INTRODUCTION*

*United Nations Secretariat**

The second half of the twentieth century has witnessed a marked increase in government use of policy instruments to effect demographic, social and economic change. And this phenomenon has become more widespread in both developed and developing countries. Along with family and child allowances, special grants and conveniences for mothers of newly born children; and proscriptions and provisions relating to abortion, sterilization and the supply and distribution of contraceptives, the national family planning programme has become widely used as a measure directed towards influencing human reproductive behaviour. It is typical, however, of policies designed to influence fertility and other demographic and social phenomena that researchers have had only from meagre to moderate success in measuring the extent to which the policy measures have achieved the intended objectives.

Over the past two decades, major population policy efforts in many developing countries in particular have been centred on large-scale family planning programmes. It is important, consequently, that the tools for evaluating the effect of these programmes should be honed to the point where they can be used with reasonable facility and efficiency and with maximum confidence as to the validity of the results that they produce.

A. Types of evaluation

The national family planning programme is a wide assortment of activities geared towards an ultimate objective. But achievement of the final goal depends necessarily upon accomplishments at different levels or, in other words, upon attainment of a variety of subobjectives. Each of the latter goals can be subjected to evaluation; and, as good administrative practice, this task is done periodically as a guide in operational strategy. Thus, evaluation is done in terms of both intermediate and ultimate objectives, as well as in light of plausible outcomes of the programme which are not specified as goals:[1]

(*a*) *Intermediate objectives.* A number of intermediate goals can be subjected to evaluation, including efficiency of workers, periodic achievements of workers, recruitment of acceptors, programme logistics and yield from communication efforts;

(*b*) *Long-term or ultimate objectives.* Such objectives may be, among other things, reduction of infant and maternal mortality and improvements in the general well-being of mothers and children; reductions in subfecundity or infertility; reductions in induced abortions; a decline in fertility; or achievement of a specified rate of population growth.

Relatively little has been done to ascertain, reliably, what the planning programmes established for various health-related purposes have accomplished. Conversely, much effort has been devoted to perfecting methodologies for assessing the impact of the programmes on fertility. Further, most programmes of as long as three years duration have been evaluated for fertility impact, even where the programmes did not have fertility decline as an objective.

B. Purpose of the manual

This *Manual* is intended as a guide by which middle-level professionals may assess the impact of family planning programmes on fertility. Its further purposes are to assist programme directors in improving management through better evaluation; to present in a single volume a complete methodological statement on the currently used evaluation tools, so that individuals engaged in family planning programme evaluation research may have a handy reference; and to provide material for teaching and training courses in the methodology commonly applied in this field. The *Manual* offers illustrations and principles designed to assist researchers who have reasonable competence in mathematics and statistics to apply the accepted methods of evaluation and offers suggestions as to how the results of evaluation obtained by each method should be interpreted. The *Manual* is to serve as an aid in the use of the following methods that have been advanced by various scholars as being appropriate tools for assessing family planning programme impact on fertility:

(*a*) Standardization approach;

(*b*) Standard couple-years of protection (SCYP);

(*c*) Component projection approach:

(i) Computerized model;

(ii) Model for desk calculators;

* Population Division, Department of International Economic and Social Affairs.

[1] For a more detailed discussion, see W. Parker Mauldin and Gwendolyn Johnson-Acsádi, "Introduction", in C. Chandrasekaran and Albert I. Hermalin, eds., *Measuring the Effect of Family Planning Programs on Fertility*, published by the International Union for the Scientific Study of Population for the Development Centre of the Organisation for Economic Co-operation and Development (Dolhain, Belgium, Ordina Editions, 1975), pp. 1–16.

(*d*) Analysis of reproductive process;

(*e*) Multivariate areal analysis;

(*f*) Simulation;

(*g*) Experimental designs;

(*h*) Fertility projection/trend analysis (illustrations of this method are given in annexes I and II).

