Part Three

STATEMENTS ON METHODS OF MEASURING THE IMPACT OF FAMILY PLANNING PROGRAMMES ON FERTILITY

Submitted by members of the Expert Group
COMMENTS ON COMPARISON STRATEGIES FOR THE EVALUATION OF FAMILY PLANNING IMPACT

William Brass*

The background paper, "Methods of measuring the impact of family planning programmes on fertility: problems and issues" (ESA/P/AC.7/1), systematically reviews the area of study. The following notes are of particular interest in relation to recent applied research. No attempt is made to examine the topics in detail; rather, points are raised for discussion.

Perhaps the most significant element in the development of family planning evaluation over the past few years has been a change of attitude. In the earlier periods, there was a strong tendency for the subject to be treated in isolation as a well-defined subdiscipline. Currently, to an increasing extent, it is seen as an integral part of fertility analysis. Thus, family planning is one of a variety of factors underlying childbearing behaviour at different levels of "depth". The explanation of the part it plays cannot be separated from the interaction with the other influences. The consequences of this change of attitude have not yet been fully realized. Among them are the need to move the focus from the programme itself towards the broader determinants of fertility, which in turn has a whole series of implications about the resources which are justified for information gathering about individual programmes and the nature of the data required.

Research strategies are also affected if the aim is to measure the contribution that a family planning programme is making to fertility change individually or, more plausibly, synergistically with other variables. In particular, there is the necessity for comparative materials for a range of situations in which these variables enter either explicitly or implicitly in combination with different degrees and/or types of family planning effects. This need sets the subject within the broad conceptual and methodological field of observational interpretation which has a much longer history of study within other disciplines, for example, epidemiology. The principles and techniques that have evolved in these disciplines can be translated and adapted for family planning evaluation.

By their nature, observations of an uncontrolled process cannot provide the same kind of certainty about cause-and-effect relationships as can well-designed experiments. Quasi-experimental approaches, for example, the use of comparative groups but without treatment randomization, can never fully eliminate the unknown factor correlations which make invalid quantitative specifications of belief in terms of probabilities. Subjective elements must enter, which makes the investigation a search for plausibility. However, there is sufficient agreement about the kind of evidence that establishes plausibility to most informed observers for a consensus to be reached.

Internal comparisons of fertility changes in appropriate subgroups of a population are much more powerful than assessment in terms of theoretical or model comparisons and the returns from the latter are weak and limited. The reason is the great variability in reproductive determinants and behaviour over populations. Models are a form of simple averaging and their use as standards against which observed measures are assessed must ignore many influences which may be critically relevant. Of course, models can sometimes contribute towards an essential adjustment of direct measures which are not immediately comparable, but reliance on such uses must be very tentative.

Within-population comparisons have the immense advantage that an important part of the determining forces will be common to all subgroups, permitting the remainder, including family planning programme effects, to be defined more precisely. (Of course, interpopulation studies are needed to illuminate the common factors.) Several broad types of comparative design are possible but some principles are applicable to all. One such principle is the need for replication, that is, the repetition of the examination of measure differences related to a characteristic, in a variety of circumstances. Another is the importance of timing since influences are directional; conclusions must be governed by considerations of the order in which changes occurred and not only by their size. Although there is little possibility of strict randomization because control over systematic effects is impracticable, it may be useful to include some randomization to guard against the introduction of subjective biases.

These points are general but they will be illustrated by reference to particular types of comparative investigation. Multiple replication can be achieved by the relation of measures of fertility change in subareas of a country or region to indexes of family planning effort, use or effect, and other putative determinants normally of a socio-economic nature. But the replication is often much less than would appear because of strong dependence of the family planning indexes upon the socio-economic indexes. In addition, it is difficult to

* Professor of Medical Demography, University of London; and Director, Centre for Overseas Population Studies, London School of Hygiene and Tropical Medicine, London, England.
take satisfactory account of the underlying relationships because of restrictions on the data available (frequently from censuses) and the complexity of the possible lag intervals. The results of such multivariate analyses are essentially superficial as there is not the detail of outcome and characteristics to permit the closer linking which could throw light on how the measurement of effects is distorted by dependence.

Some improvement may be gained by quasi-experimental approaches in which comparative design classifies in the same strata subareas "homogeneous" with respect to fertility determinants. Within one stratum or several strata, family planning programme differences are established preferably, but not necessarily, subsequent to the stratification. In the extreme, but most usual, situation there may be one stratum and two subgroups: that is, family planning activity is concentrated in one area and the other serves as a control. For example, in an experimental programme in the Matlab region of Bangladesh, the family planning area is being "saturated", household by household, with supplies of contraceptives. The degree of replication in this design is very thin for subareas and for strata. Both are important, but for slightly different reasons than those operating in the classical randomized-block experiment. The provision of many comparisons in varied situations guards against the random effects both of external influences and of possible dependence due to the "contamination" of the control by the operational areas, because such dependence is likely to be variable.

Quasi-experiments can be assessed by the same kinds of general summary indexes as used in the multivariate analysis of observations for subareas (birth or fertility rates, percentage of family planning acceptors etc.), but more detailed and sophisticated measures are preferable. Replication, in fact, need not be done by complete subpopulations in separate areas, but can also be through particular groups by age, ethnic origin, social class, education and so on. In other words, the need is for a series of comparisons of the effects of a factor in different contexts with evaluation based on the consistency of patterns. But the nature of the reasoning about the impact of family planning must now be altered because the biosocial subgroups cannot be expected to respond in the same way to other influences (apart from random or more properly erratic distortions) which is the hypothesis for designated and control subareas. The criterion must be the agreement of the differentials by the family planning indexes with a coherent set of concepts of how a genuine impact would operate, as opposed to a situation where the programme use was a substitute for action to achieve the same ends by other means. Of course, there is certain to be a subjective element in any such set of concepts and ideas are not static, but a broadly acceptable view appears to be attainable. Thus, responses of subgroups (by age, education, type of community, religion and so on) to a family planning programme to an extent that appears to reflect the allocation of effort rather than the social classification can be taken as valid evidence of impact. The quantitative measurement must, however, be extremely uncertain.

The greater the detail of significant classification, the more incisive is the evaluation since it expands the opportunities for the effects of uncontrolled factors to be registered. But detail has another important advantage. "Nuisance" influences can be eliminated and the sensitivity of the comparisons increased by concentration on the refined measurements which will show the relevant indications more strongly. This detail may, in some cases, include basic biological features which can interact with family planning adoption. Thus, in Nigeria, survey studies are being made of "modernizing" and traditional groups with the objective of separating out the effects of family planning and biosocial factors, particularly breast-feeding practice, on fertility change. The component of the birth interval associated with level of breast-feeding is so large that modifications in practice can interfere greatly with the attempt to examine other influences.

The logic of the evaluation by detailed comparison of measures in subpopulations differently affected by a family planning programme can be extended to the situation where the designated group consists of the entrants to the programme itself. The controls are those who have not entered (or a representative sample) but who are subject to the other influences towards lower fertility and who have some opportunity to respond. Of course, it is always possible to argue that entrance to the programme constitutes such a powerful selective process that the bias is overwhelming. The argument has great force in some situations but much less in others. The most favourable case is where a substantial proportion of the potential clients has entered, build-up is rapid, there is considerable social diversity and access to contraception outside the programme is relatively easy. It is also necessary to assess the change in fertility of all entrants to the programme (or a sample) regardless of whether they have continued to be clients, since, otherwise, the biases are insuperable.

The measurement of fertility decline of the entrants and of the controls on a consistent basis raises difficult problems, even if it is assumed that behavioural biases are sufficiently modest. This difficulty arises because the entrants are selected for past fertility in complex ways. They are not pregnant and differ from the average in exposure to risk, proportions sterile, intervals to last birth, births in previous periods and fecundity. Much study has been given to the problems of adjustment for these selection biases, usually under the topic heading of "births averted". The results are theoretically not too satisfactory, but the range of error is often small enough to be unimportant in the conditions where impact can reasonably be evaluated. The major difficulties are two. Some effects can be calculated quite precisely but only if there is sufficient information, e.g., on the times of last pregnancy of entrants
and also on the biological components of the birth intervals (post-partum amenorrhoea, delay to conception etc.). Others depend upon characteristics that cannot be measured directly, in particular, fecundity and sterility. Only the outcome in fertility can be observed and it includes a large stochastic element which is part of the selection bias.

The adjustments for selection biases have often been based on model constructs of the reproductive process. However, there are increasing doubts about the regularity of the measures in different populations, in particular, the relation of post-partum amenorrhoea to lactation and cultural factors and the pattern of sterility with age (and recognition of sterility) in different nutritional and social conditions. The need for comparisons within the populations becomes more and more evident. It is here that the chance elements in fertility are particularly troublesome because the measurements reflect to an unknown extent the persistent (which is relevant to the comparisons) and the transient (which is not). For example, women aged, say, 35 years enter a programme with, in general, a higher parity, greater fertility in the past five years and shorter interval to the last birth than the average for married women of the same age in the population. Part of the difference is due to sterility (which will continue), part to risk exposure and fecundity (which can change) and part to chance (where the expected future excess is zero). Barrett and Brass\(^1\) have shown that the chance effects are large and a recent calculation from historical data shows a very moderate correlation of about half between fecundity in two successive birth intervals. Thus, the attractive idea of establishing the fertility potential of the entrants by their past performance has major snags.

Nevertheless, the advantages are so considerable, in terms of the minimization of model elements in the measurements, that more study of procedures for reducing the stochastic biases is indicated. The most promising approach is through the specification of the most effective controls, either in reality or conceptually. The stochastic biases can be eliminated from the comparisons by appropriate matching of controls to entrants so that both groups have the same chance distortion. Presumably the right criteria for matching depend upon recent birth intervals but more research is needed on the best choice and its sensitivity. If the control matching is conceptual, that is, a computer simulation of the stochastic biases in a population with the "persistent" characteristics of the entrants, further problems arise of the proper specification of features, such as fecundity variation. Brass and Barrett produced a conceptual matching for the evaluation of the data from the Mauritius family planning programme and have recently developed more general results.

MEASURING THE IMPACT OF FAMILY PLANNING PROGRAMMES ON FERTILITY

Chen-tung Chang*

In measuring the impact of a family planning programme on fertility, it is generally assumed that a birth averted is identical with every other birth. In a sense, this assumption is indisputable; if the same number of births has been avoided over the same period of time, the same fertility impact has been achieved. However, if one wants to go beyond the mere statement of the number of births averted and try to relate it to prospective population growth for planning purposes, the untenability of the assumption becomes all too apparent. The same problem also arises when one attempts to compare the fertility impacts of two or more family planning programmes. Is a birth averted in a programme equivalent to a birth avoided in another? The answer hinges upon the question whether the two births have been averted for the same reason.

For the present purposes, the most relevant way to classify reasons for family planning is into spacing and limiting. To have a proper assessment of the fertility impact of a programme, it is important to ascertain whether the acceptors have come for purpose of delaying or stopping childbearing. The same number of births averted over the same given period of time, especially when the period is short, as when a programme is evaluated on an annual basis, may have substantially different implications for prospective fertility if the predominant reason for family planning differs between acceptors of two programmes. For child-spacing, by definition, implies merely a temporal redistribution of live births over a period of several years. Its impact on fertility, as far as the quantity dimension is concerned, is therefore of only a temporary nature. On the other hand, births averted because of the decision to limit family size may be avoided forever.

It will be observed from table 1 that the practice of child-spacing is very popular in Singapore, with a substantially high proportion of delayed births among all births averted in a given year.

There is no question about the fertility impact of the national family planning programme in Singapore. The national programme began in 1966, about a decade after the inception of the transition from high to low fertility. However, evidence strongly indicates that the programme has significantly accentuated the downward trend of fertility. Especially in the first few years, actual fertility was much lower than expected fertility, however it is estimated. The question is that despite the continuing recruitment of new acceptors, the fertility reduction decelerated, and there was actually a slight increase in the total fertility rate in 1972.

The most likely explanation of this pattern of fertility change is that the initial impact of the programme resulted in the recruitment of large numbers of women practising birth control for the purpose of spacing. Because many of these delayed births appeared several years later at a time when there was a decline in the number of new acceptors, there was an upsurge in births and hence the stagnation of the downward fertility trend and even a slight upward turn.

The availability of birth-order data permits a more careful examination of this process. It will be observed from table 2 that in 1971 and again in 1972, whereas birth rates for births of the fifth order and above continued to decline, lower-order birth rates showed increases for many age groups of women. This pattern of changes indicates that as the reductions of higher order birth rates reflected the intention to limit family size, the increasing practice of birth control by high-parity women continued to result in the reduction of the birth rates. On the other hand, much of the reduction of lower order birth rates was due to the wide practice of child-spacing in response to the inception of the national family planning programme in 1966. As postponement of childbearing is usually only for a short period, many of the births delayed in the first few years were appearing several years afterwards. As a consequence, despite the continued spread of the practice of birth control, in 1971–1972, increases were observed in many of the lower order birth rates.

