

EVALUATION OF FAMILY PLANNING PROGRAM PERFORMANCE: A CRITICAL REVIEW

Jack Reynolds

International Institute for the Study of Human Reproduction, Columbia University
Centro de Estudios Sociales y de Población, Universidad de Costa Rica, San José,
Costa Rica

"There is nothing more frightful than ignorance in action"
Goethe

Abstract—Family Planning Program Evaluation is not a well-developed art, much less an exact science. The evaluation of the performance of such programs has received a good deal of attention but the methodologies are still controversial and the results inconclusive. This is due to a wide variety of constraints, not all of which are technical, and most of which are not unique to family planning. The prospects for improvement of FPPE depend, to a great extent, on overcoming these constraints.

We view Family Planning Program Evaluation (FPPE) rather broadly as a process that includes measurement of goal achievement, feedback of information for adaptive decision-making, and examination of a wide variety of processes to determine how and why a program was or was not successful. Implicit in this viewpoint is an interest in the evaluation of program activities and tasks as well as the total program, and examination of unexpected and undesirable effects as well as expected and desirable ones.

We would settle for a definition recently proposed by Paulson: ". . . a process of examining certain objects and events in the light of specified value standards for the purpose of making adaptive decisions" (1970, p. 1). Few programs approach evaluation in such a comprehensive manner. There are some exceptions, however (see, e.g., Beasley, 1969; Keeny and Cernada, 1970; Han, 1970).

Elsewhere we have presented a typology for classifying evaluative and non-

evaluative studies (Reynolds, 1970b; 1971b), and we have reviewed approaches and issues in administrative forms of evaluation (Reynolds, 1970d; 1971c). In this paper we look at several types of performance evaluation, particularly those related to the measurement of the demographic impact of programs.

EVALUATIVE STUDIES OF PROGRAM PERFORMANCE

Program evaluation is concerned with two broad areas, measures of program performance (effects, effectiveness, and efficiency) and explanatory examinations of program processes to determine how and why a program was or was not successful. There has been so little of the latter that we will not deal with it here.

Evaluation of Program Effects

Performance measures are quite familiar since they relate to the attainment of goals in most cases. Effects measures are the most familiar of all. They are measures of the outcomes or impact of

program effort. Three levels of effects have been described elsewhere in detail (Reynolds, 1970b). Briefly, these are: 1) primary effects (changes in awareness, knowledge, attitudes, motivation); 2) behavioral effects (trial or adoption of a particular form of behavior); and 3) status effects (changes in fertility, health, economic, and social status). A program may have one or more of these effects and they may be positive or negative, planned or unplanned. Most FPPE has stressed the measurement of expected and desirable effects, particularly attitudes toward family planning, use of contraceptives, and births averted.

1. Primary effects have been measured mostly by Knowledge, Attitude and Practice (KAP) sample surveys. These have not often been evaluative, since they have not been used to measure changes due to programs as much as to measure the current status of knowledge, attitudes and contraceptive practice. For example, a finding that 80 percent of a population approve of family planning does not mean that this was due to program effort.

KAP studies may be the single most popular research tool employed in family planning. Well over 130 major KAP studies have been conducted since 1949 (Population Council, 1970a), and hundreds of minor surveys have been undertaken in recent years. Twenty-one KAP surveys have been conducted in Pakistan (National Research Institute, 1969). These do not include other types of family planning surveys, which may number in the thousands. A 1968 newsletter listed 241 reports of such surveys in India alone (Kapil and Saksena, 1968).

The question of the reliability and validity of these surveys has been raised repeatedly (e.g., Mauldin, 1965; Berelson, 1966; Kirk, 1969; Hauser, 1967; White, 1971). A few reports have attempted to check response error and the conclusions have not been methodologically encouraging, although the error was in the di-

rection of underreporting of knowledge and practice (Green, 1969; Stoeckel and Choudhury, 1969). Another study revealed that husbands were more likely to give dependable reports than wives (Poti et al., 1962). Berelson summarized the problems as "how to do technically solid studies as foreign advisors with few qualified nationals, in a foreign language, with largely illiterate people, in an alien culture, on a personally delicate matter that is difficult of measurement, and to do so in a way that will guide policy decisions" (1964, p. 3).

Most of these problems remain, but some progress is being made. Kirk recently concluded: "There is a need for further methodological research in this relatively new area of study. Nevertheless, with all the failings of KAP surveys it is difficult to deny the validity of the very general and consistent conclusions from their findings and their application to family planning programs" (1969, pp. 4, 5). "More striking than their deficiencies has been the extent to which well-designed family planning surveys have proved to be methodologically feasible and scientifically valid in *all* cultures" (p. 12).