C. Issues of evaluation

It is established that fertility trends are influenced by a large assortment of demographic, social, cultural and economic factors, and that determining the contribution of any one of those factors to fertility change is an exceedingly difficult task. Yet it is precisely this task that the evaluator seeks to accomplish when undertaking evaluation. If there has been a change in the level of fertility, the task is to determine what portion may be ascribed to family planning programme activities. Or, if no change is evident, it is necessary to ascertain what the level of fertility would have been had there been no family planning programme, because the stability may be due to counter-effects of lower fertility among some women due to the programme and rises in fertility among others as a result of certain modernization factors. In working through either or both of these problems with the methodologies now in use, a number of methodological issues must be dealt with, and there is no real consensus as to how this task should be done.[2]

These issues, upon which hinge evaluation studies, are: (*a*) the estimation of potential fertility; (*b*) correlated variables and interaction; (*c*) uncontrolled variables; and (*d*) independence of method. They represent the recognized limitations of evaluation methodology, and some of them are encountered no matter what method of evaluation is being applied. Evaluators must have a thorough understanding of these issues and must take them into account in applying the methods and in interpreting the results.

Except for the issue of independence of method which, when replication by different methods is sought to determine validity of results, is also one of the criteria for choice of method, these issues are not discussed here, owing to their complex nature and to the fact that they must be treated exhaustively or not at all. Such a treatment is aptly accomplished in the two relatively recent publications that the reader is urged to consult for a thorough discussion of the issues and problems of family planning programme evaluation.[3]

D. The method to choose

This section contains a brief discussion of some of the considerations that enter into the choice of a method or evaluation approach. It is not intended to be exhaustive but to alert the researcher to the problem of choosing an evaluation approach in light of the methodology currently used. A more analytical treatment of this subject is offered in a recent article which examines the circumstances in which it is suitable to apply certain categories of methods.[4] The reader might find it also helpful to examine the matter of method choice within the framework provided in that source.

Researchers developed the several methods currently used to determine how much effect a family planning programme has had on fertility within a given segment of time evidently in response to, among other things, the needs of administrators for the information, the type and amount of useful data at hand, the degree of technical expertise that could be marshalled for the purpose and the urgency attached to finding an answer. Although some of the methods were devised as much as a decade and a half ago, in large measure, data availability and amount of technical expertise or familiarity with a method are still among the principal reasons that a researcher chooses a particular method as the tool for evaluation. Clearly, knowledge of family planning programme evaluation techniques should be more widely disseminated. Only a slight introduction to the structure of each evaluation technique or method is sufficient to disclose that they are essentially dissimilar, being based on different hypotheses, assumptions, data and base populations. Thus, a choice among them should involve more scientific considerations than, for example, familiarity with the method.

The question concerning the method that should be employed is faced by any evaluator who has the minimum data required and the technical competence to execute more than one method.[5] But even if all conditions for utilizing all of these methods were fully met, selection of an evaluation approach would not be a simple matter. First, there is a lack of agreement among scholars as to the relative efficiency of the various approaches;[6] each approach has certain advantages and disadvantages. Further, some methods are applicable in some circumstances and not in others. And there is no consensus as to when each is the method to be preferred or as to the meaning and relative validity of the results obtained.

---

[2] Excellent guidance is found in C. Chandrasekaran and A. I. Hermalin, eds., *op. cit.*; and in *Methods of Measuring the Impact of Family Planning Programmes on Fertility: Problems and Issues* (United Nations publication, Sales No. E. 78. XIII. 2), especially pp. 3–42 and 137–161.

[3] *Ibid.* These sources contain extensive discussions of these methodological issues.

[4] Jacqueline D. Forrest and John A. Ross, "Fertility effects of a family planning programme: a methodological review", *Social Biology* (in press).

[5] The United Nations Expert Group Meeting on Methods of Measuring the Impact of Family Planning Programmes on Fertility, which was convened at Geneva in April 1976, found that the factual basis does not exist for choosing a method of evaluation according to strict scientific criteria. It accorded highest priority to research directed to developing knowledge in the sphere. It took note, however, that in the selection of a method certain criteria could be applied. See "Report of the Expert Group Meeting on Methods of Measuring the Impact of Family Planning Programmes on Fertility", in *Methods of Measuring the Impact of Family Planning Programmes on Fertility: Problems and Issues* (United Nations publication, Sales No. E.78.XIII.2), pp. 137–161.

[6] The United Nations is attempting to alleviate this problem by sponsoring a number of national case studies in which several or all of the measures are applied to a single set of data with the view to ascertaining the relative yield (in terms of birth rate decline due to the programme) of each method applied. For information on the first three case studies, see *Methods of Measuring the Impact of Family Planning Programmes on Fertility: Problems and Issues*, pp. 43–136.