* Lecturer, Department of Sociology, University of Singapore.
TABLE 2. SINGAPORE: BIRTH RATES BY AGE GROUP AND BIRTH ORDER, 1970-1972

<table>
<thead>
<tr>
<th>Year and age group</th>
<th>Birth order</th>
<th>1970</th>
<th>1971</th>
<th>1972</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First</td>
<td>Second</td>
<td>Third</td>
<td>Fourth</td>
</tr>
<tr>
<td>15-19</td>
<td>26.1</td>
<td>18.6</td>
<td>6.0</td>
<td>11.3</td>
</tr>
<tr>
<td>20-24</td>
<td>139.1</td>
<td>61.8</td>
<td>43.0</td>
<td>21.5</td>
</tr>
<tr>
<td>25-29</td>
<td>212.8</td>
<td>49.0</td>
<td>53.8</td>
<td>42.1</td>
</tr>
<tr>
<td>30-34</td>
<td>138.0</td>
<td>14.3</td>
<td>23.7</td>
<td>25.0</td>
</tr>
<tr>
<td>35-39</td>
<td>74.6</td>
<td>3.6</td>
<td>5.1</td>
<td>6.8</td>
</tr>
<tr>
<td>40-44</td>
<td>26.7</td>
<td>0.6</td>
<td>0.7</td>
<td>1.3</td>
</tr>
<tr>
<td>45-49</td>
<td>4.7</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>618.0</strong></td>
<td><strong>148.0</strong></td>
<td><strong>131.3</strong></td>
<td><strong>98.2</strong></td>
</tr>
</tbody>
</table>

It is interesting to note in this connexion that in a study on the fertility impact of a family planning programme in one country, a comparison was made between the number of births averted by the programme estimated from use of programme methods and that from differences between expected and actual fertility over the period 1965–1971. The expected fertility was derived from the fertility trend in 1959–1963 because 1963 was the year when the programme gained momentum through expansion. The comparison shows that the calculation, which was based on the difference in expected and actual age-specific marital fertility, over-estimates the number of births averted for the period prior to 1967 and underestimates the number after 1967. The author’s interpretation is that in the absence of the programme, the expected fertility trend after 1963 should be curvilinear, with a higher rate of decline at the beginning and a lower rate at the end, so that the difference between expected and actual fertility would be smaller at the beginning and larger at the end. An alternative interpretation, which appears to be more plausible, is that the reduction of fertility was due in part to the practice of child-spacing, and the births originally to have appeared in the early years were postponed to the later years. In other words, a phenomenon occurred which is similar to what has happened in Singapore.

The two examples serve to make a point that if evaluation is conducted as a routine operation on an annual basis, the results must be interpreted carefully for their fertility implications. Otherwise, the elation at the initial impact of the programme or the introduction of new schemes may soon be followed by alarm when the downward trend of fertility is arrested, which, when misinterpreted, may in turn lead to the development of unnecessarily drastic measures. The problem arises from the failure to make a proper distinction between the short-term and long-term effects of a programme. Strictly speaking, the practice of child-spacing also has its long term effect, inasmuch as it affects the tempo dimension of fertility. For all practical purposes, however, its annual implication is negligible when the effect is dispersed thinly over the length...
of a generation. The fertility impact along the quantity dimension is, therefore, the primary concern here.

The proper assessment of programme impact, with due recognition of the distinction between short-term and long-term effects, can be approached if the number of couple-years of protection (CYP), for example, is measured for each parity group of acceptors specifically. The number of births averted is thus estimated together with their birth-order composition. Since higher order births are averted most probably because of child-limiting, they can be given more weight than those of lower order in assessing the "permanent" fertility impact of the programme. In other words, provided with information not only on the total number of births averted but on their birth-order composition, it is possible to have a more "realistic" interpretation, on a short-term basis, of the impact of a programme on the fertility of a population.

Nevertheless, "... restricting the measurement of programme impact to change in unwanted fertility would exclude the demographic and other impact of spacing, the improvement of which may turn out to be one of the principal contributions of U.S. family planning programs." Frederick S. Jaffe, "Issues in the demographic evaluation of domestic family planning programs", in J. R. Udry and E. E. Huyck, eds., The Demographic Evaluation of Domestic Family Planning Programs (Cambridge, Mass., Ballinger Publishing Co., 1975), pp. 19-30.

The same applies to the fertility projection approach. If in comparing the actual with the expected fertility both are examined in more detail by birth order, one can then gain some idea of the proportion of their difference that is due to changes in fertility of higher parity women. Again, since the effects on higher order and lower order fertility are assessed separately, one can maintain a more balanced perspective on the over-all impact and be better guarded against the misinterpretation of the fertility implications of programme impact.

Obviously, such a more differentiated approach to assessing fertility impact is especially necessary for programmes where child-spacing is popular among acceptors or where the proportion of acceptors practising spacing changes significantly because of frequent programme changes or policy revisions. In Singapore, for example, child-spacing has been popular from the beginning of the national programme. Moreover, in the six years from 1967 to 1973, the proportion of acceptors practising spacing increased from 53 per cent to 84 per cent, while the proportion with one child or none at all more than tripled, rising from 21 per cent to 67 per cent. For a programme like that in Singapore, it is necessary to know more about the births averted, especially the birth-order composition, in order to interpret the fertility impact of the programme on a firmer basis.
NEEDED DATA AND RESEARCH ON THE IMPACT OF FAMILY PLANNING PROGRAMMES ON FERTILITY

G. Edward Ebanks*

There is no necessity to labour the point concerning the need for data and scientific research in the area of evaluating the impact of family planning programmes on fertility. In this short paper, the following assumptions are made:

(a) There is, or there is about to be launched, a family planning programme, national or otherwise;

(b) Because of emphasis, and rightly so, upon delivery of services, data collection within the everyday operation of the programme is minimal, e.g., age, marital status, parity, living children, occupation, education and religion;

(c) There is a separate research and evaluation unit with adequate funds for conducting research;

(d) Use of an electronic computer is available;

(e) There is no need to make assumptions about trends in fertility coinciding with or following upon the initiation of the programme;

(f) Because of the nature of the programme within a single national unit, it is not possible to set up an experiment with experimental and control groups. The programme, whether national or subnational, is assumed to contaminate to greater and lesser extents the subunits of the national population.

Since using service statistics as sufficient for evaluating the impact of the programme on fertility is implicitly ruled out and since a true scientific experiment is also ruled out, one is left with two appropriate research possibilities. The first is use of sample survey techniques and the second is an extension of this method to panel studies. Also ruled out is the matching of clients and non-clients because of the many problems associated with that type of design.

Sample surveys, if well planned and executed, have the potential to yield the type of data necessary for an evaluation of the impact of a family planning programme on fertility. The evaluation of the effectiveness of a family planning programme in terms of its impact on fertility is best done within a broader framework, that of establishing the determinants of fertility within a specific socio-cultural system. In other words, a study directed towards assessing the impact of a programme on fertility can be extended to an examination of the determinants of fertility, another area in which there is a need for data and research.

One can, using the Blake-Davis formulation of 11 intermediate variables and Freedman's extension of it, establish the links between the institutional variables and the intermediate variables, and between those variables and fertility. One intermediate variable, use or non-use of contraception, may take on significance all the way from just a minor one of the 11 variables up to its being the single most important determinant of fertility and hence making nought of most of the other 10. The family planning impact at this extreme can then be seen as an institutional variable influencing the intermediate variables. Properly designed and executed sample surveys of the types presented below can assess the intermediate variables, as well as the socio-cultural variables operating through them to influence fertility.

APPROACH 1: PERIODIC FIELD SURVEYS

A periodic field survey should be conducted about every five years. Assuming that each of these periodic sample surveys is intended to cover a period of evaluation of, say, 5, 10, 15 or 20 years, the initial survey may try to assess the impact of the programme over the past period of, say, anywhere from 5 to 20 years. The longer the time period to be covered, the more difficult is the problem, the larger the sample required and the greater the problems of reliability and validity. By way of an example, one may assume that the initial survey is directed towards evaluating the impact of the family planning programme over the first 10 years or the most recent 10 years. The appropriate age range of women to be included in the sample is 15–59 years, or should, in the present author’s view, be that range. At the beginning of the period, the women currently aged 59 were 49, the upper limit of the childbearing ages. These women and those aged 45–48 years are included in order to have completed fertility for a five-year age cohort at the initiation of the programme. The 15–19 age group was aged 5–9 years at the initiation of the programme. They are now included because they have just entered the childbearing ages. By first including the age range 15–59, one has women aged 15–49 for all the 10 years for which the evaluation is being conducted.

---

* Associate Professor, Department of Sociology, University of Western Ontario, London, Ontario, Canada.


171
The aim here is to be able to calculate age-specific fertility rates for each of the 10 years under consideration. This calculation should be done for single years of age, which can then be aggregated into five-year age groupings for each of the 10 years under consideration. The calculation of age-specific fertility rates is seen as a necessity for an evaluation of the impact of a family planning programme on fertility.

If the evaluation is to cover a period of five years, one would select into the sample women aged 15–54 years, which would give an age range of 10–49 at the beginning of the period. Similarly, if it is for a period of 15 years, the sample could still include women aged 15–59 years, which would give at the beginning of the period an age range of 0–44, thus missing the 45–49 age group at the beginning of the investigation, an omission which could be accommodated. If it is for 20 years and one includes women aged 15–59 years at the beginning of the period, one would have women aged 0–40 years, thus missing only the two last five-year age groups within the childbearing ages.

Since the objective is the calculation of age-specific fertility rates, one must be concerned with the number of women at each age over the period of investigation. For argument’s sake, one may say that 100 women are required for each single year of age. Because one is including ages 15–59, that is, 45 single years of age, 4,500 women would be required for the sample. However, one would at least want to compare ever-users of contraceptives with never-users, and hence twice as many women, say 9,000, would be needed. If one wants to compare the age-specific fertility of clients of the programme, non-clients but users of contraceptives and non-users, one should again increase the sample size by a significant number, even if not by another 4,500. If one wants to subdivide further the sample units into, say, active clients, inactive clients, active users who are non-clients, inactive ever-users who are non-clients, and never-users, the sample size should be even larger. To cover adequately all these subgroups, one would need a very large sample, which then becomes expensive. However, research is not cheap and this type of evaluative research, if it is to accomplish its goal, is even more expensive.

A problem not yet discussed which is related to size of the sample is one of identifying the subsamples mentioned above, at each point in time. That is, in the case of the 10-year evaluation, one needs to be able to identify the subgroups in each of the 10 years, so that a woman who is a client in year \( x \) may have been a non-client in year \( x - 1 \), and an inactive client in year \( x + 1 \). The recording of the retrospective data should afford this opportunity and this factor should be considered when deciding on the necessary sample size.

A national probability sample is proposed because in so doing, it is hoped that the other independent variables will be randomly distributed, making it less necessary to control for them and hence further necessitating large sample sizes. On the other hand, a national probability sample must consider the proportion of women in the childbearing ages who are members of the family planning programme, since this will affect the number of possible clients included in the random sample.

The minimal amount of desirable data is a contraceptive history, a pregnancy history and sexual union history. Other useful histories would be a work history and a migration history. The questionnaire should be designed in such a way as to make integration of these histories relatively easy and also to facilitate the identification each year of the various subgroups of interest, so that age-specific fertility rates may be calculated for each of them.

Repeated at about five-year intervals, this research procedure would, in addition to evaluating the effectiveness of the family planning programme, allow the researchers to examine changes in the intermediate variables as well as their societal determinants and hence to provide data on the determinants of fertility.

**Approach 2: The Panel Study**

Approach 2, the panel study, is seen as not very different from approach 1 and hence is an alternative to it. The data to be collected are the same. The aim is the same: evaluation of the effectiveness of the family planning programme through its impact on fertility. Once again, it is done by calculating annually the age-specific fertility rates for single years of age, for some or all of the subgroups identified under approach 1.

One should begin with a national sample covering women in the age interval 15–54. In so doing, one has at first two five-year age cohorts for which one can calculate completed fertility. Data obtained on this group should be similar to those obtained in approach 1. This initial national survey is the base-line study and should preferably coincide with the launching of the programme. However, if the programme is already in existence, it would mark the beginning of the panel study and would retrospectively look at fertility and contraception, among other things. Although covering the above-mentioned age range, there should be oversampling of the 15–19 age group, as those women should become the cohort which will be followed in order to record their social, economic, demographic and contraceptive statuses and changes as they go through the childbearing ages.

These women aged 15–19 years should be followed for 30 years, being reinterviewed every two years, and hence interviewed a total of 15 times. It is proposed that with the two-year interview interval one would include each time women aged 15 years and over. So at the second interview, that is, the interview following the base-line study, one would interview the panel members, who would currently be 17–21 years of age, and also a sample of women aged 15 and 16 years. At the third interview, the original panel members would be 19–23; and one would reinterview them, plus a sample of the women aged 15–18 and so on, until the fifteenth interview, 30 years from the beginning when
the original panel members would be 45–49 years of age and one would include an additional sample of women aged 15–44. In so doing, one would have the benefit of a panel study and also each time of being able to calculate age-specific fertility rates for the 15-year to highest age among the original panel members.

A variation on this technique would be to keep adding each time the younger age groups as additional panel members rather than as described above, where there is an independent sample each time except for the original 15–19 age cohort of panel members. Another variation is to take every two years an independent sample rather than following the panel members. That is, at the second interview or the one after the base-line study, a sample of women aged 15–21 years is drawn and interviewed. On the next round, two years later, another independent sample is chosen of women aged 15–23. This approach enjoys much of the benefits of the first panel type described and avoids the high attrition rates of panel studies as well as the sensitization effect. It is also less costly and more manageable.

Regardless of the format of the panel study, it has several advantages. By following the panel members and interviewing them at regular intervals, one is more likely to obtain reliable information. In all cases of this type of research, validity is a major problem. In following the panel, one can assess the effect of the programme on a group of women passing through the childbearing ages. Both the direct and indirect effects of the programme are measurable within this approach.