More important for this discussion is the contention that "the KAP survey provides a baseline against which later studies can measure the effect of family planning programs" (Population Council, 1970a, p. 1). Aside from the fact that "relatively few studies give comparability by repeated studies of the same population over time" (Kirk, 1969, p. 6), that is, "as evaluation, their use has thus far been infrequent" (Mauldin, 1969, p. 229), I can only repeat my own view that the evaluative power of KAP studies is "particularly questionable. At best KAP surveys show correlations between program effort and KAP changes, they do not demonstrate cause-effect relationships, even though this may, in fact, be the case. If 'proof' of the impact of family planning programs is what we

want, then we need to devote some of our evaluative resources to experiments and/or in-depth investigations of client reactions to family planning programs" (Reynolds, 1970e, p. 48).

2. Evaluations of changes in behavior have been measured largely in terms of contraceptive acceptance and continued use. These evaluations generally suffer from a lack of definition of terms and controlled follow-up (Nortman, 1970, p. 15). The term "acceptor" has at least three different meanings: first acceptance of contraception; first acceptance of a particular contraceptive method; first acceptance of a program. Some programs also count women served in a particular clinic for the first time as new acceptors; others count re-acceptors as new. This results in a large amount of double counting in many programs. It has not been unusual for a Ministry of Health to count as new acceptors all women who were transferred from private programs. It is also common practice to count as new a woman who returns to a program after an absence of several months. In almost all cases it is not known how many acceptors actually began *using* a method. There is no simple indirect way at the moment to determine how many of the women who were prescribed oral contraceptives actually went home and took them. A study in Taiwan reported that "six percent of pill cases never took a first cycle and 24 percent had stopped before starting the second cycle" (Population Council, 1970b, p. 11). All of this means that the most basic data, new acceptors, must be qualified. The need for standard definitions is apparent.

The problems of determining continuity of use are even more difficult because: 1) the term is more poorly defined than "acceptor" (often as women "served" or assumed to be using contraceptives), and 2) few careful follow-up studies have been conducted to determine patterns of contraceptive usage. Again, there are different types of users: people who con-

tinue to use the *program*; people who continue to use *contraception*; and people who continue to use a prescribed contraceptive *method*. We could also add people who continue to use a particular clinic. The distinctions between these are not often made and this results in a good deal of confusion as to the meaning of continued use.

There have been some careful follow-up studies conducted in experimental IUD and oral contraceptive projects. These studies have uncovered high method discontinuation rates (see, e.g., Studies in Family Planning Nos. 18, 24, 54). Such surveys are liable to the same response error problems as KAP surveys, since they rarely include reliability and validity checks. Checks can result in significant adjustments. Rechecks on acknowledged IUD insertions in Korea found an underreporting error of 40 percent (Population Council, 1970b, p. 4).

However, even if it would be possible to determine patterns of use, it would still be necessary to sort out the relative influence of the program from other intervening variables in order to determine how much credit the program can take for continued use of contraceptives. For example, a study in Taiwan found that at least two-thirds of women who became pregnant after dropping out terminated their pregnancies by induced abortion (Population Council, 1970b, p. 10). Another study found that three-fourths of pregnancies among all acceptors were terminated by induced abortion (Chow et al., 1968, p. 231). The implication is that these women would have controlled their fertility anyway without the program. On the other hand, one could hypothesize that the program made them aware of abortion as an alternative (Koya, 1963).

Given the limitations of funds and personnel, most programs estimate continued use on the basis of acceptance data and estimates of discontinuation (Nortman, 1970, p. 70, footnote 1).

These are often based, in turn, on arbitrary definitions of "active status". For example, a woman may be presumed active until three months after missing her last scheduled appointment. Granted that operational definitions are necessary, most of these are not methodologically rigorous. They are rarely derived from empirical data and they do not usually account for intervening variables (Polgar and Kessler, 1966, p. 38). This means that most measures of program effects that are based on acceptance and continuation data must be viewed with some skepticism.

It is helpful to conduct careful follow-up studies and then to construct life tables of continuation (see Potter, 1967; Reynolds and Ramaprasad, 1970, in appendix, pp. 25-29). When clear definitions of acceptance and continuity are combined, significant improvements in effects evaluation can be made. The computer feedback programs of such groups as the Center for Disease Control (CDC) have been used to provide information to clinics on specific women who are late for their appointments (Allen, 1970). These are useful control systems that should give data on patterns of program usage, contraceptive practice, and fertility outcome if supplemented with follow-up studies.