Some programmes will have been in existence for only a few years and others much longer; some will have been implemented in a relatively stable society and others in a climate of social, economic and cultural upheaval. These conditions influence the method of choice. The researcher should therefore, to the extent possible, become thoroughly familiar with all methods before making a choice, and the text of this *Manual* is designed to facilitate this undertaking. The following discussion is not intended to serve as a guide to choosing a method, but to point out some of the reasons why the researcher might choose or reject a particular approach.

In addition to such considerations as the information needed and the data available and technical feasibility of application, some of the principle questions determining the choice of methods are: (*a*) what it is desired to measure; (*b*) the population for which evaluation is needed; (*c*) whether it can separate programme from non-programme effects; (*d*) the time period to which the measurement relates; and (*e*) the independence and reliability of the method. It is generally recognized as a short-coming of the methodology of evaluating programme impact on fertility that it is difficult to determine the degree to which the reproductive behaviour of persons who do not accept a method from the programme is none the less influenced by it is not an easy matter. Hence, there is concern about what a method actually measures.[7] Further, in certain circumstances the aim of evaluation is to determine births averted for those who accepted a method from the programme; in other cases the intent may be to determine births averted, i.e., programme impact on fertility, in the entire population as a whole. And it is difficult to know whether and to what extent the change in fertility has occurred in response to the programme or is a result of changes in social, demographic and economic conditions. Thus, when this question is of interest, the method should be capable of separating these effects.

Because time can be seen either as a short- or a long-run element with differing consequences, it is crucial in evaluation. First, there is the question how long a programme must run before its impact can be measured. Then the researcher must decide whether the interest is in short-term effects, which might be deduced from changes in period rates brought on by birth spacing, or in long-term effects suggested by the completed fertility of women 45 or 50 years of age. Of course, if the programme has been in effect for a sufficient period of time, the researcher would in most cases wish to measure both its short-term and its long-term effects.

The current state of knowledge about evaluation is not such that the results of evaluation exercises can be accepted without question. This situation is due to a variety of problems and issues which are dealt with in the chapters that illustrate the methods. Confidence in the product derived by application of a specific method is enhanced if there is replication of results when one or more additional evaluation approaches are used. Consideration of a method to be applied in establishing replication is whether it is independent of the other methods used. Reliability of results obtained by a method may also be established in other ways, as will be observed from material provided in the succeeding chapters.

1. *Objective of the evaluation*

The evaluator may be interested in determining either the impact of the programme on acceptors' fertility or in the total effect of the programme. Fertility projection/trend analysis is a means of assessing total programme effect, for it attributes a close association between projected fertility and the movement of measures of actual fertility to the effects of the family planning programme; if no association were evident, it would be assumed that the programme had not had an effect.

Standardization, which, generally speaking, should be a first step in evaluation, is capable of establishing how much of the change in fertility was due to demographic factors of age structure and marital status, and thus also to changes in marital fertility. The latter result would indicate whether further analysis was warranted; and if so, the researcher would then choose a method to determine how much of the change in marital fertility could be attributed to the programme.

Methods that deal with acceptors only, such as experimental design, are an aid to assessing the direct impact of the programme. The indirect effect, i.e., the impact of the programme on the fertility of non-acceptors, may be determined by a number of approaches. A practical method is to obtain a measure of total programme effect and to calculate the indirect effect as the difference between that measure and the direct effect obtained by use of a method dealing with acceptors only. A note of caution is in order, however, as the approaches should be independent of each other and conflicting assumptions should be avoided.

2. *Period covered*

If the evaluation is to cover only a short period, say, less than one year, none of the evaluation approaches that has fertility as a variable would suffice. Instead, it would be required to apply some yardstick such as numbers of acceptors, with appropriate assumptions as to continuation rates and effectiveness of the contraceptive methods. The couple-years of protection (CYP) method,[8] though flawed in methodology, has been indicated as a method suited for short-period evaluation.[9] Because it is very crude, that method is not illustrated in this volume; instead, the "standard" couple-years of protection approach is offered.

When evaluation covers periods of up to five years, the standard couple-years of protection approach, which reduces programme acceptances to a single measure

[7] *Ibid*, pp. 137–161.

[8] The couple-years of protection method yields an index of the prevalence of use of specific methods. It estimates the number of couples protected against the risk of pregnancy during one year.