SUMMARY AND CONCLUSION

Both approaches briefly described above have much in common, although they differ in certain important aspects. In both approaches, one must be concerned with problems of reliability and validity. Both require the same type of data but get at it in slightly different ways. They are both costly and time-consuming. But in the present author’s opinion, there is no cheap or easy way of evaluating the impact of family planning programmes on fertility. In addition to evaluating the family planning programme, the approaches described above provide micro data for the study of the determinants of fertility. While this study is going on, the macro changes in the society could be carefully observed and recorded so that macro analyses relating to the determinants of fertility could be undertaken. Combining the micro and macro approaches would provide all the necessary data for whatever evaluation techniques may be desired. One would then have data for technical analyses as well as data for the policymakers and programme personnel.

Both approaches have the potential for providing data on such areas as voluntary childlessness, infertility, and changing ideals, norms and values as they relate to childbearing and family sizes. Family planning evaluation is best viewed within the broader framework of the determinants of fertility and these approaches afford the opportunity to do so.

In both approaches, as the aim is the calculation of annual age-specific fertility rates for various subgroups, sample sizes must be large and the possibility of interrelating the various data components must be built in. Only brief mention was made of the types of data required and the techniques for obtaining them because of the limited space and because examples of these needs and methods exist in the literature. The space limitation contributed to the lack of clarity on some of the points made and also to the omission of some of the underlying assumptions.
AVOIDING AN EMBARRASSMENT OF RICHES

Albert I. Hermelin*

It is a well-known reflex among researchers when faced with limitations of inference from a body of data, to call for more data. Though the limitations encountered may be real enough, less clarity usually attaches to the way in which the additional data will relieve the problem. Analysts of cross-sectional data, for example, often send out a plea for panel or longitudinal data but usually without a detailed statement of the structure of the new analysis which will incorporate the desired data. Those concerned with measuring the effect of a family planning programme have been in a similar position, often restricted in what can be said from a single method applied to a single instance and wishing for multiple approaches in a variety of settings. Thanks to the efforts of the United Nations, these wishes have been granted in part, and it is to be hoped that further material of this nature will soon be forthcoming. If one is to avoid an embarrassment of riches, however, the Expert Group meeting might well devote some attention to how best to analyze this increased data base. The purpose of this brief note is to suggest some directions this discussion might take without at all pretending to resolve the matter. More specifically, the aim will be to review some developments about reliability and validity from the sociopsychological literature which appear isomorphic to the concerns of this meeting as a first step in appraising their potential utility. It is assumed in the following discussion that the type of data available to this meeting will be enlarged to more trials so that it makes sense to talk about the correlations over settings of variables emerging from the application of various methods for evaluating the effect of family planning programmes.

ELEMENTARY CONCEPTS

On the most general level, reliability refers to the consistency of a method or test, while validity relates to the ability of a method to measure what it intends to measure. In the "classical" approach, reliability is established through "the agreement between two efforts to measure the same trait through maximally similar methods", while validity is seen as "the agreement between two attempts to measure the same trait through maximally different methods".1 It is rec-

1 Donald T. Campbell and Donald W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix", Psychological Bulletin, vol. 56 (March 1959), p. 82.

ognized, of course, that independence of methods is a matter of degree and an alternative formulation of this distinction will be developed below. Though interest centres on validity, reliability is important because it limits the degree of validity that may be established. It can be shown, for example, that "the validity of a test with respect to any criterion cannot exceed the index of reliability" 2 (where validity is viewed as the correlation of a test score with a second measurement, and reliability is defined as the correlation of a test score and its true score). Attenuation formulae that attempt to estimate the correlation between the true scores of two tests involve the observed score correlation as well as the index of reliability of each test.

ESTABLISHING THE RELIABILITY OF METHODS FOR EVALUATING PROGRAMS

One may thus ask as a first step how the reliability of the various methods for measuring the effect of family planning may be established. Most of the well-known methods, such as test-retest, parallel forms and split-half, have been developed in the field of mental testing. Some of these methods have little applicability to the subject-matter of this meeting, but others may be adaptable to the present concerns. For example, as an analogue to the split-half technique, it should be possible to determine contraceptive continuation rates for random subsamples of acceptors or to carry out areal regression for subsamples of areas, where the number of subdivisions is large enough. If this procedure were done over sufficient instances, the correlation between the measures obtained from each subsample would provide an estimate of reliability.

The parallel-form technique suggests that alternative approaches for obtaining the same measure should be considered. For example, continuation rates may be determined by follow-up of acceptors or by obtaining the necessary information from a KAP survey. Again, the correlation between measures derived from each source would speak to the issue of reliability. One can go further and sort out the various methods for measuring the impact of a programme according to their degree of similarity by analyzing the measures produced as well as the variables employed. One would expect greater correlation among similar measures than among those more dissimilar. Turning

around the inequality between validity and reliability, mentioned earlier, it is also true "that the index of reliability cannot be less than any validity coefficient of a test".\(^3\) Thus, the degree of correlation between methods deemed similar should provide a lower bound to the reliabilities of the methods in question.

**Concepts of validity**

Various approaches to validity have been advanced in the literature: content validity; predictive and concurrent validity; construct validity. The opinion here is that construct validity holds the major interest for this meeting, though the other concepts have some relevance. One may note in passing that it makes sense to judge whether the proposed methods of evaluation bear sufficiently on what they are intended to measure. For predictive validity, it is clear that an independent criterion of the effect of a program is not available, for if it were, all these efforts would not be required. Nevertheless, if a program is evaluated as having a strong effect over a period of years, it is reasonable to expect that this effect will be reflected in an appropriate fertility measure, suitably lagged and adjusted for any demographic confounding effects. This would appear to be a necessary, but clearly not sufficient, condition to establishing the validity of the method in question.

Lord and Novick hold that "for scientific purposes the most important characteristic of a test is its construct validity".\(^4\) Another source states: "Construct validity involves relating a measuring instrument to an overall theoretical framework in order to determine whether the instrument is tied to the concepts and theoretical assumptions that are employed."\(^5\) As such, construct validity is not established in a single step but requires a chain of reasoning in which hypotheses concerning the construct are derived from theory and tested. Campbell and Fiske\(^6\) have advanced the understanding of this process by noting that the validation must be convergent as well as discriminant. That is, the results of a test or method should correlate appreciably with other results that the hypotheses suggest as related (for example, other methods designed to measure the same construct), while there should not be appreciable correlations among variables that theory identifies as relatively independent. In carrying out these steps, one must be alert to spurious correlations which may inflate the correlations observed and lead to incorrect inferences with respect to both the convergent and the discriminant criteria. Spurious correlations may arise because two methods "measure in part something other than the construct of interest".\(^7\) This statement introduces the important concept of method variance as a non-random factor affecting observed measures.

Campbell and Fiske state: "Each test or task employed for measurement purposes is a trait-method unit, a union of a particular trait content with measurement procedures not specific to that content. The systematic variance among test scores can be due to responses to the measurement features as well as responses to the trait content."\(^8\) Lastly, one must recall that the reliabilities of each method come into play when assessing the level of correlation between two variables.

**The multitrait-multimethod matrix**

In order to carry through the ideas on construct validity described in the previous section, Campbell and Fiske state that more than one variable and more than one method must be used. They suggest that to separate the relative contribution to observed scores of trait variance and method variance, it is convenient to set up a multitrait-multimethod matrix in which all of the intercorrelations resulting from the measurement of several traits by each of several methods are displayed.\(^9\)

In the context of this meeting, one might think of a situation where for a reasonably large number of countries, three methods of evaluation have been applied, each of which produces a measure of the impact of the programme on fertility (X), a measure of the effect of a socio-economic variable on fertility (Y), and the effect of some possibly confounding demographic variable, such as age structure or nuptiality (Z). Thus, one has three variables, each measured by three different methods. In this case, the multitrait-multimethod matrix can be visualized as was done by Alwin\(^10\) in his excellent explication of this approach (see figure I).

Carrying on the example suggested, \(X_1\) would be the measure of the effect of the programme obtained by method 1, \(X_2\) by method 2 etc., so that \(r_{X_1Y_1}\) in figure I would be the correlation between programme effect and socio-economic effect obtained by method 1, while \(r_{X_1Y_2}\) would be the correlation between programme effect measured by method 1 and socio-economic effect measured by method 2. Figure I indicates two different blocks of correlations:

1. The monomethod blocks which show the intercorrelations among variables obtained by a single method, the diagonals of which represent the reliabilities for each measure for that method \(r_{X_1X_1}, r_{Y_1Y_1}\)

\[^{3,4}\text{Ibid., p. 72.}\]
\[^{6}\text{Loc. cit.}\]
\[^{7}\text{F. M. Lord and M. R. Novick, op. cit., p. 279.}\]
\[^{8}\text{D. T. Campbell and D. W. Fiske, loc. cit., p. 81.}\]
\[^{9}\text{Ibid.}\]
Figure I. Multitrait-multimethod matrix for three traits and three methods


Note: Values in validity diagonals (MTHM) are in bold-face type.

The entries in the validity diagonal should be significantly different from zero and sufficiently large to encourage further examination of validity. This requirement is evidence of convergent validity. Second, a validity diagonal value should be higher than the values lying in its column and row in the heterotrait-heteromethod triangles. That is, a validity value for a variable should be higher than the correlations obtained between that variable and any other variable having neither trait nor method in common.

A third common-sense desideratum is that a variable correlate higher with an independent effort to measure the same trait than with measures designed to get at different traits which happen to employ the same method. For a given variable, this involves comparing its values in the validity diagonals with its values in the heterotrait-monomethod triangles. A fourth desideratum is that the same pattern of trait interrelationship be shown in all of the heterotrait triangles of both the monomethod and heteromethod blocks.\textsuperscript{12}

\textbf{GENERAL CAUSAL MODEL FOR ANALYZING THE MULTITRAIT-MULTIMETHOD MATRIX}

The scheme proposed by Campbell and Fiske is clearly a promising first step in establishing construct validity and for analyzing a series of results obtained by alternative methods. Evaluation of this proposal, however, has detected a number of weaknesses. For example, although Campbell and Fiske note the relevance of reliability, they provide no systematic way of taking this factor into account.\textsuperscript{13} As stated by Siegel

\textsuperscript{11} Loc. cit., pp. 82–83.

\textsuperscript{12} D. T. Alwin, loc. cit., p. 82.

\textsuperscript{13} Ibid., p. 83.
and Hodge,\textsuperscript{14} it is desirable to develop models that would simultaneously address questions of validity and reliability. Of the alternative approaches to the Campbell-Fiske methodology well reviewed by Alwin, outlined below are the features of the general causal model since it clearly brings out this distinction.

As against the definitions of reliability and validity given above in the section on elementary concepts, one may reformulate the distinction in terms of the unobserved construct one is attempting to measure and the factors that cause the observed values to depart from it. From this standpoint, questions of reliability focus on the random errors \((e)\) that cause observed scores \((O)\) to depart from the \textit{true} scores \((T)\) represented by the formula:

\[ O = T + e \]

Questions of validity arise when variation in the observed scores is seen as arising also from other factors—the methods of measurement or other constructs—which are non-random or systematic. Under certain conditions,\textsuperscript{15} the effects of both types of variation can be estimated through the factor-analytical techniques developed by Jöreskog.\textsuperscript{16}

In the formulation relevant to this discussion, each measured variable would be treated as arising from the unobserved construct, the method used to obtain the measure, and random variation. Figure II represents this situation for the hypothetical situation posited earlier of three methods each used to measure three constructs. The correlations obtained are the basis of estimating a series of parameters \(\alpha_i\) and \(\beta_i\) for structural equations of the form:

\[ X_i = \beta_i X + \alpha_i M + \mu_i \text{ etc.} \]

in which "each measured variable has non-zero loadings on one trait factor (\(\beta\)) and one method factor (\(\alpha\)) but zero loadings on all other trait and method factors".\textsuperscript{17} It should be noted that this model allows both trait and method factors to be correlated, and these correlations to be estimated.

**Some final observations**

The foregoing observations were designed to show that developments in the study of reliability and validity, particularly through a causal model in which each source of variation is identified, are structurally similar to the type of data that may become increasingly available from the evaluation of family planning programmes. It would appear to be worth while, therefore, to carry out further analysis of the potential applicability of these methods to these data. Can one identify three or more methods of evaluation which produce comparable measures on three or more constructs thought to influence fertility? If not, is it possible to alter the analytical framework to fit the type of data produced? Or, alternatively, is it more promising to consider modifications of current methods of evaluation so that these fairly powerful techniques can be utilized? It may turn out that these methods for treating reliability and validity prove more useful for studying alternate ways of measuring the effect of a programme, not on fertility, but on variables antecedent to it, such as the proportion of the population currently using contraception and continuation rates. For example, prevalence of contraceptive use along with characteristics of users may be obtained through surveys or derived from service statistics and the general causal model outlined above may be applicable to examining these alternative approaches to measuring these key aspects of a programme. Questions of this type might well occupy some of the attention of this meeting if increased data on evaluation are to be analysed to maximum advantage.


\textsuperscript{15} The conditions relate to the question of identifiability, which may be thought of as the number of unknown parameters to be estimated in relation to the number of observed correlations.


\textsuperscript{17} D. T. Alwin, loc. cit., p. 84.
SOME ASPECTS OF DETERMINING APPROPRIATE METHODS OF EVALUATION FOR MEASURING THE IMPACT OF THE FAMILY PLANNING PROGRAMME ON FERTILITY IN THE REPUBLIC OF KOREA

Kap Suk Koh*

The recent demographic history of the Republic of Korea is well known to those interested in the study of human fertility. The total fertility rate in the Republic of Korea declined from an estimated 6.17 in the early 1960s to 3.9 in the early 1970s. The crude birth rate declined by approximately 45 per cent between 1960 and 1975, falling from 42 to 24. This decline in the fertility of women in the Republic of Korea took place in conjunction with several other important developments. Over-all economic development has increased greatly since the early 1960s. During the period 1965–1975, the gross national product grew at an average annual rate of approximately 11 per cent. Per capita income increased from $83 in 1960 to $513 in 1975. In addition, urbanization increased. Approximately 50 per cent of the population currently live in cities of 50,000 or more; in 1960, the comparable figure was 28 per cent.