3. The final level of effects, status changes due to the program, has received much attention. Most of this has been directed at the measurement of the effects of family planning programs on birth, growth, and fertility rates, often popularly referred to as "births averted".

It is interesting that most of this effort is directed toward the measurement of *declines*. Since many programs emphasize the voluntary aspect of family planning, and particularly the idea that family planning will allow couples to have as many children as they desire, it is strange that "Little attention has been given to date to the actual or potential impact of family planning services on

increasing natality" (Polgar and Kessler, 1968, p. 38). Polgar and Kessler pointed out that higher population growth "would be a consequence of two different factors: first the provision of treatment for infertility problems and second the improvement of survival rates" (p. 38). To this we might add a third factor, poor family planning service. It is our personal impression that poorly organized and administered family planning services contribute to the *increase* of unwanted fertility. It is an entirely plausible hypothesis that poor education on how to use contraceptives coupled with intermittent service and irregular provision of contraceptive supplies could result in unwanted pregnancies. Such hypotheses are at least worthy of consideration. To the best of our knowledge there is neither evaluation of such possibilities nor adjustment of program effects to account for their impact. Are FPP assumed to be incapable of such negative (or positive?) effects?

There is general recognition, even by family planning proponents, that "it is exceedingly difficult to demonstrate incontrovertibly that a given program has, in fact, been responsible for a given decline in fertility" (Harkavy, n.d., p. 2). Freedman and Takeshita stated that "We will never know conclusively whether the program reduced Taichung's birth rate" (1969, p. 308), and Bogue wrote "there is absolutely no way to answer directly the question, 'What would have happened to the birth rate if there had been no family planning program in Korea?'" (1970, p. 67). Having acknowledged that direct measurement is unlikely, they set out to measure the impact indirectly.

Some of the more popular methods for measuring negative demographic effects of FPP have been described by Bogue (1970) and Mauldin (1968). These approaches are summarized in Figure 1 and discussed below.

Changes in Fertility Status of One

FIGURE 1.—Measures of Demographic Impact of Family Planning Programs

Impact of Family Planning Programs	Used to Measure	
	Fertility changes	Births averted
I. CHANGES IN FERTILITY TRENDS WITHIN ONE GROUP		
A. Trends in Actual Rates	yes	no
1. Actual Births among Total Population		
2. Actual Births among FPP Contraceptors		
3. "Numerator Analysis"		
B. Trends in Actual vs. Expected Rates	no	yes
1. Expected vs. Actual Births: Total Population		
2. Expected vs. Actual Births: FPP Contraceptors		
II. CHANGES IN FERTILITY TRENDS IN EXPERIMENTAL GROUP		
A. Actual Births to Study and Control Groups	yes	yes
B. Actual Births to Study and Matched Groups	yes	yes
C. Actual Births to Study Group and Total Population	yes	yes
III. INDIRECT CALCULATIONS OF EFFECTS ON FERTILITY		
A. Calculated vs. Expected Births among FPP Contraceptors; "Couple Years of Effective Contraception"	yes	yes
B. Other	yes	yes

Group.—One approach measures the trends of actual birth rates of a group over time. Another compares these actual trends with expected trends. We will take these in turn.

1. Trends in Actual Rates.—In the simplest form of analysis the birth rates are plotted on a graph and the point at which the FPP was introduced is noted. An increased rate of decline after the introduction of the program is taken as evidence of program impact. One version measures declines among the total population while another restricts itself to the measurement of declines among FPP contraceptors. The analysis may or may not include allowance for other variables, such as migration, age-

sex redistribution and the impact of non-program fertility regulation measures.

A third version, "numerator analysis", measures the decline in reported births on an age-parity grid (Ravenholt and Fredericksen, 1968). This latter version attempts to avoid the problems associated with poor denominators. However, those countries that have poor denominator data are likely to have poor numerator data also. For example, one of the major problems in computing fertility trends is the lack of reliable data on vital events. A recent United Nations Mission to India (1969) noted that studies in 1945 and 1950 estimated that 40-55 percent of births were not registered. The Mission concluded that "there is no

evidence that vital registration has improved during the intervening period . . ." (Population Council, 1970c, p. 13) and that "the inadequacies of the civil registration system and other sources of fertility data preclude the precise measurement of any limited fertility changes that could be expected at this point in the family planning programme" (p. 18).

