fertility was not to be confused with "natural fertility", which was the level of fertility exhibited by a population that made no deliberate attempt to control its fertility. Natural fertility was not the biological maximum. In most populations, a number of socio-cultural checks tended to operate in such a way as to depress fertility well below what was believed to be near the biological limit. Therefore, defining potential fertility at the level of the biological limit would produce an overstatement of potential fertility in most populations. As previously stated, in the framework of family planning programme evaluation, potential fertility referred merely to what would have happened to the fertility of a population in the absence of the programme.

Another aspect of the definitional problem was that of determining the group of women on whose fertility performance estimates of potential fertility should be based. The Expert Group discussed the question of the various groups of women for whom potential fertility could be calculated and suggested possible time dimensions for the analysis. The results of their discussions all shown below in table 2. The time periods chosen indicate that measures of potential fertility could be calculated for a period either prior to the initiation of a programme or after a programme had begun to operate. If the measure of potential fertility to be used was for some period prior to the initiation of the programme, it could be based either on observed fertility rates, which were preferable, or on some estimates of those rates. But for many countries, the requisite data were not available. Thus, the measures of potential fertility would have to depend upon some estimates of the fertility level. Alternatively, that measure could be calculated for the period following the beginning of the programme. As indicated in table 2, either observed or estimated fertility could be used as the measure of potential fertility for the first three categories of women. If the measure was based on observed data, such a measure would, however, presumably produce underestimates of potential fertility since acceptors would be included in those groups.

If the measure of potential fertility desired was for acceptors only or for other more specially defined groups of women (as suggested in the lower part of table 2), it would be necessary to estimate what the fertility of those groups of women would be if no programme had been in existence. It would be required to base the estimates on a number of assumptions, and the type of assumptions that should be made posed difficult problems. If potential fertility was calculated for acceptors, it could be based on specific methods of contraception; women selecting different methods differed in terms of their motivation; and, for that reason, their potential fertility would differ. Another dimension that might be introduced into the table was that of other characteristics of acceptors (e.g., educational level or place of residence). Thus, other columns could be added to the table to show the measures of observed fertility that might be subtracted from the potential fertility measures.

There was great difficulty in classifying women as acceptors. Many women entered a programme, left it and then re-entered. To facilitate research on evaluation, acceptors might be classified into various subgroups, such as ever-users and current users. The term "use" implied prevalence and the term "acceptor" implied incidence, and the Expert Group felt that there was no need to link the term "acceptor" with that of "user".[7]

Problems of measurement

Opinions varied as to the appropriateness of measuring the potential fertility of different groups of women

[6] W. Parker Mauldin, "Births averted by family planning programs", *Studies in Family Planning*, vol. 1, No. 33 (August 1968), pp. 1–7.

[7] The general definitional problem of "acceptors", "users" etc. has been discussed in detail by another United Nations Expert Group, and the definitions recommended by that Group should be followed. See *Assessment of Acceptance and Effectiveness of Family Planning Methods*, report of an Expert Group meeting, Bangkok, 11–21 June 1968; Asian Population Studies Series, No. 4 (United Nations publication, Sales No. E.69.II.F.15).

TABLE 2. POSSIBLE POTENTIAL FERTILITY MEASURES FOR VARIOUS GROUPS OF WOMEN

Group of women	Time period: Pre-programme initiation ($t-x \ldots t-3 \quad t-2 \quad t-1$)	t_0	Post-programme initiation[a] ($t+1 \quad t+2 \quad t+3 \ldots t+x$)
All women aged 15-44 (49)	Observed or estimated		Observed or estimated
All married women aged 15-44 (49)	Observed or estimated		Observed or estimated
All fecund women aged 15-44 (49)	Observed or estimated		Observed or estimated
Acceptors: Method 1, Method 2, Method 3, . . . etc.	Observed or estimated		Estimated
Women by postpartum status	Observed or estimated		Estimated
Other special populations (ever-users, current users etc).	Observed or estimated		Estimated

[a] The time periods might also relate to intervals after the programme was initiated, such as from t_1 to t_3, and not merely to a period from the beginning of the programme to some years after, such as from t_0 to t_1, from t_0 to t_2 etc., as specified in the table.