Of more immediate interest is the fact that in 1961, the Government of the Republic of Korea established a national family planning programme. Since then, approximately 8 million women have received programme-supported services. In addition, the age at marriage has risen steadily and the use of induced abortion increased considerably throughout the period 1960–1975.

As a consequence of these developments, the Republic of Korea has become something of a battleground for those arguing about the impact of national family planning programmes on fertility. Because of the obvious importance of the effect of family planning programmes on fertility, as well as the interest in the experience of the Republic of Korea, this meeting is especially welcomed by experts from that country. It is unfortunately the case that those experts do not know as much as they would like to know about how to measure the impact of family planning programmes on fertility. It is hoped that this meeting will clarify at least some of the issues involved and lead to increased efforts in this important area.

Several of the methods discussed in the background papers for this meeting have been employed with data of the Republic of Korea. Cho and Retherford 1 have used the standardization approach. Smith and Koh 2 have carried out a trend analysis. Yang 3 has reported on an experimental study. Lee and Isbister 4 employed data of the Republic of Korea in a report. The Korean Institute for Family Planning 5 has calculated couple-years of protection (CYP). Analyses have been made of the reproductive process, following the leads of Potter and others. Mode and Littman 6 have worked on a simulation model. Analysis along these lines continues at several institutions both in the Republic of Korea and elsewhere.

All of the studies done so far have suffered from various short-comings. It is impossible to make specific comparisons here of the results obtained with the various procedures because of the different data sources and temporal reference periods employed in each study. Instead of detailing particular failings of past research, the present author would like to point out some of the problems that are most troublesome in the continuing analysis of the impact of the family planning programme on fertility in the Republic of Korea.

The selection of a given evaluation procedure depends upon one’s orientation towards certain measurement problems and the availability of the necessary data. In the Republic of Korea, the existence of a body of high-quality survey data and the problems with programme statistics limit the approaches that can be employed in a straightforward fashion.

MEASUREMENT METHODS

The standardization approach, at least as it has been utilized to date, has failed to come to terms with an important problem. It has missed the influence of socio-economic factors which are probably crucial determinants of fertility, both directly through their effect on marital fertility and indirectly through their impact on fecundity and the probability of having a live birth and a given level of infant mortality. One aspect of this problem is that observed changes in marital fertility are difficult to allocate to specific causal factors, such as the family planning programme. Although the standardization procedure is useful as a measure of changes in fertility due to such factors as age structure, age at marriage, proportions marrying and marital fertility, it is less helpful in specifying the contribution of the family planning programme to changes in these variables, most importantly marital fertility. In short, there is need to take account of the complexity of demographic change and socio-economic development and their interrelations. Standardization procedures probably can be modified to take into account additional factors. These modifications are necessary if one is to measure the impact of family planning programmes. A hopeful sign is the increased availability of sample data obtained from censuses that can be manipulated in a way that allows one to control to some extent the impact of socio-economic factors and to study changes in fertility among various subpopulations.

The projection approach has not been popular in the Republic of Korea. Some informal attempts have been made to utilize this procedure but the results have not been widely distributed. In part the projection approach suffers from the same shortcomings as the standardization procedure. For the most part, available projection models have no completely adequate way to incorporate socio-economic and development variables that may influence the level of fertility independent of the family planning programme.

The specification of hypothetical fertility trends has usually been an unsatisfactory exercise. The complexity of recent demographic developments in the Republic of Korea means that the trends one assumes would have existed if there had been no family planning programme must be very carefully specified. In a country where the family planning programme has been influential, as is the case for the Republic of Korea, it is extremely difficult to specify what would have happened in the absence of the programme. Attributing changes in fertility to programme and non-programme factors is a major problem. For example, induced abortion,7 in part supported by the programme but also widely available through non-programme channels, complicates any analysis. It is especially difficult to estimate what the impact of abortion would have been without the programme. The question arises whether the programme has prevented a higher level of abortions through the provision of contraception or whether it has helped to create the atmosphere and provide the resources that have made abortion so popular.

The Republic of Korea has been the location of several well-known experimental studies.8 The difficulty with the experimental-design approach is that it is nearly impossible to assume that control and experimental areas are not influenced by unmeasured or even unknown factors. The "Hawthorne" effect is probably important in treatment areas where villagers are not used to and do not know how to deal with the considerable attention given them by programme workers. The rapid social changes of the past 15 years makes it extremely difficult to evaluate the differences between experimental and control areas. Moreover, because of the irregular pattern of change it is hard to judge the extent to which changes in one area are a reflection of changes in other areas of the country. Lastly, it must be pointed out that the experimental approach, although perhaps the most ideal procedure in theory, is extremely expensive and time-consuming in practice and is obviously limited by ethical considerations.

Couple-years of protection has become the central methodology in the official evaluation of the programme in the Republic of Korea. The CYP approach has been used for evaluation purposes since 1974. Currently, family planning acceptor targets also are calculated on the basis of the CYP approach. Several problems have arisen with this approach. Among the most troublesome are the accuracy of service statistics and the limited and too quickly outdated data on contraceptive continuation. Measuring the impact of abortion is also difficult because it is heavily influenced by the national programme. The most severe problem encountered in the Republic of Korea concerns the accuracy of the data collected through problem channels. What is needed is some way to measure both the reliability of service statistics and the continuation of contraceptive use. Several surveys along these lines have been conducted in the past. However, there is reason to believe that recent changes in target-setting procedures and payment to workers may have changed the situation sufficiently so that new surveys are needed. The case in the Republic of Korea is one example of the general problem of data availability and reliability, and the changes in these over time that limit the application of the CYP procedure.

The last four procedures, component projection, analysis of the reproductive process, regression analysis and simulation models, have received less attention in the Republic of Korea. The component projection approach, although it has been employed with national data, is in some respects unattractive because of the difficulty of directly including a measure of the influence of socio-economic factors. This

---

7 On this subject, see S. H. Han, The Study on Induced Abortion (Seoul, Korean Institute for Family Planning, 1973).

8 J. M. Yang, loc. cit.
method suffers from the same problem as the procedures discussed earlier. The transformation of the Republic of Korea over the past 15 years makes the inclusion of socio-economic factors and some estimate of the impact of their change on fertility an important element in any analysis of the change in fertility.

The modes of analysis developed by Potter and Wolters are useful in the Republic of Korea because they allow the analyst to maximize the information obtained from a series of surveys carried out at regular intervals since the early days of the national programme. In addition, this approach appears to be better suited than others for handling the impact of socio-economic factors independent of the family planning programme.

Like the CYP approach, those developed by Potter and by Wolters estimate programme impact directly from the number of acceptors. This is a limitation since the data available are known to be faulty with respect to both the number of acceptors of each programme method and their continuation. A special advantage of this approach is the fact that it incorporates a way of dealing with non-programme contraception. This factor is important in the Republic of Korea, where the combined influences of economic development, rising personal income and urbanization have increased the private market for contraception.

Regression analysis is another approach that is useful for the analysis of changing fertility in the Republic of Korea. A major problem that has limited the application of this method is the lack of relevant data. New efforts have recently been initiated to collect longitudinal measures of both programme activity and socio-economic changes for small areas. It appears likely that before long researchers will be able to provide new insights into the factors that have influenced the fertility decline, especially in the rural areas of the country. A particular advantage of the regression approach is becoming clear even during the preliminary stages of analysis. Because of the variety of measures of social and economic change that can be calculated from the data collected by various government agencies, it appears possible not only to estimate, albeit in a crude way, the impact of social and economic changes but to differentiate between the impact of socio-economic changes of various types (as indicated by different measured variables).

The last approach covered in the background papers, simulation models, has been employed only in a limited way. Mode and Littman provide a useful analysis, although limited because of its inability to take account of the complexity of factors influencing fertility in the Republic of Korea. The model does incorporate sex preference as a key variable determining the use of contraception. In this sense, it represents an improvement over the techniques that either cannot handle attitudinal factors or ignore their impact. Given the current levels of interest and technological skill it is unlikely that simulation models will develop rapidly within the Republic of Korea as a means to measure the impact of the family planning programme.

**SUMMARY AND CONCLUSIONS**

All of the methods to be discussed by this meeting could be usefully employed in the context of the Republic of Korea. However, because of the rapid development of the country, it is most desirable that procedures capable of explicitly treating the impact of socio-economic changes be developed and applied.

The complexity of the interaction between programme and non-programme factors remains the most significant barrier to a completely satisfactory estimate of the impact of the family planning programme in the Republic of Korea. Sophisticated experimental designs can measure fertility decline and programme impact in a precise way. However, national programmes in family planning and development cannot be experimental. One must settle for approximations often obtained through indirect observation. It will not be possible to solve the "substitution" problem although it is hoped that it can be dealt with in a more satisfactory fashion. (One advantage of the analysis of the reproductive process and of various regression procedures is that they represent improved ways of dealing with this "substitution" problem.) In the Republic of Korea, it is known that acceptance and use through programme channels has been substantial and that there has been a substantial fertility decline due in large measure to a reduction in marital fertility. The problem of how to relate these two facts deserves more attention.

Two additional problems should be mentioned in closing. The measurement of the impact of family planning on fertility must be done in a way that makes sense to administrators and others. Rapid feedback to administrators and those voting on programme budgets is very important. Likewise, being able to apply methods quickly and within the context of limited resources is extremely important. One should not concentrate on measurement problems to the extent that one forgets about programme workers and the importance of the programme impact for management and over-all social and economic development.

---


11 C. J. Mode and G. S. Littman, *op. cit.*
NEEDED RESEARCH FOR MEASURING THE IMPACT OF FAMILY PLANNING PROGRAMMES ON FERTILITY

W. Parker Mauldin*

Family planning programmes probably have been more carefully studied and examined than any other large-scale programme effort in history. During the past 10–15 years, a great deal has been learned about such programmes, and in the process, social science research has become more sophisticated as well as contributing to the improvement of programmes. The following discussion seeks to identify some of the more important research needs related to measuring the impact of family planning programmes on fertility. A question is also posed whether enough is currently known about potential fertility so that it can be set aside as an issue.

RESEARCH NEEDS

The substitution effect

Since the earliest days of discussion and analysis of the effects of family planning programmes on fertility, there has been recognition that:

(a) Users of fertility control methods prior to initiation of a national or large-scale family planning programme sometimes switch to the family planning programme for supplies and services. Such a switch could involve adoption of:

(i) A less reliable method, though that is thought to occur rarely; or
(ii) The same method because programme supplies are more convenient or less expensive; or
(iii) A more reliable method.

In spite of recognition of this type of substitution and a considerable amount of work that has gone into the problem, no really satisfactory accounting procedure has been developed to allocate appropriate proportions of such acceptors to the programme;

(b) A closely related problem is how to treat non-users at the start of a programme who subsequently have the number of children they desire and then accept a programme method. The question is how one estimates the proportion of those persons who would have adopted a fertility control method regardless of whether there had been a programme, and at what point in time they would have become users. The programme could, for example, influence one to become a user months in advance of what would have happened in the absence of a programme. One may then ask whether these programme and non-programme effects can be allocated;

(c) Another similar problem concerns the way in which to treat programme users who discontinue programme use but continue the use of fertility control. Such use is sometimes called the "extended use-effectiveness" of a given method (and sometimes is calculated on the basis of the first method adopted even though the person switches to another method);

(d) There is the reverse problem, or what might be called the catalytic effect of a programme. Many observers note that some national and large-scale family planning programmes create an awareness of family planning through their information and education programmes and through the activity of field-workers. Partially as a result, many persons adopt fertility control methods but do not become a statistic in the programme because their adoption takes place outside of the formal programme itself.

The two principal approaches to these problems, particularly to those outlined in (a) and (b), have been to question a subsample of acceptors within the family planning programme as to whether they were using fertility control methods before coming into the programme, and, if so, what method. This information permits an analyst to subtract such cases from the total number of acceptors and to exclude them in calculating births averted by the programme, though this is rarely done. Another method is regression analysis, which presumably takes account not only of users prior to initiation of the family planning programme but of non-users at the beginning of the programme who subsequently accept a programme method but would have accepted such a method in the absence of such a programme.

Fertility as a function of development

Perhaps the major theme of the World Population Conference held at Bucharest in 1974 was the interrelatedness of fertility and development, and the World Population Plan of Action recommends that "countries wishing to affect fertility levels give priority to implementing development programmes and educational and health strategies which, while contributing to economic growth and higher standards of living, have a decisive impact upon demographic trends, in-

cluding fertility. This is a needed area of research, and the guidelines for such research are spelt out in paragraph 32 of the Plan, which indicates that specified development goals, enumerated in that paragraph, generally have an effect on the socio-economic context of reproductive decisions that tend to moderate fertility levels.

Similarly, there is a need for research to determine the extent to which fertility is affected by family planning programmes in different socio-economic settings. Ideally, one needs to know the effect of a specified type of family planning programme administered with a specified intensity in different settings. More conventionally, one needs to know what the effect will be on fertility, within a specified period of time, of a family planning programme costing X per capita, at different levels of socio-economic development, assuming that: (a) all other factors are constant; or (b) socio-economic development continues at the same pace as in the past; or (c) it proceeds at some other specified rate. Research along these lines is particularly pertinent at this time because the United Nations Fund for Population Activities (UNFPA) is considering what strategy it should apply in allocation of its funds and basically is considering allocations on the basis of need within a country rather than on the basis of cost effectiveness or the ability of the county in question to use such funds effectively.

