

## I. THE METHODS

At the outset of the discussion of the eight methods commonly applied in the evaluation of the effect of family planning programmes on fertility, it was made clear that evaluation could usefully be undertaken only in light of what precisely was being measured and the time to which the measurement related.

To specify what was being measured was not a simple matter, for family planning programmes generated both direct and indirect effects; and those effects were not easily differentiated, nor was there a satisfactory method of separating the effects of a family planning programme from those of socio-economic development. A direct programme effect would be births delayed or forgone by a group of women who had become acceptors of a contraceptive method in a programme. An indirect programme effect would be births delayed or forgone by women who had adopted some contraceptive method obtained from a source other than the programme, but who had been influenced by the programme to do so. An effect of socio-economic development might be a reduction of fertility traceable to increases in the employment of women outside the home in non-familial activities or to improvements in the health and nutrition of women.

The relevance of time involved two perspectives: how long a programme must run before the measurement of impact was feasible; and whether an observed impact was of short- or long-term significance. Long-run impact was viewed as an effect on the completed fertility of women currently in the reproductive ages. Short-term impact might represent changes in period rates, due to spacing of births, with or without change in completed family size.

As stated above, the principal focus was on the methodological issues and the problems encountered in the application of the methods, along with certain advantages and disadvantages of the specific methods.

### A. STANDARDIZATION APPROACH

In the standardization approach, an attempt was made to answer the question whether there had been a real change in the fertility of the population, by determining changes that were due to shifts in age structure and marital status. Use of that method indicated the proportion of change attributable to structural factors (e.g., age structure and marital status), and the proportion attributable to fertility behaviour. The method did not, however, separate the effects of a family planning programme from those of socio-economic factors, due to the interaction of the structural factors with the development variables.

The standardization approach was an essential first step in the measurement of the effect on fertility of family planning programmes, and it should be routinely applied in all evaluations. It might also be used to study factors that had affected trends in fertility prior to the introduction of a family planning programme. However, because the method produced only a residual after variables had been standardized for, it could not be used to determine how much of the change in fertility had resulted from the effects of the family programme and how much from socio-economic changes. Further, if the residual was, in fact, taken as a measure of the impact of the family planning programme, an over-estimate of impact would be likely. The results of the application of the standardization procedure in the case studies presented as background papers for the meeting indicated that in no case could all of the observed decline in fertility have been accounted for by the variables included in the standardizations, but that the decrease could be attributed to the effects of the family planning programme as well as to factors that had not been taken into account. In summary, standardization could only indicate that the programme might have had some impact.

The choice of variables to be controlled by standardization was a crucial step in application of the method. Age and marital status were considered essential, and rural-urban residence and socio-economic status indicators added refinements that provided invaluable information to programme administrators. However, the choice of variables was in part a theoretical problem, in that the determinants of fertility must be identified. Consequently, the variables must be selected according to hypotheses concerning their relation to fertility. It was noted that the statements by Ebanks<sup>1</sup> and Wells<sup>2</sup> deal directly with that problem.

One of the greatest problems encountered in using the standardization approach was the availability of data. In each of the three country case studies provided for the Expert Group meeting, standardization was carried out for age and marital status. The case study for Karnataka State, India, also standardized for urban-rural residence, while the Chilean study included area of residence, educational level and female labour force participation. The overriding role of availability of data as a determinant of evaluation procedures was seen in the fact that, in each study, the

<sup>1</sup> G. Edward Ebanks, "Needed data and research on the impact of family planning programmes on fertility" (Conference room paper 8).

<sup>2</sup> H. Bradley Wells, "Notes on causal relationships in measuring fertility change" (Conference room paper 9).

inclusion or exclusion of variables depended upon what statistics were available.

Data availability might prove a crucial factor if it were desired to undertake a more sophisticated application of standardization. For instance, if the intent was to standardize several factors simultaneously, the unavailability of required data could be a major hindrance. Interaction between various factors might be viewed as a problem. For example, the unobserved variables might influence those employed in the standardization, while, at the same time, the variables used in the standardization might interact with one another; and it was exceedingly difficult to determine their separate effects. In either case, it would be difficult to ferret out the amount of fertility change attributable, for example, only to changes in proportions married, if women had also advanced in education or had engaged increasingly in economic activities. For, improvements in education had tended to raise age at marriage and, consequently, to lower the average duration of marriages, while equipping women for non-familial activities.