[9] "Report of the Expert Group Meeting on Methods of Measuring the Impact of Family Planning Programmes on Fertility", *loc. cit.*, p. 155.

combining mixes of method, age and expected fertility level, may be appropriate. It differs from CYP mainly in that its product is births averted in terms of a standard unit of contraception.

When the evaluation is to cover periods in excess of five years, both the standardization approach and the component projection approach are particularly appropriate. The latter method shows the probable course of the crude birth rate given a certain regimen of contraceptive practice. Fertility projection/trend analysis may also be considered suitable and especially if there is evidence that a strong fertility trend was under way prior to initiation of the family planning programme.[10] Regression analysis may be applied whenever the time period is at least one year.

3. *Population covered*

The population for whom the evaluation is needed is also a factor determining the evaluation approach. Case-by-case matching is a productive means of studying acceptors, but this method is not illustrated in this text. However, the approach has many commendable aspects and is to be preferred if resources and conditions permit.[11] If the evaluation is to deal with acceptors of specific methods, several approaches are possible. The component projection approach treats the behaviour of acceptors only, as does the standard couple-years of projection method.

Reproductive process analysis might be the approach to choose, provided that the data demands can be satisfied. With this method, the units of analysis are segments of contraceptive use. As summarized by Forrest and Ross,[12] this approach follows acceptors from acceptance through continuation to termination or pregnancy and birth events, and the yield is an estimate of the births averted due to utilization of programme contraception.

Some methods attempt to measure the effect of the programme by analysing the behaviour of the entire national population. If the interest is in this direction, the researcher may apply fertility projection/trend analysis, standardization or multivariate or other areal analysis.

4. *Separation of programme and non-programme effects*

Evaluation of family planning programme effects upon fertility obviously implies that the researcher will determine the change in fertility that is attributable to programme activities as distinguished from the amount of change due to factors not directly related to the programme. Separating programme and non-programme effects is one of the more difficult problems associated with family planning programme evaluation. Some evaluation approaches either do not attempt this task or merely carry an assumption, with certain conditions, that any fertility below what is expected or what has been determined to be the potential is due to the programme. Programme and non-programme effects can, of course, overlap; the latter effects can facilitate the former. However, deciphering relative effects in these circumstances calls for a different level of analysis.

Several approaches are designed to separate programme and non-programme factors. Reference has already been made to the standardization approach, by which it is possible to separate the effects of age structure and nuptiality and to determine how much change is due to changes in marital fertility, which may possibly, but not certainly, be due to the programme. Multivariate areal analysis can be applied so as to separate changes due to the programme from those brought about by social and economic improvements or by other factors, including cultural changes that may be indicative of modernization. This approach can also distinguish changes in national fertility due to the altered behaviour of acceptors from changes due to the reproductive behaviour of persons who are not participating in the programme. One view is that the most appropriate procedure for understanding the relative importance of family planning programmes vis-à-vis development is regression analysis.[13]

Another type of areal analysis suited for this purpose is experimental design, in which the populations of different geographical areas are subjected to different programme treatments. It is a method most easily implemented in the early stages of a programme, when different areas can be subjected to different degrees or quality of programme activity; and, as a control, one or more areas may be left without a programme. However, for political and moral considerations, it may not be acceptable to withold from programme treatment the population of an area merely for the sake of evaluation research.

It has been said that simulation is generally inappropriate as a method of family planning programme evaluation,[14] although some models of the component projection method incorporate features of simulation. But the impact of family planning programmes on fertility has been investigated by means of simulation models, and those investigations have been able to focus on effects attributable to factors other than programme influences.[15] However, their major strength in evaluation research appears to be in determining probable results of alternative programme strategies, in target setting and in testing the validity of results obtained by different evaluation approaches. They are said to approximate only crudely the demographic process[16] and therefore

[10] *Ibid.* See also J. D. Forrest and J. A. Ross, *loc. cit.*

[11] See H. Bradley Wells, "Matching studies", in C. Chandrasekaran and Albert I. Hermalin, eds., *Measuring the Effect of Family Planning Programs on Fertility*, published by the International Union for the Scientific Study of Population for the Development Centre of the Organisation for Economic Co-operation and Development (Dolhain, Belgium, Ordina Editions, 1975), pp. 215–244.