**Impact as a function of availability**

To what extent is the success of a family planning programme affected by the availability of different fertility control methods to large segments of the population? About two years ago, the present author asked a number of colleagues and associates to rank all of the countries with national family planning programmes on the extent to which various methods of fertility control were generally available to the population. Five classes of methods were used: conventional methods; orals; the intra-uterine device (IUD); male and female sterilization; and abortion. The method of rating availability of methods was subjective and crude, but there was, none the less, a remarkable relationship shown between the number of methods generally available and the extent of change in fertility by country. Observers differ on their evaluations of how much difference it would make if fertility control methods were generally available to the population, but the strategy of the largest single donor in the field is predicated on the assumption that the more generally available fertility control methods are, the more likely fertility is to decline, and, as a consequence, a great deal of money is being invested on this important assumption. Research is needed as a guide to Government and donors in the allocation of resources and the development of programme strategies.


**When to stop**

A much debated issue is how far a family planning programme will take one, and related to this issue is the question when such a programme becomes self-sustaining. It is in the self-interest of a Government and of the donor community in general to know when to stop investing in family planning programmes. Donors typically decide that they will no longer support activities in a given country when that country has reached a certain **per capita** gross domestic product, and in the family planning field, when it has developed a moderately mature programme and the growth rate has declined to the order of 2 per cent. This writer has suggested to programme personnel in one of the more successful programmes that an experimental design should be worked out and that programme activities should be stopped in different socio-economic settings, and the results measured over a period of time. Results should be measured, of course, in terms of changes in fertility, which is the crucial question with reference to the impact of family planning programmes designed to reduce rates of population growth. In addition, however, the assessment should take into account the extent to which underprivileged persons are either denied access to fertility control methods because of the change in policy or are faced with an undesirably heavy burden because of either the cost or the inconvenience of obtaining services and supplies. There are ethical issues involved here as well as demographic and programmatic issues, but it is the present author’s assumption that wherever family planning efforts have been moderately widespread and successful, the abandonment of a programme, or the application of welfare criteria, would not mean cessation of supplies and services to that population. If such a programme were undertaken, it would be important for those involved to ensure that supplies and services should be available in the private sector, possibly including the provision of some types of contraceptive at subsidized prices through community distribution schemes or other more conventional commercial outlets.

**Validation of births averted**

Is there a gap, an inconsistency, between the reported number of users by method, continuation rates, the calculation of births averted, and actual changes in and levels of fertility? There has been a marked improvement in models relating to the calculation of births averted in recent years, but it is notable that each of these models, in effect, pays little or no attention to coitus interruptus, abstinence or rhythm. Each, in effect, assumes that a discontinuer is returned to the pool of non-users, but it is known from a number of studies that such persons have very different characteristics than never-users and that they more effectively control their future fertility than do never-users. Also, a recent study indicates that with current contraceptive methods it is not possible to attain a very
low crude birth rate without rather widespread use of abortion. This latter conclusion has a number of implications, both ethical and programmatic. Accordingly, it is suggested that this area deserves considerable research attention.

One of the major areas that needs investigation is the definition of use of a fertility control method. If a person has an IUD inserted or is sterilized, clearly that person is a user of a fertility control method; but if a person accepts a gift of pills or condoms or even purchases those contraceptives, the individual is not necessarily a user. Surveys ask some questions relating to use, but this is a sensitive area and has not been widely researched. Some little attention is needed in this area. As was mentioned above, not much attention is paid to *coitus interruptus* or abstinence and relatively little to rhythm, to breast-feeding and perhaps also to other practices.

**Biological determinants**

Mahalanobis wrote in the foreword to *Couple Fertility* that malnutrition was sufficiently widespread in India in the mid-1950s that one could expect a rise in fecundity with an improvement in economic conditions and that that increase would tend to offset the effects of a family planning programme. This thesis is currently receiving increasing attention, particularly by Rose Frisch and her colleagues, but also by the World Fertility Survey, which has developed a module in this field. The basic argument is that in many countries of the world malnutrition is so widespread that fecundity has been appreciably impaired even though fertility rates are moderately high. The expectation, according to this thesis, is that as food supplies become more plentiful and more equitably distributed, fecundity will be less impaired and fertility will increase. In most of the countries where malnutrition is widespread, there are national family planning programmes and the effect of such an improvement in nutrition could be to mask the effect of the national family planning programme which is in the direction of lowering fertility. Consequently, a country could come to the conclusion that its programme was being ineffective when, in fact, it was reducing expected fertility appreciably.

**Potential fertility**

Potential fertility is listed here as an area needing research although a colleague of the present author states that potential fertility is understood rather well now and that the problem is that the analyses do not sufficiently take into account what is known. The background paper on problems and issues (ESNP/AC.7/1) contains a good section on potential fertility, and it states that simulation studies have shown that chance factors in fecundity operating on a group with identical probabilities will produce a subgroup with above-average fertility which is thereby disposed towards acceptance. Also, there has been some work that shows that a tendency exists for high fertility to regress towards the mean (and excluding subfecundity, the reverse; i.e., for persons with low fertility to regress towards the mean and thus to have higher fertility). The work of Brass and Wolfers is noted,3 appropriately so at that point. In addition, the work of Sivin4 has refined this concept a little further, and he has shown the expected fertility of women who accept a contraceptive method during the post-partum period and subsequently discontinue a method. This analysis shows that it takes several years to return to "expected" fertility for a group of such women. There remains, however, a residual question whether there is sufficient understanding of potential fertility and how to analyze it; perhaps this is one issue that can be resolved at this meeting.

**Early indicators of change**

There is a very considerable lag between the occurrence of events, their collection and tabulation, and analysis. It would indeed be helpful if it were possible to establish a short-cut to the conventional method of registration of vital events and the subsequent tabulation of such events. The attention that is being paid to birth intervals is a move in this direction. The query is whether other measures could be developed, and with the application of several of those methods, whether one could tell at an earlier time than is currently the case if fertility had changed and, if it had, also could give an order of magnitude of that change.

**Validity of one-time surveys**

There have been many improvements in surveys during the past decade or two and some of the methods of analyzing such data have been markedly improved. The World Fertility Survey is the largest undertaking of its kind and it is based exclusively on one-time surveys. None the less, a question remains as to how valid these surveys are.

A number of studies undertaken a few years ago indicated that one-time surveys often, perhaps almost always, underestimated fertility levels, and as a con-

---


sequence, dual-purpose surveys were introduced. Indeed, the POPLABS project at the University of North Carolina (United States of America) was begun and is being carried forward because of the belief of a group of scholars associated with it that single-purpose surveys are typically deficient and unreliable. The National Sample Survey of India has instituted a dual collection system for vital statistics and those efforts clearly indicate that any one method by itself understates vital events appreciably. A question remains whether a more sophisticated analysis of the results would permit one to arrive close to the values produced with the dual system, but, to the best of this author's knowledge, such investigations are not being carried out.

There is the intriguing likelihood that the World Fertility Survey will show a remarkable decline in fertility because of two factors. First, historically one-time surveys have understated fertility; and, secondly, the Potter or domino effect results in a heaping of births about five years previously and gives an artificially high level for that period, but more important, a lower figure for the current period, thus indicating a recent decline in fertility. At a minimum, a serious analytical work is needed to analyze results in recent years and to acquaint scholars in the field both with the results and, to the extent that those results indicate that one-time surveys are satisfactory, with the specific analytical methods that should be used in the application of such surveys.

Satisfying demand

Most of the research relating to the impact of family planning programmes has concerned itself with the extent to which stated targets have been met or the extent to which fertility has fallen. More important is the extent to which family planning programmes have provided information, supplies and services that are needed and wanted by the population. As a part of a national family planning programme, a real effort should be made to determine what couples want, what women want and the extent to which those needs and wants are being met. A part of this relates to the extent to which coverage is being adequately supplied, including the location of facilities within some convenient time-distance frame, and also to the extent to which information and services are provided reasonably promptly and courteously. Cost also is an important factor. There is a tendency to say that the large-scale family planning programmes have invested so much in time, money and personnel that enough is being done in this direction; in fact, however, very little research has been undertaken to assess the extent to which such wants are being met. It is known that abortion is moderately widespread in many areas under appalling conditions, and this fact would appear to indicate that fertility control methods are in no sense generally available to the population. This is a major need, one on which very little research has been done to date.

Fertility of the entire population

When one focuses on measuring the impact of family planning programmes on fertility, there is a tendency to look only at acceptors within the programme and within limits this tendency is desirable. More important is what is happening to fertility in the population at large. If one assumes, as one must, that the family planning programme is only one element, albeit an important factor, in fertility reduction, it is important to know what is happening to fertility as a whole because presumably not only the family planning programme is working but social and economic development is occurring, and there is an important interaction between the two. If one looks only at the impact of the family planning programme, or rather looks only at acceptors within the programme, one may well be missing an important element of the programme effect, which is the interaction of the programme with social and economic factors.

Data needs

Because of the length of the foregoing section, the author had intended to omit any specific references to data needs other than those implicit in research needs, but two particular needs should be identified:

(1) Improvement in vital statistics. It is recognized that the improvement of vital statistics is a long-run process and perhaps because of this aspect, the major actors in the field have done virtually nothing during the past 25 years to improve vital statistics. The United Nations system is, in this writer's opinion, regrettably deficient in this area. UNFPA is putting some funds into the general field, although this writer considers that a questionable use of UNFPA funds. The Statistical Office of the United Nations states that the improvement of vital statistics is important, but it attaches very low priority to this activity. An independent effort is being made to organize an international vital statistics group to focus attention on this area and perhaps this is the best way to proceed. In any event, there is a very great need here and very little attention is being given to it;

(2) Five-year censuses. In the absence of good vital statistics, a programme that takes censuses every five years would be an important addition to the current knowledge about population and population change. Censuses are, of course, relatively expensive, but they have the virtue of requiring little or no foreign exchange if the countries involved feel a need for such data. It is suggested that a programme to encourage countries to take censuses every five years, particularly those with poor vital statistics, could result in a marked improvement in the knowledge about population change.
COMPONENT PROJECTION VERSUS OTHER TECHNIQUES FOR ASSESSING PROGRAMME ACHIEVEMENT TOWARDS A TARGETED FERTILITY REDUCTION

Robert G. Potter*

A number of the countries with national family planning programmes have set demographic goals, usually expressed in terms of specified reductions of either the crude birth rate or the growth rate within a given target period. In this context, there is likely to be not only an interest in the impact on fertility of past programme effort but a need to estimate the required acceptances of programme methods necessary to realize the current demographic goal. A variety of techniques have been utilized, conditioned heavily by the data and analytical skills locally available, but most of which represent varying complete realizations of what in the background paper on problems and issues (ESAP/AC.7/1) is called the component projection approach (CPA). It is argued here that CPA does possess special advantages for target setting and that, on balance, none of the other approaches listed in the background paper are quite as appropriate.

ADDITIONAL WORK ON THE COMPONENT PROJECTION APPROACH

Besides the applied work recapitulated by Ross, there has been extensive work systematizing CPA since the pioneering effort of Lee and Isbister. A project of the Economic and Social Commission for Asia and the Pacific (ESCAP) led to the development of a computerized component projection scheme (PROJ5), which was applied to Pakistan. Meanwhile, Nortman and Bongaarts were designing their first uncomputerized version of TABRAP, which for the first time offered a direct rather than a trial-and-error means of deriving acceptor targets. In a subsequent ESCAP project, the Multinational Study in Methodologies for Setting Family Planning Targets in the ESCAP Region, planned and at first directed by K. Srinivasan, TABRAP has been amplified into a nine-programme, documented computer package called the ESCAP Target-Setting System and applied to data from eight Asian countries. Another study presents a more theoretical treatment of CPA, based on stable population theory.

BASIC FACTORS

Three sets of basic factors have to be considered when bridging demographic goals and contraceptive targets. They comprise demographic setting, demographic goal and family planning programme factors:

(1) Demographic setting:
(a) Size of population;
(b) Initial age-sex-marital status distribution;
(c) Male and female age schedules of mortality and changes over the target period;
(d) Age-specific proportions of those married among women of reproductive age and their trend over the target period;
(e) Age schedule of marital fertility among non-users of contraception and changes over the target period;

(2) Demographic goal:
(a) Length of target period;
(b) Change in crude birth rate (or rate of natural increase) aspired to;

(3) Programme factors:
(a) Continuation and effectiveness of each method (programme or private sector) of contraception offered and the age distribution of the users of each method;
(b) Age of acceptors;
(c) Relative popularity of the different contraceptives, in each age class of acceptors;

* Professor, Department of Sociology and Anthropology, Brown University, Providence, Rhode Island, United States of America.


Logic of the Component Projection Approach

The logic of CPA may be illustrated by the two computer programs, TABRAP, which gauges acceptance requirements to realize a pre-designated crude birth rate path, and CONVERSE, which treats the converse problem of determining consequences of specified acceptance schedules on the crude birth and growth rates.

Both computer programs embody conventional component projection schemes that, utilizing five-year age classes and projection steps of one year, are elaborated to handle age schedules of acceptance and continuation of contraception. Only married women 15-44 years of age and not currently using a method are defined as eligible to accept.

In CONVERSE, births allowed by family planning are estimated in five main steps:

1. Partition acceptance into private and public sectors and within each sector by age class and contraceptive method;
2. Convert acceptance into woman-years of protection classified by age of user;
3. Multiply woman-years of protection by age-specific potential fertility rates of users, which products, appropriately lagged and summed, yield annual births prevented;
4. Subtract births prevented from "potential" births, i.e., births expected without family planning, to obtain births allowed.

Two projections are always computed: one based on allowed births, and the other based on potential births.