#### B. TREND ANALYSIS

The Expert Group considered that trend analysis, the fertility projection technique, permitted some indication as to what might have happened without a family planning programme or in the absence of social, economic and demographic changes. Moreover, with that method account could be taken of changes in fertility levels which had occurred prior to the initiation of a family planning programme. For instance, it was possible to extrapolate into the future declines in fertility observed before a programme was begun. Consequently, the method indicated whether a change had taken place beyond what would be expected on the basis of past experience. But the results derived by the projection approach rested upon a very gross assumption, namely, that future fertility would be equivalent to what would have happened if past trends of fertility had continued. The impact of the programme was seen as the difference between projected fertility and recorded fertility.

The method required data on past trends in fertility and good observations on the trends in fertility following inauguration of a programme. But the results derived from it were even more uncertain than those of the standardization technique, which at least permitted the identification of some factors that had contributed to an observed change. Mention was made of the fact that there were a number of projection techniques, each of which differed somewhat in respect of the data required for application. For example, the statistics needed for the projection technique could be quite formidable if the number of components employed was large, as the requirements might include data on sex, age, marital status, residence, education and other characteristics over a considerable period. But owing to the great variety of problems encountered and the

means of coping with them that were applied in each instance, the results could not be viewed as leading to any firm conclusions as to the efficaciousness of the method for gauging the impact of a programme.

#### C. EXPERIMENTAL DESIGN

The extent to which truly classical experimental design could be carried out was considered by the Expert Group to be the major issue involved in the application of that method. The opinion was that truly classical experimental design could be approached, but rarely achieved in family planning evaluation. However, that method was a powerful scientific tool, setting a standard against which all other methods must be judged.

In classical experimental design, the researcher selected randomly a member of the control group and one of the experimental group. But in evaluation, it was all but impossible randomly to allocate individuals or areas to experimental and control groups, and the problem would be different for areas and individuals. A drawback to application of experimental design was that it was implemented over a considerable period of time and thus was not suited to be a measure of programme impact after short intervals.

The difficulties of applying the method were considered virtually insurmountable, but there were various ways in which the classical experimental design might be approached. Matching studies were one means of affecting the more rigorous experimental design method. One variation would be to use individuals or areas matched on a number of basic characteristics. Retrospective data for the matching of the groups would be collected and the groups established at some point prior to a survey. Follow-up surveys would then produce the prospective data whereby changes could be assessed.

More generally, the full potential of surveys had not been realized in terms of analyzing the data with a view to applying the experimental design as the model. Further, the use of stratified sampling in survey design could assist greatly in achieving the basis for more rigorous analysis. Regression techniques also were important analytical tools that would permit an approach to the experimental design by taking advantage of the fact that family planning programmes were usually begun in phases, e.g., only some parts of a country were included in a programme when it was initiated and the programme was gradually extended to other parts of the country. If areas of a country could be randomly assigned prior to the time that the family planning programme was initiated, using as criteria the projected time when the programme would be phased in, fairly good measures of the impact of a programme after the elapse of a given time period could be obtained. Thus, something approaching a classical experimental design could be achieved.

An indication of the difficulties of carrying out experimentally designed studies was provided by the applications of the method in case studies of India and

Tunisia, which approached the method by utilizing available survey data. Both applications were considered good examples of what could be achieved with survey data. The utilization of the method in the Indian case study might be regarded as an example of the use of careful matching on a number of important reproductive characteristics. Both applications illustrated the problems of interpreting the results.

#### D. COUPLE-YEARS OF PROTECTION

The central issues in regard to couple-years of protection (CYP) was what it purported to measure. It was originally conceptualized as a prevalence index but became widely used, even by those who had developed it, as an index of births averted and, thus, as a method of assessing family planning programme impact. It was the consensus of the Expert Group that if that method was used for the assessment of the impact of a family planning programme on fertility, the formula should be modified to give a more precise measure both of couple-years of protection and of births averted.

Among the advantages of the CYP method was its relative simplicity in terms both of data requirements and of calculation. If data were available, many refinements might be introduced, such as a CYP measure specific for age, residence and education.