Where registration is more complete, changes in fertility trends have been noted as evidence of FPP impact (see, e.g., Freedman, *et al.*, 1969). Berelson expressed cautious optimism in his last status report: "The evidence is not full or finally convincing, but it is beginning to come in and to show a moderate, but still heartening result in the desired direction" (1970, p. 17).

Some reservations to such optimistic conclusions are based on methodological issues, particularly on questions of interpretation of the data. Hauser expressed his reservations several years ago (1967, p. 407), and they are not yet out of date: If one restricts himself to empirical data, then, he must conclude that:

1. There are as yet no satisfactory methods of measuring small changes in fertility and growth rates over short periods of time in the developing areas containing the mass populations of the world.
2. There are as yet no experiments in family planning which have precisely measured the impact of an action program on fertility differentiated from other forces as embodied in secular trend.
3. There are as yet no significant examples of declines in fertility by reason of action programs in areas in which secular decline in fertility has not already occurred.
4. The examples of fertility declines to date in areas in which action programs are underway (Taiwan, Hong Kong, Singapore, South Korea, Ceylon) are not only areas in which declines in birth rates had already occurred but are areas with special characteristics that preclude extension of the results to mass populations in Asia, Latin America, or Africa that are still steeped in illiteracy and in poverty, and that are predominantly still 'traditional societies'.

5. The evidence of differential fertility in the developing regions provides, at best, a basis for wishful thinking about the potential of action programs in such areas. It is not yet known just what the forces are that produced the differential patterns and it is a great leap into the unknown to assume that action programs will therefore increase the differentials or change the behavior of the 'traditionalists'.

Recently Kirk pointed out that minor modifications in the assumptions can make major differences in the comparisons between pre- and post-program fertility rate decline. He shows that "the average annual reduction in the birth rate of Taiwan was 1.1 points before the program (1959-1963) and 1.2 points after the program (1964-1968)" (Kirk, 1969, p. 4). The claimed reduction appeared to be much larger, from 2.3 before the program to 4.2 after. Kirk concludes later "(a) that present methodology, though rapidly gaining in sophistication, is still inadequate to provide a reliable measure of the direct effects of a family planning program and (b) that the demographic impact is probably exaggerated by methods of analysis now widely used" (p. 6).

2. Trends in Actual vs. Expected Rates.—Again the birth rates are plotted and the point of entry of the FPP noted. Then an estimate is made of expected births and this line is placed on the chart. By comparing actual and expected births we have the difference, "births averted". The computations may be made for the entire population (e.g., Wolfers, 1968) or for FPP contraceptors only. Some variations may appear, such as "later live births", but it is the "births averted" idea that has been used most (see, e.g., *Studies in Family Planning* Nos. 45, 54, 57, 2:1, 2:2 appendix B).

The problems involved in this type of analysis include all of those noted above plus a few more. Seltzer (1970, p. 15), among others, has reviewed the procedures. He argues that the estimation

of expected births is particularly difficult, and brings into question attempts to measure births averted: ". . . since they refer to non-events, statistics on births averted are particularly prone to misinterpretation and hence, at times, to misrepresentation. The apparent simplicity of the concept tends to obscure the complexities of estimation, and vagueness about the time reference and the population covered is common." He concludes that "It is preferable to measure program accomplishment in terms of changes in fertility rather than in terms of births averted, because the latter are subject to greater measurement errors and more misuse than are conventional data on fertility" (p. 16).

There is also a discomfoting tendency in many evaluations to neglect to account for the effects of intervening variables other than the FPP, such as migration, age-sex redistribution, foetal mortality, sub-fertility, etc. In fact, many of the eleven intermediate variables described by Davis and Blake (1956) as affecting fertility are never considered in FPPE. A few studies that have taken these factors into consideration have identified variables other than those related to the FPP that have contributed to fertility decline (TinMyaing-Thein, 1969, 1971; Nag, 1970). Kirk noted that "Even where programs have been very successful, their direct effects may have been less than the effects of other influences on the birth rate. In both Taiwan and South Korea, for example, rise in age at marriage accounted for a substantial part of the reduction in birth rates" (1969, p. 8). And on abortion he noted, "It is clear that abortion has been a very important 'silent partner' in the success of programs in Taiwan and South Korea and probably elsewhere" (p. 9).