In TABRAP, the problem of deducing acceptance requirements to achieve a pre-designated crude birth rate path presupposes a number of constraints to make the solution direct rather than iterative. The constraints chosen for TABRAP relate to the relative popularity of programme and private sector methods within five-year age classes and to relative age-specific dispositions to accept among eligible women. Analysis proceeds year by year. The sequence of steps is as follows:

1. For any programme year \(t\), determine total births to be prevented in order to realize the crude birth rate targeted for that year;
2. Determine births already averted by users who accepted during programme years \(t - 2\) or earlier;
3. Difference these two quantities as an estimate of births to be prevented in year \(t\) by new acceptors in year \(t - 1\);
4. Convert this last quantity into corresponding woman-years of use;
5. Transform this protection time into required numbers of new acceptors classified by age and method.

Assessment of the Component Projection Approach

In favourable circumstances of data and computerization, the strengths of CPA are its ability to encompass the factors enumerated above, its provision of detailed output and its being cheap enough in operation to run many projections, when desired. By recourse to the programming technique of overlaying, either TABRAP or CONVERSE can be fitted onto a 32K computer.\(^6\)

Several weaknesses exist, of course. First, data requirements are heavy. Especially problematical are the estimation of continuation rates and the age-specific potential fertility rates of future users and married non-users, these latter two sets of quantities being incapable of direct observation. Secondly, CPA operates at the level of births, not pregnancies, making it awkward to incorporate induced abortion. In some manner each abortion has to be made equivalent to an appropriate amount of contraceptive protection time.

Thirdly, CPA shares with several of the other techniques listed in the background paper (ESAP/AC.7/1) a lack of any systematic procedures for estimating indirect effects of family planning programme on private sector birth control or for assessing to what extent programme contraception is substituting for fertility control that would have occurred anyway. Fourthly, if data on the private sector are missing, then one is hard put to apportion roles for private sector and programme; and even if one succeeds in contriving an estimate of comparative contributions at the beginning of the target period, their evolution during the remainder of the target period remains necessarily speculative.

**Comparative utility of other techniques for target setting**

Most of the techniques listed in the Secretariat background paper (ESAP/AC.7/1) are designed to evaluate the impact of family planning on fertility for purposes other than target setting.

For example, standardization is typically employed to resolve an observed past change of fertility into components attributable to concurrent changes in age structure, proportions married and marital fertility. The part played by family planning in the last-mentioned change is not rendered explicit.

Because trend analysis (the projection approach) is directed towards estimating "net" effects of programme activity, it has to be speculative; and the interactions of contraceptive acceptance, use and potential fertility within and between private and public sectors are left implicit. Though subject to problems of matching and contamination, the approach of experimental designs has the advantage of offering direct estimates of net programme impact. Nevertheless, a comparison of the fertility change of a non-random, self-selected experimental sample with that of a purposedly matched control sample is hardly the same as measuring the fertility consequences of a programme for an entire population or its acceptance requirements to attain a prescribed demographic goal.

What is called the analysis of the reproductive process constitutes a specialized and elaborate tool for assessing births averted per segment of intra-uterine devices (IUD) and, less conveniently, of other contraceptives whose practice is less continuous. The high data requirements of this methodology rule out application except in rare instances.

The couple-years of protection (CYP) approach represents a set of rules for converting the acceptances of a mixture of contraceptives into an absolute number of births averted, but the technique lacks age specificity and fails to locate births averted in time. Moreover, the conversion rules, derived originally for Pakistan, lack general applicability to other cultures and times.

If family planning service statistics are available but only limited demographic data exist, appeal may be made to linear regressions predicting the crude birth rate or growth rate as a function of the proportion of married women of reproductive age using contraception. However, a regression crudely summarizing the experience of many countries cannot be expected to reflect the idiosyncrasies of a particular country. A much more powerful regression analysis becomes possible when comparable family planning and demographic data are available for subareas of a country. The implicit interaction among acceptance, use, and potential fertility and apportionment between private and public sectors is an appropriate aspect, namely, that belonging to the immediate past of the country. How relevant the resultant regressions are for target setting depends upon the indexes available as independent and dependent variables. Unfortunately, few countries have the data to support this type of analysis.

Most simulation models of the family-building process are disqualified for target-setting relevance by virtue of representing cohorts instead of period populations. Exceptions are the micro-simulation model POPSIM and the stochastic population model. As family-building models operating at the pregnancy level, both models are better able to handle induced abortion than is CPA. With a monthly time unit rather than a year, they can be more precise about the timing of events. However, data requirements as well as development and operating costs are much higher. The latter are so high, in fact, as to be incompatible with the substantial number of runs typically necessitated by the several objectives outlined above. Nor does the extra precision really justify the added costs inasmuch as the basic weaknesses of CPA—estimation of continuation rates and of potential fertility along with apportioning shares to programme and private sector—are not being solved by the simulation models.
A NOTE ON TWO APPLICATIONS OF MICRO-SIMULATION MODELS TO THE PROBLEMS OF EVALUATING FAMILY PLANNING PROGRAMMES

Jeanne Clare Ridley*

The purpose of this brief note is to report on two recent applications of micro-simulation models to methodological problems encountered in the evaluation of family planning programmes. Not only do these applications illustrate the types of methodological questions that may be investigated but they are suggestive of a wide range of questions of a methodological nature that can be investigated with such models. Further, the results of these applications indicate that the more conventional fertility measures may be more adequate for assessing changes in fertility than is commonly assumed.

The two micro-simulation models, POPREP2 and REPSIM-B, employed in the two applications summarized here, have been described in detail elsewhere. Both models employ Monte Carlo methods. Briefly, POPREP is an elaboration of POPSIM4 in that it expands considerably the fertility component of POPSIM by including a number of the basic biological determinants of human reproduction. POPREP generates the reproductive histories of a hypothetical female population of all ages. The model provides for the demographic factors of marriage, marital dissolution, remarriage and death. The biological factors included are fecundability; various outcomes of pregnancy, including induced abortion; the post-partum non-susceptible period; and sterility. In addition, it also provides for the use of contraception, surgical sterilization and induced abortion.

REPSIM-B is an extension and elaboration of REPSIM-A. REPSIM-B simulates the detailed reproductive history of a hypothetical cohort of women. It provides for a woman marrying, dying, becoming sterile and becoming pregnant, and for varying outcomes of pregnancy including induced abortion. Provision is also made for the adoption of contraception or surgical sterilization.

The specific objective of the POPREP application summarized here was to explore the effects of different patterns of contraceptive use on alternative measures of fertility. The measures studied were age-specific birth rates, the total and general fertility rate and various measures of birth intervals.

The initial population assumed in the simulation was derived from the census data for India in 1961. Assumptions regarding marriage, mortality and other parameters were assumed to be similar to those characterizing Indian women. For example, over half of the women were assumed to marry by age 16; life expectancy at birth was assumed to be approximately 50 years; the mean length of the post-partum period was assumed to be about 12 months; and fecundability was assumed to vary between women and by age with a mean maximum level of 0.20.

The initial population was first simulated for a period of five years. During that period, no use of contraception was assumed. Beginning in the sixth year, three different patterns of contraception were postulated. Acceptance of contraception was assumed to vary with age and with parity. The first pattern of contraceptive use assumed a moderate rate of acceptance of contraception with a contraceptive effectiveness of 100 per cent. The second pattern had higher acceptance rates and also had a contraceptive effectiveness of 100 per cent. The third pattern postulated the high acceptance rates with contraceptive effec-

* Professor of Sociology, Kennedy Institute, Center for Population Research, Georgetown University, Washington, D.C.


2 A manual describing the computer program of POPREP has recently been completed; see User's Manual for POPREP and INTRVL (Chapel Hill, N.C., University of North Carolina, Department of Biostatistics, 1975).


6 This summary based on C. M. Suchindran and others, Sensitivity of Alternative Fertility Indices, final report on contract No. NICCH-2187 (Chapel Hill, N.C., University of North Carolina, 1976).
tiveness assumed to be only 90 per cent effective. The simulation for the three patterns of contraceptive use were carried out for 10 years.

Not surprisingly, for the entire 10-year period of contraceptive use all the fertility measures reflected the predicted impact of the three different patterns of contraceptive use. As stated by Suchindran and his colleagues, however: "What is more surprising, and perhaps disquieting, is the fact that, when these indices are calculated at earlier time points, they fail to indicate the relative effectiveness of the three patterns, and, in some instances, fail to provide conclusive evidence of a change in fertility." Only after the third year of contraceptive use did any of the various measures begin to reflect accurately the various patterns of contraceptive use. The measures most sensitive to showing change at the earliest points were the age-specific fertility rates, general fertility rates and total fertility rates. The open birth interval showed little change in the early years of contraceptive use although it did show considerable change towards the end of the 10-year period of contraceptive use. Closed intervals were not expected to show any change when contraceptive effectiveness was 100 per cent; they did not, however, show much change when contraceptive effectiveness was only 90 per cent. Life-table estimates of birth intervals were striking in that they showed little or no change.

Supporting the above-mentioned results as to the usefulness of the more conventional fertility measures are results obtained with an application of REPSIM-B. The original purpose of the particular experiments by Ridley and Clague that are summarized here[^1] was not to assess the usefulness of various fertility indexes but rather to assess various methods of estimating births averted. Using data from two cohorts simulated under conditions characterizing women in India and in Latin America, the methods of estimating births averted developed by Lee and Isbister, by Potter and by Wolfers were evaluated against criterion values derived directly from two simulated cohorts. In the first cohort of each experiment, no contraceptive use was postulated; and in the second cohort, contraceptive use was postulated for a period of five years. Acceptance of contraception was at a moderate level and depended upon desired family size; contraceptive effectiveness was assumed to be 95 per cent.

Although the results obtained are limited by the experimental conditions assumed, the simpler method first developed by Lee and Isbister, which depends upon estimating age-specific birth rates, reflected most accurately changes in fertility resulting from contraceptive practice. The results also indicated that the effects of contraceptive use in a particular year were not clearly reflected in any of the estimates. The estimates for the total five-year period were considerably better.

Although the results obtained in these two applications must be viewed most cautiously because of the restricted assumptions under which these investigations were carried out, they do suggest that future investigations might be profitably pointed in certain directions. First, more attention should be placed on studying the usefulness of more conventional fertility measures in evaluating fertility changes. The development of better data for estimating these measures, therefore, becomes of highest priority. Secondly, more investigations of the sensitivity and robustness of various fertility indexes are needed, as are studies of their sampling variability. Simulation models afford an important tool for carrying out such investigations.

[^1]: Jeanne Clare Ridley and Alice S. Clague, "Experiments with a microsimulation model of human reproduction" (forthcoming).
SOME ISSUES IN DETERMINING APPROPRIATE METHODS OF EVALUATING THE FERTILITY IMPACT OF FAMILY PLANNING PROGRAMMES

Ismail Sirageldin*

The purpose of the exercise is to choose an index of family planning programme efforts; to choose an index of fertility; to measure changes in both indexes over a given period of time—usually, five or fewer years; to choose an appropriate theoretical frame and an appropriate methodology to relate family planning programme efforts to changes in estimated fertility after allowing for the effects of all other relevant factors, be they individual, group and/or societal; and, lastly, to make evaluative statements about the fertility impact of the family planning programme. This three-way relation is schematically illustrated below in its more general and gross way.

As is well known by now, behind this seemingly simple diagram are some of the most complex theoretical and methodological questions, many yet unresolved in the social and behavioural sciences. Each of the two boxes in the left-hand side of the figure includes a large set of factors for which there does not yet exist an established body of theory about their separate and/or combined effects on fertility, especially in the context of developing countries. Space does not allow a review of current knowledge. However, there are several recent surveys available. This state of affairs undoubtedly increases the probability of introducing specification errors (not only of missing variables but of functional forms) when developing statistical models to estimate the basic structural parameters of the system. In the case of regression analysis, the exclusion of relevant variables from the regression may be a very serious error since the estimated coefficients may be seriously biased and the estimate of the residual variance could be biased upwards. This problem of specification error is further compounded by the related problem of measurement errors in some of the key demographic and socio-economic variables, mainly because of the lack of adequate data base, and accordingly analysts resort to inadequate proxy variables. The relevant questions appear to be: whether it is possible to develop some general criteria to rank the various evaluation approaches; and whether useful evaluation can be made without adequate specification and given the inadequacy of available data.

In the useful background paper prepared for this meeting, a survey of eight measurement methods was presented: standardization approach; trend analysis (fertility projection approach); experimental designs; couple-years of protection (CYP); component projection approach; analysis of the reproductive process; regression; and simulation models. Some of these measures vary widely in their theoretical foundation and specification details, in their levels of analytical sophistication and accordingly in the level of manpower skills required, their population coverage, their required data, their potential errors; and in the limitations imposed by their specifications on interpreting their derived conclusions and on their level of generalizations.

Ideally, in a cost-benefit framework, one should be able to say that for a given purpose method A is better (or worse) than method B if both methods produce the same service or “benefit” but the “cost” of A is lower (or higher) than that of B; or, alternatively, if both incur the same “cost” but A produces more (or less) of the same “benefit” than B. What one has, however, is a multiple-cost criterion and a multiple-benefit criterion that cannot be reduced to a common criterion without having a common yardstick for measuring the various cost and benefit elements. It is evident that in order to arrive at a rational criterion for choice among these alternative methods, one needs, first, to enumerate the various elements of these costs and benefits or at least the important ones; and, secondly, to have (or to develop) a system of weights assigned to

---

* Professor, Population Dynamics and Political Economy, The Johns Hopkins University, Baltimore, Maryland, United States of America.