Nevertheless, since such refinements would require additional data, it was felt that a major drawback existed. If the measure was to be refined in a meaningful way, fairly detailed data on continuation rates (derived from a follow-up of acceptors) would be needed, and such data were not usually available. Further, the calculations were made, as a rule, for a calendar year, and realistic assumptions regarding continuation rates of contraceptive use on a calendar-year basis were particularly difficult to obtain.

Although, as noted above, the CYP method was capable of refinement, the additional calculations would require statistics cross-classifying women by fertility and by acceptance of specific contraceptive methods according to age, residence, marital status etc. Those data were not usually available, and without such refinements, interpretation of results of the method was very difficult. Further, lacking refinements based on reliable data, a number of gross assumptions must be made. Owing to the rough nature of the method, it might yield unreliable estimates of programme impact.

Each of the case studies applied the CYP method. A fairly detailed appraisal of the method was given in the Tunisian study. In the Indian case study, the method was calculated on only one set of assumptions regarding continuation rates although two different population distributions were used. But it was considered noteworthy that similar results had been obtained from applications of the component projection approach and the CYP method.

#### E. COMPONENT PROJECTION APPROACH

The major issues raised concerning the component projection approach appeared to be its sensitivity and robustness. It was felt to be conceptually sound, workable and easy to apply, and was considered to hold the best potential for single-method evaluation.

(b) Mention was made of the following advantages: (a) age was taken into account; (b) the estimation of births averted for acceptors was possible; (c) it could be used for target-setting by providing estimates of the number of acceptors needed to achieve a specified reduction in the number of births; (d) it permitted the disaggregation of the population into components and, consequently, estimates might be obtained for important subgroups of the population (e.g., rural-urban, educational groups); (e) it also could be used to estimate the impact of a programme in the future under varying assumptions.

But the component projection approach also posed significant disadvantages, including the necessity of estimating potential fertility. The method also required continuation rates and, hence, a good follow-up of acceptors.

Although the method was thought to be conceptually good, many of the applications had merely been simplified versions of it, owing mainly to the various simplifying assumptions required in the absence of concrete data. Ideally, programme administrators should plan towards the use of the method by collecting the basic data at the time the programme was begun and periodically thereafter, for it was the lack of such bench-mark data that necessitated resort to a number of arbitrary assumptions.

The Expert Group observed that the statement by Potter,<sup>3</sup> which summarized the recent developments in the use of the method, pointed out that the component projection approach fulfilled a variety of evaluation needs.

The difficulties of obtaining the requisite data for that method were emphasized in each of the case studies. Estimation of potential fertility was also considered a major problem in those studies. For example, it had become clearly evident from those studies that because no clear guidelines existed as to what data or group of women the estimate should be based upon, the quality of evaluation depended on the skill and judgement of the evaluator; there were no standards for assessing those qualities.

#### F. ANALYSIS OF THE REPRODUCTIVE PROCESS

Analysis of the reproductive process had not been widely used, owing mainly to the many detailed refinements incorporated into the approach. And it was considered that those refinements constituted the principal

<sup>3</sup> Robert G. Potter, "Component projection *versus* other techniques for assessing programme achievement towards a targeted fertility reduction" (Conference room paper 6).

issue with respect to the method. Questions were raised concerning what the conditions were that made the extra refinements incorporated into that approach really essential and when those additional adjustments were expendable. The refinements consisted of subtractions from woman-years of protection that allowed for secondary sterility, overlap of contraceptive practice with amenorrhoea and failures of contraception causing accidental pregnancies. In addition, alternative estimates of potential fertility were provided in order to represent a range of substitution effects.

The method provided a highly accurate estimate of women-years of protection and of potential fertility of acceptors, which were its major strengths; and it had proved a fruitful research tool in that it focused research on a number of significant factors affecting evaluation.

The Expert Group was of the opinion that a basic disadvantage of the approach was its extensive data requirements. Also, the assumptions for one population with respect to secondary sterility, post-partum amenorrhoea and time to conception in the absence of contraception, derived in Potter's application,<sup>4</sup> were not necessarily valid for other populations. Another problem was that, in the Potter application, the approach was designed to estimate births averted per insertion of an intra-uterine device (IUD). Some modification was necessary for application in respect of other contraceptive methods.

Analysis of the reproductive process as a method of evaluation, as formulated by Wolfers,<sup>5</sup> had been designed for evaluation of a post-partum family planning programme. Its application to other types of family planning programmes would necessarily require modification.