(as shown in table 2). One view held that it was not of importance which group was used to derive a measure of potential fertility and that it was not necessary to take account of the higher fertility of acceptors, since that fertility quickly regressed to the mean. Although it might be appealing to estimate potential fertility of acceptors on the basis of their previous fertility, it could be viewed as a transient phenomenon and would not persist, as the regression to the mean was extremely strong. However, it was of interest to know the number of years required for that regression to the mean to occur. After all, it was only by chance that the women fell into two groups, i.e., whether they had high or low fertility. Furthermore, even the differences among various groups of women in continuation rates would tend to disappear after a certain period. Thus, it was unnecessary to calculate the potential fertility specific for various contraceptive methods. Any selectivity of women that operated at the beginning of a programme would tend to wash out and acceptors would become more homogeneous. At higher ages, however, the higher fertility of acceptors might be a problem; but because an increasing number would become sterile, that factor might not pose too serious a problem. If it were possible to use only fecund women, some of the major problems of selectivity might be overcome. That grouping, however, was difficult to arrange, inasmuch as some acceptors, even in a post-partum programme, were sterile when they accepted.

It was noteworthy, however, that it had not been shown conclusively that differentials in fertility and continuation rates, in fact, regressed to the mean, particularly in the short run. Rather large differentials in fertility as well as contraceptive practice continued to be observed in developing countries where contraception had long been practiced.

Simulation models might aid in answering some of the questions posed regarding the groups for which potential fertility should be calculated. Perhaps, in certain conditions, it would be more advantageous to use one group than another. Simulation would, perhaps, give some indication of which groups should be used. It was pointed out that, in the view of some researchers, the exact value of potential fertility was not essential and that a range of values within which potential fertility probably lay might be sufficient.

Opinions differed also regarding the necessity for introducing corrections for such factors as the overlap of contraceptive practice with the post-partum non-susceptible period, accidental pregnancies and the effect of acceptors in a programme merely substituting for other contraceptive practice that might have occurred anyway. On the one hand, considerable error could be introduced if corrections were not made for those factors. Imposing a penalty per accidental pregnancy could have a significant impact on calculated protection time when considerable emphasis was placed on conventional contraceptives, with irregular practice leading to many failures. Given a post-partum programme of contraception, especially in a context of lengthy lactation, it was critical to subtract from credited protection time the average period of overlap between practice of contraception and lactation amenorrhoea. Lowering the estimate of useful protection time in order to reflect the proportion sterile might be important for the older age-classes of acceptors (e.g., ages 35–39 and older). If mortality or divorce levels were high, it might be worth while to correct continuation rates for marital dissolution by death of a spouse and by divorce or separation.

On the other hand, a number of those corrections would not influence the results to any important extent, and the lack of data on which to base the correction factors was, in any case, a major hindrance. In addition, in many instances, relatively poor data were being subjected to rather refined corrections and thus such corrections were not warranted. Only when there was good evidence that a particular factor, such as a particularly long post-partum period resulting from lactation, was operating, should certain correction factors be introduced.

A number of other factors could be identified as affecting the measurement of potential fertility. The effect of many of those factors would be to produce increases in the level of fertility while the effect of other factors would be to produce declines in fertility. Accordingly, the operation of some factors would offset the effect of other factors, and, therefore, no change in fertility levels would be observed. In either instance, the estimation of potential fertility could be particularly difficult.

If the birth rate remained stable as a result of a number of offsetting factors, such as changes in population structure, the impact on fertility of a family planning programme might not be discerned. That situation would be particularly true if the effect of a number of factors was in the direction of increasing fertility while the effect of the family planning programme was to offset the potential increases in fertility, with the result that no change in the birth rate would be observed. In such a circumstance, to conclude that the family planning programme was having no effect would be erroneous.

Among the factors that could have a positive effect on fertility were the removal of the traditional checks on fertility, changes in socio-economic conditions, changes in mortality and migration, and changes in marital patterns.

The weakening of traditional checks might take such forms as decreases in the practice of breast-feeding or a shortening of its length, the rèmoval of taboos on intercourse during certain periods and the decline in prolonged visits by the wife to her mother's home. Such factors might have contributed to a rise in potential fertility in one state of India. In urban areas of many developing countries, the decline in breast-feeding practices might be especially significant. In general, however, the effect of changes in those factors was considered to be highly conjectural, as very few studies had been carried out that clearly demonstrated their effects.