3 “Methods of measuring the impact of family planning programmes on fertility: problems and issues” (ESA/PAC.71/1); see part one of the present publication.
these various elements. It is argued below that the assignment of these weights, and to some extent the full enumeration of the cost and benefit elements, cannot be developed independently of the "purpose" of evaluation; and that that "purpose" will vary greatly, depending upon the type and the life cycle of the programme under consideration, i.e., upon the intended use of the evaluation findings. Lastly, an attempt is made, for purpose of stimulating discussion, to develop a simple and preliminary typology that relates programme objectives, evaluative needs and appropriate methods of evaluating the fertility impact of family planning programmes.

The "total cost" of a given method could be conceptualized as the market or imputed value of all inputs required to produce an "estimate" of a given precision where the cost elements include:

(a) The type of data required—whether they are already available (e.g., service statistics or census data) or must be produced independently;

(b) The time needed to produce such estimate (of a given precision);

(c) The level of technical skills required relative to the available pool;

(d) The alternative uses of the data and analyses produced—comparability with other estimates in terms of scope and coverage, its future use or its utilization by other governmental agencies. Such alternative uses might justify relatively large overhead costs.

Clearly, the type and magnitude of these cost elements will depend upon the type of "estimate" required and upon its desired level of precision.

There are various reasons for examining the relationship between family planning efforts and changes in fertility. For example, the focus might be on one or more of the following purposes:

(a) To examine the "aggregate" effect of the programme on "aggregate" fertility changes; this objective is usually applicable when evaluating the total investment of a national effort;

(b) To examine the "aggregate" effect of the programme on "differential" fertility, where the focus of the programme is on serving special target population groups, e.g., high parity, rural, poor or teen-agers;

(c) To examine the interaction between programme and non-programme factors on fertility in order to develop optimal population strategies; this emphasis becomes important especially where there are rapid social and structural changes (e.g., Venezuela, Iraq and Iran) or random shocks (e.g., Bangladesh) which have independent influences on the basic determinants of fertility. An important example would be the interaction between health, nutrition and fertility;

(d) To examine the relative effects of individual programme components or strategies on fertility, as, for example, trying a new contraceptive method or a new component in the delivery system, or assessing whether the effectiveness of a given method(s) has changed over time.

Furthermore, some countries (e.g., Egypt) have broadened the scope of their family planning programme to include inputs that influence fertility indirectly, e.g., influencing age at marriage; female labour force participation, education and literacy; or rural cottage industries, which makes it harder to define programme factors and to measure their independent influence on fertility, mainly because of the lack of adequate understanding of many of these linkages. It becomes difficult to identify "programme acceptors", especially through service statistics.

At one time or another during the life cycle and development of a programme, its organizers or planners will need to have answers to these various questions.

At the early stages of a programme, information about the relative effectiveness of the various methods is needed both for initial planning or target setting and for ongoing evaluation of performance. Probably the CYP method or analysis of the reproductive process, both of which focus primarily on programme acceptors and are based on service statistics and some economical retention surveys, would be appropriate at the initial stages of programme operations and when programme acceptors are reaching the anticipated target. But for adequate specification, data are not usually immediately available.5

However, when acceptance rates are very low and continue to be low, the efficiency of these techniques appears to be reduced. In such cases (e.g., Bangladesh and Pakistan), what is needed is to understand why programme factors do not influence fertility. Information about "non-clients", what determines their fertility norms and behaviour, becomes of primary focus. For that purpose, some theoretical frame that includes the socio-economic and biological determinants of fertility should serve as a guide for this type of research. Multivariate analyses (e.g., regression analysis, including path analysis; automatic interaction detector technique) appear to be needed for this type of approach, and probably survey samples, if adequately

---

4 See, for example, W. Butz, "Socioeconomic and biomedical analysis of nutrition, birth interval and infant development", Santa


designed, could be an efficient way of collecting the relevant data.\(^6\)

When the time period of evaluation is relatively long, say, 10 years, and accordingly the expected change in fertility is relatively large, then the standardization approach along with a multivariate analysis that controls for the main socio-demographic variables might be appropriate, especially when significant changes in these variables have also occurred.

To summarize the discussion thus far, appropriate methods will depend upon programme needs and development, or, more generally, upon the way in which the findings will be utilized. It appears that evaluating the fertility impact of a family planning programme should go beyond aggregate relations—it should identify pockets of high fertility in society and their response to programme factors. It should also examine whether short-run apparent changes are delayed pregnancies or are there to stay. These are questions that cannot be answered in the abstract but need specific reference-setting (e.g., teen-age pregnancies in Jamaica, the health-nutrition-fertility relationship in Bangladesh, the State/method differential in India).

When comparing the relative effectiveness and utility of a given method, the objectives of evaluation should be indicated, for example, whether its purpose is for short-run management use (1–4 years) or for long-run planning use (say 5 or more years); and also whether the focus is on the relative efficiency of programme input or on the effectiveness of programme operation. Furthermore, the purpose of evaluation should be specified with respect to the objectives of the programme in terms of family planning only or other objectives, such as family health and welfare. A schematic illustration of a possible classification is given below.

The difficulty of choosing an evaluation method is not unique in demographic analysis; as Jaffe noted “Our articles of faith—and the public and private policies and programs which express them—remain mostly unevaluated in any formal sense”.\(^7\) The case

---


---

### A POSSIBLE CLASSIFICATION OF EVALUATION OBJECTIVES, PROGRAMME OBJECTIVES AND CRITERIA FOR COMPARING METHODS OF MEASURING THE FERTILITY INPUT OF FAMILY PLANNING PROGRAMMES

<table>
<thead>
<tr>
<th>Objective of evaluation</th>
<th>Objective of programme and methods of measurement</th>
<th>Short-run/management (1-4 years)</th>
<th>Long-run/planning (5+ years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Input focus; efficiency (e.g., method, specific, clients’ age characteristics)</td>
<td>Population focus: effectiveness (pregnant/non-pregnant acceptors)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population focus</td>
<td>Alternative investment (e.g., cost/benefit)</td>
</tr>
<tr>
<td></td>
<td>Short-run/management (1-4 years)</td>
<td>Long-run/planning (5+ years)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Input focus; efficiency (e.g., method, specific, clients’ age characteristics)</td>
<td>Population focus: effectiveness (pregnant/non-pregnant acceptors)</td>
<td>Alternative investment (e.g., cost/benefit)</td>
</tr>
<tr>
<td></td>
<td>Some selected criteria*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feasibility variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data base</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Independence of variability, reliability, cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>

**Note:** Other considerations: 1. Age of the programme in relation to the level of participation; 2. Rate of structural change: e.g., industrialization, urbanization, health/mortality.

*Possible criteria to compare and evaluate the various methods of measuring the fertility input of Family Planning Programmes.

192
can be made, however, for population programme evaluation, but a balance should be made between perfecting a given technique and its utility. There is no question that the development of highly refined models for intra-uterine device (IUD) decay are useful if the IUD is highly accepted and could correct planners' conception about what parameters to use in their CYP, as was the case in Pakistan. In the latter case, however, it was only dealing with less than 1 per cent of the fertile women and the method acceptability was losing popularity. Analysis of the reproductive process, on the other hand, appears to have made an important contribution towards understanding changes in fertility, especially short-run changes and especially in interpreting socio-economic theories of fertility. Further empirical testing of this theoretical framework is needed (current attempts are being made in the Matlab region of Bangladesh and in Guatemala).
INTERACTION OF SOCIO-ECONOMIC CHANGES WITH FAMILY PLANNING PROGRAMMES: AN ASSESSMENT MODEL

K. Srinivasan*

It has been a frustrating experience for family planning administrators, demographers and social scientists alike not to be able to assess, with the desired level of precision, the impact of family planning programmes on fertility, especially in the developing countries where there is a pronounced need for such an assessment. The frustration arises in part from the inadequacy of the currently available models of evaluation for dealing with the combined effects of social change and family planning programmes on fertility, and in part, the paucity of reliable data in developing countries even for the application of simplistic methods of evaluation. In recent years, there has been a spate of publications on the methodological aspects, analytical problems, case study applications and methodological issues pertaining to this topic. One comprehensive review of the current state of the art in this field, with an identification of gaps in knowledge, has recently been published under the auspices of the International Union for the Scientific Study of Population (IUSSP).1 The major problems and issues in the evaluation of impact are also succinctly reviewed and discussed in the background paper (ESAP/PAC.7/1) prepared by the United Nations Secretariat for the meeting.

In the present paper, an attempt is made to identify and discuss some special methodological and data problems faced in the developing countries in the evaluation of the impact of family planning programmes in the context of rapid socio-economic changes and to develop an assessment model in order to cope with the problem.

As has been well recognized by now, the plethora of social, economic and communication development programmes undertaken in developing countries constantly interact with family planning programmes and lead to a variety of changes in knowledge, attitudes and behaviour related to fertility modification. Paucity of necessary data makes it difficult to quantify the changes as they occur even in the fertility levels of the population and almost impossible to isolate the net effects of family planning programmes after partialling out the effects of social and economic changes on fertility. The question arises whether even at a conceptual level, there exists what can be termed a family planning programme which can be identified as a separate entity from other developmental programmes, with separate effects on fertility that could be measured. This note attempts to highlight some crucial issues in this area of interaction, to develop a model for evaluation and to indicate a plausible method of compilation of necessary data for the application of the model.

CONCEPTUAL SCHEME

The demand for family planning services in a developing country is a function of a number of interdependent factors. These factors can be conveniently categorized under three major groups, recognizing that the groups are not necessarily mutually independent. The first category includes the family planning programme variables, such as inputs of money, men, materials and methods into the programmes, activities and systems of contraceptive delivery services; the second group includes attitudinal variables on people's perception of the programme, which is dependent upon the socio-cultural background of the population as well as upon the type of personnel employed in the family planning programme and the nature of facilities made available; and the third category includes the various socio-cultural and political variables in the population.

An operational definition of the family planning demand structure, which will facilitate quantification, can be made as "the proportion of eligible couples in the population at different stages of readiness to adopt family planning". From past experience with family planning programmes in developing countries, it has been found that in every community a certain proportion of eligible couples will adopt a modern method of family planning as soon as the services are offered to them; a certain proportion will do it when the programme is supported by their community leaders; a certain proportion when it is backed by a scheme of incentives and disincentives and a certain proportion only when there are changes in their socio-economic conditions. An ordered scale of readiness to accept family planning under different levels of stimuli can be considered and would form the basis for measuring the structure of demand for family planning. The three sets of factors mentioned above, which are not mutually

---

*Director, Population Centre, Bangalore, India.

independent, influence the demand structure which in turn affects the number of couples accepting family planning methods, their continuation rates and fertility change. A schematic diagram of the conceptual model of structure of demand for family planning services is given below.

Adopting the conceptual framework mentioned above, it follows that the demand for family planning in any society changes continuously, depending upon the interaction of the family planning programme variables and the socio-economic structural variables. Further, it is obvious that the demand at any time is also a function of the demand and contraceptive practices at earlier times because the potential couples available for fresh recruitment to contraception depends not only upon the total number but upon the proportion who have been recruited earlier. In such a framework, evaluation of the impact of family planning programmes on fertility should be stochastic in nature and cannot be deterministic wherein analysis is restricted to the situational position of the variables at any one point of time or at most some variables are measured as changes during a given period of time.

There are continuous changes both in the social structure and in the programme variables which have to be taken into account in any analysis of the impact of the programme. For example, in India, many traditional customs and practices have long operated as checks on fertility. Since the distant past, the taboos on sexual intercourse on selected days for religious or social reasons, the practice of the pregnant woman going to her mother's home for delivery and staying there for a number of months after delivery, the long periods of breast-feeding of children, and the social taboos on widow remarriage have all contributed, though indirectly, to keeping down crude birth rates to about 40, which is much below the biological potential. There has been a considerable variation even in the natural fertility, fertility in the absence of wilful contraceptive practices among different groups. During the past two decades, however, owing to the processes of modernization unleashed as a part of development strategy, many of the traditional checks have begun to be released and the potential fertility in the absence of contraception can be expected to increase in many groups.

The existing techniques for evaluation of the impact of family planning programmes do not adequately deal with this problem conceptually or analytically, though they recognize the existence of such interaction. For example, the standardization approach assumes that whatever differences in fertility level between two points of time remain as a residue, after standardization for certain structural factors, such as age-sex...

---

**Conceptual model of structure of demand for family planning services**

**Tentative scale of demand structure**

---

Note:

\[ X_1 + X_2 + X_3 + X_4 + X_5 = \text{total eligible couples}; \]
\[ X_1 = \text{currently contracepting couples}; \]
\[ X_2 = \text{couples likely to adopt family planning with modifications in the programme but without additional inputs}; \]
\[ X_3 = \text{couples likely to adopt family planning with additional inputs into the programme}; \]
\[ X_4 = \text{couples likely to adopt family planning with modifications in the programme and additional inputs into it}; \]
\[ X_5 = \text{never-adopters}. \]
marital status distribution and possibly rural-urban differences, is due to family planning programmes. Whatever change in fertility that cannot be accounted for, after consideration of a few factors for which data are available, is considered to be due to family planning. Similarly, in the fertility projection approach, the fertility in the absence of family planning programmes is projected on the basis of pre-programme levels or of some previous trend extrapolated. In the component projection or analysis of reproductive process methods, serious assumptions are involved regarding the potential fertility of acceptors, fertility that would have prevailed had they not come into the fold of the programme or into family planning. Such assumptions are generally made on the basis of the pre-acceptance fertility of the acceptors or a comparison of the fertility of acceptors with non-acceptors assuming that these differences would persist in future also. The method of computing the couple-years of protection (CYP) and converting to births averted, on the basis of the formula

\[ \text{CYP} = C \text{ births averted} \]

again ignores the socio-economic variables and their effect on potential fertility. Among the existing evaluation methods, the two which explicitly take into consideration socio-economic variables and their interaction with family planning programmes are experimental designs and regression analysis. In the experimental-design approach, two populations matched for a number of social and economic variables, including fertility and mortality levels, are selected, and one is exposed to the family planning programme and the other kept as a control. Here also, there is a tacit assumption that the future trends in socio-economic factors in both the populations are nearly the same and what accounts for the differential change in acceptance rates or the fertility levels during the period of study is the programme.