Theoretically, data permitting, the procedures of the analysis of the reproductive process approach could be incorporated into the component projection approach in order to measure family planning impact in a period population, though that merger had not yet been formally carried out.

Significantly, analysis of the reproductive process was the only method not applied in any of the three case studies, and the reason given in each case was unavailability of the requisite data. In light of the considerable problems of applying it, as it was currently

formulated, that highly specialized method had limited applicability to general evaluation efforts and research was needed to facilitate its wider application.

## G. REGRESSION ANALYSIS

The Expert Group felt that regression analysis depended upon a conceptual model, and, therefore, some of the more important issues in regard to its application related to the variables that should be included. The method posed a number of statistical and theoretical problems, none of which could be easily resolved. Regression analysis permitted estimation of the relative impact of the programme, and it could be used with both macro and micro data, although in evaluation studies macro applications were the more common.

Among its disadvantages, regression analysis required adequate measures not only of fertility but of non-programme and programme variables. Because the choice of variables to be included was so crucial, a fairly detailed conceptualization was required. Most regression analyses dealt only with different moments in time, but the method could also be applied to time series data and thus could take account of time lags.

The attempted application of the method to Chilean data in the case study prepared for background material was cited as a good example of the difficulties encountered in applying it. In practice, there was not much choice in terms of variables to be included; the difficulty arose with respect to availability of data on those variables.

Additional problems confronted the researcher when the method was applied to areal units; a range of areal units was required, with variability among the units, for the method would break down if there was a lack of variance.

It was mentioned that in applying regression analysis, variation within the units must be accounted for, which was rarely done. The fact that the method required a fairly high level of statistical sophistication could lead to its misapplication.

## H. SIMULATION MODELS

The Expert Group felt that some question existed about the usefulness of simulation models for estimating the impact of family planning programmes on fertility, and the cost was considered to be high in relation to that of other methods. Simulation had an important role in research on evaluation, for simulation models could be useful in the study of factors that were not readily observable. Some of the basic biological factors were not amenable to direct observation and, consequently, simulation could aid in the study of such variables as age at sterility. One of its important uses had been—and was likely to continue to be—to indicate the role of chance. Simulation could also be used to estimate potential fertility, and the method could aid in testing estimating procedures under different condi-

<sup>4</sup> For a description of the method, see Robert G. Potter, A technical appendix on procedures used in manuscript "Estimating births averted in a family planning program", paper prepared for Major Ceremony V University of Michigan Sesquicentennial Celebration, 1 June 1967. See also Robert G. Potter, "Application of life-table techniques to measurement of contraceptive effectiveness", *Demography*, vol. 3, No. 2 (1966), pp. 297-304; and *idem*, "Estimating births averted in a family planning program", in S. J. Behrman, Leslie Corsa, Jr., and Ronald Freedman, eds., *Fertility and Family Planning: A World View* (Ann Arbor, Mich., University of Michigan Press, 1969), pp. 413-434.

<sup>5</sup> David Wolfers, "The demographic effect of a contraceptive programme", *Population Studies*, vol. XXIII, No. 1 (March 1969), pp. 111-141.

tions. It could be a powerful methodological tool for validating different methods of measuring the impact of programmes on fertility, and it should also be useful for studying such topics as measurement error, response error and sample size. Simulation models permitted the carrying-out of experiments and might in that way provide insights into the implications of various strategies in family planning programmes. Clearly, simulation was primarily a research tool, and it should be treated as such.

In the opinion of the Expert Group, a disadvantage of the models was that they required a great deal of data, some of which were practically non-existent. Moreover, many of the more sophisticated models (i.e., micro models, using the Monte Carlo method) required large-scale computers which, as a rule, were relatively expensive to utilize. In addition, the devel-

opment of such models was highly expensive; and, for that and other reasons, the models currently available should be more fully exploited before any new models were developed. Because the structure of many of the models was relatively complex, potential users must devote a great deal of time to understanding the definitions or assumptions built into them. It was pointed out that not only was the development of simulation models costly, but the costs of running experiments with them were high.

Only one of the case studies, that of Karnataka State, India, included an application of a simulation model. That model, it was noted, had already been developed and thus was available for use in the case study. The results obtained with the model indicated the possibility that the CYP method over-estimated the number of births averted.