[12] *Loc. cit.*

[13] "Report of the Expert Group Meeting on Methods of Measuring the Impact of Family Planning Programmes on Fertility", *loc. cit.*

[14] *Ibid.*

[15] For references, see John A. Ross and Jacqueline Darroch Forrest, "The demographic assessment of family planning programs: a bibliographic essay", *Population Index*, vol. 44, No. 1 (January 1978), pp. 8–27.

[16] See Jane Menken, "Simulation studies", in C. Chandrasekaran and Albert I. Hermalin, *Measuring the Effect of Family Planning*

are less than satisfactory as means of measuring programme impact.

5. *Availability of data*

It is well established that the developing countries, among which are the majority of countries with national family planning programmes, do not possess adequate demographic and related data. The deficiency of data for family planning evaluation is a part of the general problem of data quality and supply for these countries. Because some or all of the data needed to apply a preferred evaluation method may be lacking or defective, it may be necessary for the researcher to develop estimates or to make assumptions that would otherwise be unnecessary and frequently to do so on a tenuous basis. Often the solution will be recourse to an evaluation approach that is theoretically and technically less suitable but for which most of the required data are available. Consequently, even if all technical conditions for using a given evaluation approach are satisfied, the researcher may be forced to follow an alternative course for want of the necessary data. It is emphasized, however, that only in rare cases will there be such an insufficiency of data that it will not be possible to apply any means of evaluation. In these cases, periodic field surveys and improvements in the quality of the service statistics should be given highest priority.[17]

6. *Independence and reliability of method*

None of the evaluation techniques satisfies all evaluation requirements, and it is not yet known which method yields the most valid results. One study[18] suggests that two or more evaluation approaches should be used and that the replication of results should be taken as evidence of validity; but in certain circumstances, two different methods could yield different results, each of reasonable validity if, for example, there were differences in time span, assumptions, coverage or other important conditions. As a rule of thumb, ". . . two methods may be viewed as independent if they utilize different frames of reference in assessing programme impact".[19] Considerations in this respect are: (*a*) number and type of assumptions; (*b*) whether the factors utilized were demographic, biological and/or other; (*c*) the estimating technique employed; (*d*) whether the approach assesses programme impact directly or as a residual; and (*e*) whether the method is independent as a coverage.

An outstanding problem in respect to evaluation methodology is whether the evaluation approaches currently employed meet the requirement that they quantify the phenomena which they are designed to measure. The methods outlined in this text are generally thought to be adequate in this respect.[20]

The validity of the results that an approach produces depends not only upon the postulates embodied in the method but upon the assumptions that the researcher makes in order to accommodate data deficiencies and to estimate parameters that are lacking. Users of this *Manual* should therefore pay careful attention to the implications of these assumptions.

Experience has shown that evaluation results cannot necessarily be taken at face value, but must be interpreted in light of: (*a*) the quality and reliability of the statistics used in the evaluation, including biases attributable to sampling and non-sampling sources; (*b*) the methods used to correct unreliable, missing or incomplete data, which may have introduced still other problems; (*c*) the definitions and estimates of variables; and (*d*) all assumptions, as these factors affect the final product. Wherever possible, the researcher should test the effect of alternative definitions, estimates and assumptions as an aid to achieving valid interpretation of evaluation results.

Programs on Fertility, published by the International Union for the Scientific Study of Population for the Development Centre of the Organisation for Economic Co-operation and Development (Dolhain, Belgium, Ordina Editions, 1975), pp. 351–380.

[17] For a more extensive discussion of this problem, see "Report of the Expert Group Meeting on Methods of Measuring the Impact of Family Planning Programmes on Fertility", *loc. cit.*, p. 159.

[18] "Methods of measuring the impact of family planning programmes on fertility: problems and issues", in *Methods of Measuring the Impact of Family Planning Programmes on Fertility: Problems and Issues* (United Nations publication, Sales No. E.78.XIII.2), pp. 35–36.

[19] *Ibid.*, p. 35.

[20] See, for example, *ibid.*; and C. Chandrasekaran and A. I. Hermalin, eds., *op. cit.*; "Report of the Expert Group Meeting on Methods of Measuring the Impact of Family Planning Programmes on Fertility", *loc. cit*; and J. A. Ross and J. D. Forrest, *loc. cit.*