In regression analysis, fertility at any point of time or fertility change between two points of time in different areal units is considered to be a dependent variable and regressed on a host of exogenous variables, which include such socio-economic factors as literacy rate, rates of employment of women and mortality levels; such programme variables as family planning; and such input variables as family planning personnel per 1,000 population, number of clinics, money spent on family planning in each of the areal units etc. Areal analysis takes a geographical area, such as a county, block, district or state, as a unit for carrying out the regression or path analysis. The variance or differences that exist in the exogenous and programme factors within each areal unit and their association with fertility is ignored in the analysis. It has to be recognized that such fertility differentials on the basis of socio-economic groups are of fundamental importance in fertility change and averaging these differentials on the basis of areal units may vitiate the pattern of relationship. This possibility has been recognized in the earlier works on areal analysis.² In the following section an attempt is made to develop a research design which is focused on the measurement of interaction effects of social change and family planning programmes.

### AN ASSESSMENT MODEL

In any analysis of the effects of interaction of family planning programmes with socio-economic variables, the basic information that is needed is the probability of an eligible couple belonging to a particular socio-economic class, accepting a family planning method under a given régime of family planning programme. The variance in these probabilities has then to be divided into three portions, one due to socio-economic class, the second due to the family planning programme and the third due to the interaction between the two. A simplistic model of this procedure is presented below for the purposes of illustration.

Suppose that in any population at time \( t \), \( N_i \) represents the total number of couples in reproductive age group and that they can be classified into \( k \) socio-economic groups, which can be placed in an ascending order from 1 to \( k \). Let the number of couples in the \( i \)-th socio-economic group be \( E_{it} \). It is further assumed, for the sake of simplicity, that the population of couples in the reproductive ages is stationary, i.e., \( N_i \) is independent of \( t \) and equals \( N \). Let \( p_{ij} \) be the probability that an eligible woman in class \( i \) will accept a family planning method at time \( t \), and \( A_i(t) \) the total number of acceptors at time \( t \). It is further assumed that the proportion of a cohort of acceptors continuing the method at time \( t \) after acceptance can be defined by the function \( e^{-rt} \).

It has to be recognized that in the model, though the number of couples with wives in age group 15-44 remain the same over time, the proportion in different social classes may keep changing because of development and social mobility. Let the matrix \( M = \{m_{ij}\} \) denote the probability of transition from class \( i \) to \( j \) in any unit of time. Then, the following set of equations can be formulated to express the interaction.

\[
A_i(t) = \sum_i E_{it} p_i(t) \tag{1}
\]

where \( E_{it} \) denotes the number of couples at time \( t \) who are available for family planning acceptance because they had not accepted any method before or they had accepted earlier but discontinued in year \( t \) and are available for reacceptance in the same year. If data are available on the number of acceptors, \( A_i(t) \) in any year from any social class \( i \), then:

\[
\hat{p}_i(t) = \frac{A_i(t)}{E_{it}} \tag{2}
\]

and \( p_i(t) \) can thus be estimated for different \( i \) and different points of time or different areal units, and a matrix \( P \) of \( (\hat{p}_i(t)) \) could be formed.

An analysis of variance could then be performed in this $P$-matrix to determine what proportion of the variance in $p_i(t)$ could be attributed to the differences in social classes, what proportion is due to changes in programme inputs and the balance due to interaction. Such an analysis would be useful in determining the impact of interaction on family planning acceptance. On the other hand, if data are not available on $A_i(t)$ but only on the total number of acceptors $A(t)$, then $p_i(t)$ can be estimated indirectly adopting the following procedure.

Let the row vector $E^*(t)$ denote $(E^*(t), E^*_1(t), \ldots, E^*_k(t))$ giving the number of eligible couples in different classes in year $t$. Then it can be shown that:

$$E^*(t) = E^*_q(t-1)M + e^{-rt}(e^r - 1) \sum_{s=1}^{t-1} e^{rs}A(s)(M)^{t-s}$$

(3)

where $E^*_q(t-1)$ denotes a row vector $(1 \times k)$ of eligible couples not accepting any family planning method in time $(t-1)$ and $E^*_q(t-1)$ equals $E^*_q(t-1)(1-p_i(t-1))$ and $A(s)$ is the row vector $(1 \times k)$, $(A_1(s), A_2(s), \ldots, A_k(s))$

where $A_i(s) = E^*_i(s)p_i(s)$.

The first term in the expression for $E^*(t)$ is the number of couples who were eligible at time $(t-1)$ but did not accept the programme in $(t-1)$ and thus were eligible in time $(t)$; and the second term gives the number of couples who had accepted the programme earlier but discontinued during year $t$, making themselves eligible again. Now if $r$, $M$ and $E(t)$ are available and $A(t)$ is known for $t = t_1, t_2, \ldots$, equation (1) together with equation (3) becomes a system of non-linear equations in $(p_i(t))$. These could be solved and the matrix $(P)$ obtained. This matrix can be used for an analysis of variance to determine the effect of interaction.

The method given above is capable of being developed further, relaxing the assumptions of stationarity of the population. A preliminary application of the model is currently being attempted from the data collected through a survey in Karnataka State, India.
Identification and quantification of the effects of family planning programmes on fertility are difficult because it is usually impossible to take all related variables into account. In this respect, the problem of measuring the impact of family planning (or any other variables, e.g., economic developments) on fertility is analogous to measuring the impact of a health programme on mortality. In either case, one usually has only data from observational studies, such as census, vital statistics and surveys, rather than results of bona fide experimental studies in which subjects are randomly allocated to treatments.

Family planning programmes, at best, work in conjunction with numerous other factors to reduce and/or to increase fertility. In this paper, an attempt is made to illustrate the difficulties inherent in choosing models and to point out the need for improvement of the quality of data. Admittedly, the discussion cannot be presented in complete detail in a brief statement. Perhaps it will be sufficient to stimulate further discussion on the problems of decision making in the face of considerable uncertainty.

MODELS OF CAUSAL RELATIONSHIPS

Even if reliable measures of time trends in fertility levels and in the many variables that affect the fertility levels of a society or of a stratum or category within that society were readily available, demographers and other data analysts would have much difficulty in agreeing on models showing causal relationships between all of these variables and fertility. These difficulties exist whether the model relates to individual women or to groups, e.g., all women in a county. First, there would be some disagreement on how many variables and the level of detail at which they should be included in any model; and, secondly, there would be disagreement regarding the functional form(s) of relationships. (Note that the foregoing statement implicitly assumes that problems of definition and measurement do not exist although in the real world they are major deterrents to determining causes of fertility change. They are considered below in the section on data availability.)

The "intermediate variables" of Davis and Blake, which Freedman included under his classification one, may be used to illustrate the difficulties in specifying the level of detail required for independent variables in a model. This list of "intermediate variables" would include programmes as well as other means of limiting fertility. The intermediate variables may be classified into 11 groups as follows (by taking some liberties with the original):

A. Time spent in sexual unions by women:
   1. Age at entry into sexual union;
   2. Proportion never entering sexual union;
   3. Time spent outside union due to death of husband, divorce, separation or desertion;

B. Exposure to intercourse within union:
   4. Voluntary abstinence;
   5. Involuntary abstinence;
   6. Coital frequency (excluding periods of abstinence);

C. Conception variables:
   7. Fecundity or infecundity as affected by involuntary causes;
   8. Use or non-use of contraception;
   9. Voluntary fecundity or infecundity (sterilization etc.);

D. Gestation variables:
   10. Foetal mortality from involuntary causes;
   11. Foetal mortality from voluntary causes (abortion).

Continuing with the present assumption that reliable data are available, one should consider what level of detail one would use in an analysis. This issue would, of course, be related to the functional model postulated. Assume further that one has been asked to determine the magnitude of the effect of variable C (8) on fertility in country X at time t.

Since the above-mentioned data are available over time and in the detail required, it remains only to make three other general decisions regarding:

(a) The level of detail in the response or dependent

*Professor of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, United States of America.


variable(s), i.e., what measure(s) of fertility should be analyzed; whether it is appropriate to analyze the total fertility rate or should age-specific fertility rates be analyzed; whether age-parity specific rates should be analyzed instead; and how fertility measures from previous years should be handled, i.e., whether they are response variables or intervening variables. (The present writer thinks they are response variables, as is explained below.) Obviously, multivariate response variables are more difficult to analyze than a univariate summary measure, such as the total fertility rate.

(b) The level of detail in the independent variables listed above as well as other independent variables, including: social norms about the intermediate variables listed above; family size preferences; societal rewards and punishments which support or negate social norms regarding family size and the intermediate variables; mortality levels; net migration levels; and environmental factors. Clearly, this list will be longer than the list of intermediate variables—how much longer will depend upon the analyst and the model. Family planning programme variables other than contraceptive use, e.g., education and propaganda, can be included here.

(c) The model(s) must be specified showing how the independent variables relate to the dependent variables. Models for individual women would be somewhat different than models using areal aggregate measures. A general model can be specified as follows:

$$R = F(A,B,C,D,AB,AC,AD,BC,BD,CD,ABC,ABD,BCD,ABCD)$$

where

- $R$ = a matrix of specific fertility rates in the required detail by age and parity and by calendar year;
- $F$ = an indicator that $R$ is a function of the variables in parentheses which are defined as follows:
- $A$ = a time-series matrix of intermediate variables, excluding use or non-use of contraception, and their interactions with one another;
- $B$ = a time-series matrix of measures of use and non-use of different types of contraception. These types should be further subdivided into contraception provided through the programme and those obtained from other sources if the effect of programme contraception is to be measured;
- $C$ = a time-series matrix of the other dependent variables, excluding any non-contraceptive programme input variables;
- $D$ = a time-series matrix of the non-contraceptive family planning programme input variables, such as levels of staffing, educational efforts to change social norms regarding acceptance and use of contraception. It is necessary to include these variables if the model is to provide a measure of programme effects on non-programme use of contraception;

$AB$ and other two-factor terms refer to first-order interactions between the four sets of variables, $ABC$ and other three-factor terms refer to second-order interactions and $ABCD$ refers to third-order interactions.

Even if one restricts the model to a single response variable, say, the total fertility rate, and the other variables to one per set and a single year, perhaps year $t - 1$, there still are 14 possible effects to be estimated after specifying the analytical model. Ten of these effects are, of course, the interaction effects which in practice are often ignored. Thus, a major problem in any analysis is deciding not only what dependent variables to include but what interaction effects can be included (or eliminated).

Nevertheless, it is useful and necessary to utilize less than a full model in most situations. It should be re-emphasized that careful consideration of all factors in the full model at the design and analysis phases of study is required regardless of whether they are included in the final model(s). One must not overlook the possibility that excluded variables may influence results.

**Replication**

The most acceptable method of determining cause and effect requires randomized assignment of individuals or groups of individuals in controlled experiments. This type of experiment is difficult to do in measuring programme impact on fertility itself. Hence, even if reliable observation, as opposed to experimental, data are available in the detail required, and one agrees on the model(s) relating independent variables to fertility, the conclusions must be limited to the detection of relationships and, hence, to developing additional hypotheses to be tested rather than estimating with specified levels of confidence how much the programme reduced fertility. This limitation applies to any statistical model at this point in time.

In observational studies, replication of different tests of the same hypothesis must be repeated using different models and different sets of data in different circumstances in order to systematically compile evidence for the effects of family planning programmes with different levels of the other covariables. Alternative hypotheses which could explain observed results through excluded variables should be tested to rule out the possibility of hidden effects of interactions with included variables.

**Availability and Quality of Data**

The foregoing discussion was intended to illustrate difficulties faced when reliable data are available. In fact, the data required for a desired model are seldom
all available at a given time and place. Hence, one bravely makes do with what is available. This is acceptable as an interim approach pending more systematic planning and collection of data and analysis in a forward-looking time frame in order to examine the question of causal relationships from as many angles as possible. Even so, the number of dependent variables that can be included in a specific analysis is usually severely limited.

In the "make-do" mode, in the "forward-time frame" mode and even in the experimental-design mode, the magnitude of measurement errors and biases in the data are usually unknown. This writer strongly suspects that the inconsistency of results from different times, places and/or analysts may well be due to different levels of measurement errors and/or biases for the different variables.

It is becoming fashionable to take account of sampling errors, and procedures for doing this are fairly straightforward; indeed, the only justification of tests of significance in observational studies is to determine the probability that observed results could have occurred by chance in an experimental design with the same number of observations. Techniques for detection of and adjustment for measurement errors are not as well developed as for sampling errors.

There is much to be done in determining how measurement errors influence demographic analysis and decisions under different levels of fertility trends and different levels of independent variables. In the meantime, analysts would be well advised to consider whether measurement errors of included variables might also have "explanatory" effects in a particular model.
Cómo conseguir publicaciones de las Naciones Unidas

Las publicaciones de las Naciones Unidas están en venta en librerías y casas distribuidoras en todas partes del mundo. Consulte a su librero o diríjase a: Naciones Unidas, Sección de Ventas, Nueva York o Ginebra.

Comment se procurer les publications des Nations Unies


HOW TO OBTAIN UNITED NATIONS PUBLICATIONS

United Nations publications may be obtained from bookstores and distributors throughout the world. Consult your bookstore or write to: United Nations, Sales Section, New York or Geneva.