Improvements in socio-economic conditions, which had an over-all effect on levels of living, could act differentially in increasing or decreasing fertility. For example, improvements in nutrition were believed to increase fertility, especially when previous levels of nutrition had been very low.

Changes in mortality and migration also played some role in affecting the level of fertility. Declines in infant mortality were not believed to have much effect on decreasing fertility in the short run,[8] but decreases in the mortality of spouses had the effect of increasing marital durations and might therefore contribute to rises in fertility. On the other hand, increased migration that involved separation of husbands and wives for relatively long periods could have a depressing effect on fertility.

D. Correlated variables and interaction; uncontrolled variables and independence of method

The Expert Group recognized that the effects of a family planning programme on fertility must be viewed as a complex system of many interconnected parts, a point noted explicitly in the statements by Ebanks[9] and Brass.[10] Both writers stated that the evaluation of family planning programmes should be viewed in the larger context of studying the determinants of fertility.

The problem then was to identify the boundaries of the system, its component parts and the linkages among these components. Stated in other terms, that task was in the realm of theory building or model construction. Such efforts must involve the identification of the key concepts or variables and the form of their relationships. The Expert Group noted that the papers by Sirageldin,[11] Wells[12] and Srinivasan[13] took up that problem in some detail.

The need existed for a conceptual model of the underlying process regardless of the analytical technique used—whether it was regression or simple tabulations. Also, problems of data availability should be divorced from those of model construction. If the needed data were not available, substitutes should be devised or the model modified. In that way, it would be possible explicitly to recognize what variables were not included.

The Expert Group recognized that a theory was an arbitrary construction, and it was best viewed not as true or false, but as fruitful or less fruitful. Theories sought to explain known relationships; if possible, they did so in a parsimonious manner, and they led to new observations. Theories might be developed by beginning with a few basic concepts and adding factors

---

[8] See Committee for International Co-ordination of National Research in Demography, *Seminar on Infant Mortality in Relation to the Level of Fertility* (Paris, 1975).

[9] G. E. Ebanks, *op. cit.*

[10] William Brass, "Comments on comparison strategies for the evaluation of family planning impact" (Conference room paper 14).

[11] Ismail Sirageldin, "Some issues in determining appropriate methods of evaluating the fertility impact of family planning programmes" (Conference room paper 4).

[12] H. B. Wells, *op. cit.*

[13] K. Srinivasan, "Interaction of socio-economic changes with family planning programmes: an assessment model" (Conference room paper 7).

found to be important—an incremental strategy. On other occasions, one might have a large number of factors and proceed by removing those found not to be salient in explaining the phenomena under consideration. Although it was desired to avoid bias, a strategy of including every possible variable was often not wise; the introduction of many variables might lead to other statistical artifacts equally serious.

Given that a complex system was involved, there was also the likelihood of correlated variables, i.e., effects produced by two or more factors which were highly interrelated. The implication of that possibility was that one could not hope to partition out the variance of a dependent variable explained uniquely by those "independent" or causal factors. Some portion of the variance would be due to the compounded effect of two or more factors. Though the nature of that limitation was most clearly formulated and quantified in terms of regression analysis, it was a common limitation running through all the methods and it arose from the complexity of the system under study. For example, in standardization, when the difference between two rates was decomposed, it was often found that some portion was due jointly, say, to age and marital status, and could not be attributed uniquely to either factor. The correlated variable problem might be more severe with residual techniques, since there was clear omission of variables known to be important and which might be correlated with those included. Unlike some of the other methods, the regression model at least attempted to include the family planning inputs explicitly.

Interaction effects could be seen as a function of the form in which the relationships were expressed, to use the terminology of regression analysis. If a linear additive model applied, it was possible to say how much change in the dependent variable was associated with a unit change in an independent variable, taking into account the other independent variables. If there was interaction, so that the effect of variable A depended upon the level of variable B, then that deduction is no longer possible. Although interaction was often suspected, it was worth while to test the adequacy of the simpler model (i.e., without interaction) to gain some of the benefit of the unambiguous allocation of effects.

Further, it might be desired to distinguish operational interaction—where a programme varied the type of input by the characteristics of the area, e.g., urban-rural distribution or income levels—from statistical interaction. With regard to the latter factor, interaction might not occur over part of the range of the variables involved, but might appear at another part of the distribution.

The uncontrolled variable problem arose when one or more factors that affected the dependent variable were omitted from the model. The penalty depended upon the importance of the omitted variable to the dependent variable and whether it was correlated with the independent variables included in the model. When a variable was omitted, there occurred a loss in explanatory power that was proportional to the importance of that variable. If that variable was not correlated with the variables in the model, then the effects attributed to the included variables were not biased or misleading; but if there was correlation with the omitted variable, then the measured effects would be biased.

It was desirable to explore ways of determining the degree of independence of methods. Just as the variables in any one model were correlated, there was likely to be correlation among the methods of evaluation, arising from common data, assumptions etc. It might be useful to envisage an observation arising from the application of a method as due in part to what was being measured, in part to random error and in part to method variance, i.e., systematic error attributable to the data and procedures employed. If a number of observations were obtained from different methods applied over a sufficiently large number of settings, it might be possible to decompose the relative magnitude of each of those factors attributable to each method. That point was discussed in detail in the statement by Hermalin.[14]

Many of the problems of correlated variables, interaction and uncontrolled variables were discussed by the Expert Group in the context of regression techniques. One current problem with areal regression models was that often it was not known what variables should be included, or it was unclear what the variables included represented. In addition, the results appeared to be very sensitive to the inclusion or omission of particular variables. The results, therefore, were very difficult to interpret.

As one solution, the problem of proxy and surrogate variables might be alleviated by including a number of indicators of an underlying concept in such a way that they would be recognized as indicators. Developments with "unobserved variable" models and confirmatory factor analysis were also suggested as promising in that respect.

An advantage of regression might be that the method bypassed the problem of potential fertility and births averted, in that it was not necessary to develop measures of these concepts explicitly. But that viewpoint did not find consensus among researchers, as some would maintain that it was a disadvantage and that measures of births averted provided more validity than did the coefficients obtained with regression techniques.

The method used depended upon the purpose of the evaluation; if the purpose was to understand the relative importance of family planning vis-á-vis development, the appropriate procedure was regression analysis. In contrast, if the aim was to trace out the effects of a regimen of contraceptive use, births-averted analyses making detailed use of data on various aspects of the reproductive process were prefera-

[14] Albert I. Hermalin, "Avoiding an embarrassment of riches" (Conference room paper 10).

ble. However, it might be possible to study the effects of development on certain aspects of the reproductive process through regression techniques.

It was pointed out that several difficulties arose with the regression approach. Areal regression typically did not separate programme acceptance from non-programme acceptance, but sought to establish the over-all effect of the programme in comparison with other factors that were operating. However, data on non-programme acceptance could be utilized in a regression model. Sometimes by partitioning the areas at the beginning of the programme by different levels of fertility, it was possible to achieve controls for non-programme contraceptive use and implicitly for potential fertility.

The allocation of effects in temporal terms was very difficult in regression analysis, even if one worked with a model containing various lags. If there were strong, interareal correlations among the areal structural variables, then the results would not be too sensitive to the particular lags used. But that fact did not make temporal allocation any easier. Methods that made use of a series of cross-sectional data might be a means for improving the understanding of the time dimension.

Multiple regression was a form of multiple comparison and one should contemplate other types of multiple comparison, which could get at more detailed aspects of the process. It was possible, for example, to set up multiple comparisons from individual data and to study the effect of non-programme factors, such as breast-feeding. That procedure could serve to validate some of the results for areal regression and also to answer additional questions. For example, some aspects of development might increase fertility as well as reduce it; different educational groups might react differently to a family planning programme. Because it was difficult to observe such differentials through regression analyses, there was a need to give more attention to the reliability of regression results, e.g., by splitting the sample in half or by other means.

Areal regression results were likely to be sensitive to the number of observations and the particular measures used, as well as to a number of other factors. Thus, it was important to replicate results with different models and different measures, as well as to validate the results with alternative techniques. The results of a regression analysis should be compared with the results of other analytical methods, such as controlled cross tabulations. Cross tabulations made less severe demands on the nature and form of data. Attention should also be given to studying the effect on the results of areal regression at different levels of aggregation. Distinctions between categorical and interval data usually were not as important as they had been. Recent developments permitted the testing of linear models for categorical data.

In so far as areal regression was hampered by lack of data, improvements could come through the use of the community-level data obtained in surveys. Also, it should be possible to aggregate the individual level data of surveys and to form areal measures that would not ordinarily be available from the usual statistical sources.