An interesting approach to isolating FPP impact from other variables has been reported by Hermlin based on data from Taiwan. Using path analysis

(a multivariate technique), he demonstrates a positive relationship between IUD acceptance rates and fertility decline in 282 urban and rural areas. He concludes that "the program effect is a new and added intervention whose influence is largely beyond that of the socio-economic variables" (1968, p. 10).

Changes in Fertility Status of an Experimental Group.—Experimental designs offer the best hope for valid measurement of program effects on fertility. In these approaches two or more groups are compared (experimental vs. control group or matched groups). Sometimes an experimental group is compared to the total population. The evaluation is of the effect of the experimental variable (the FPP) on the actual numbers of births.

One approach is described by Bogue (1970). The *accidental* pregnancy rates of persons in the program is compared with the accidental pregnancy rates of those outside the program. The difference in the rates is attributable to program impact. Although the problem of self-selection is obvious, this could be reduced somewhat by matching, or even further with a control group. How one would identify "accidental" pregnancies, especially in the general population, is not explained. In Taiwan the fertility of the experimental group (FPP contraceptors) was compared with that of the general population (Chow, 1968). The obvious problems again are the incomparability of the two groups and the lack of consideration of intervening variables.

A few matching studies have been reported from Taiwan (Takeshita, et al., 1964; Chow et al., 1969) and from the District of Columbia (Okada, n.d.). These focus on total pregnancy rates and are among the best available to date, although they are subject to some of the same criticisms that have been noted above.

Controlled experiments are rare in FPPE because they are not easy to

mount. It is difficult to find suitable control areas and to keep them uncontaminated by the study variable. It is also difficult to withhold services from control groups (it is "unethical"), especially for long periods of time. An experiment is being attempted in Guatemala among four villages (INCAP, Growth and Development Project). Other possibilities are "natural" experiments and "small" experiments as suggested by Chandrasekaran and Freymann (1965). Campbell and Stanley (1963) have suggested a variety of "quasi-experiments" that could be adapted to FPPE.

Indirect Calculations of Program Effects on Fertility.—Some of the more complex attempts to measure demographic effects of FPP are based on estimates of "couple years of effective contraception" (Mauldin, 1968) that can be attributed to contraceptives distributed by the FPP. Formulas attempt to transform service statistic data (IUDs inserted, pills distributed) into estimates of gross births averted. The net number of averted births (or the additional births averted by the FPP) are estimated by taking the difference between expected births without the program and gross births averted with the program. This approach differs markedly from the former two in that no empirical fertility data are used in the final calculations and the formulas are largely educated guesses of how long different classes of contraceptives will protect an individual woman against pregnancy.

The method proposed by Lee and Ibister for Korea (1965, p. 738) consisted of formulas for calculating "(1) the effect of a given birth control program on the fertility of a future year, (2) the total effect that IUD's will have during all the years they remain in use, and (3) the scale of an IUD program required to achieve specified objectives".

A method developed by Potter (1969) for Taiwan, and a similar method developed by Wolfers (1969) for Singapore

present formulas for calculating births averted from IUD programs. A method developed by Wishik (1968) for Pakistan calculates contraceptive "prevalence" and "achievement" as well as a "Protection Prevalence Index" that relates program output to program goals. Wishik's approach tries to convert each quantity of contraceptives distributed into a common denominator of program output called "couple-years of protection" (CYP). Although Wishik did not intend that the CYP would be used to calculate births averted, it was probably inevitable that it would be used for that purpose (see Berelson, 1970, p. 5).

One of the problems with these indirect-deductive approaches is that of deciding what types of adjustments to make. Another is trying to assign realistic weights (periods of "protection") to each contraceptive. Some formulas adjust for mortality or birth intervals, others for sub-fecundity, post-partum sterility, etc. Some of the weights are based on empirical data, particularly IUD retention rates, but others are estimates, for example, condoms and sterilizations.

The complexities of the calculations and the many assumptions involved have led to repeated criticisms (see, e.g., Mauldin, 1967; Bean and Seltzer, 1968; Seltzer, 1970; Wolfers, 1969; Kirk, 1969). Harkavy reported that "A recent publication of the Demographic Training and Research Centre of Chembur in India illustrates the hazards involved in explaining a low birth rate by the prevalence of contraceptive practice" (1967, p. 2). He concludes that the "prevalence of contraceptive practice cannot be claimed as the cause of Bombay's low birth rate". Rather, it is due to age-sex redistribution. This again points out the problems associated with neglect of other intervening variables.

Some concern has been raised that if optimistic estimates of "achievement"

are proven false by actual data on birth rates, this may lead to a sense of futility among FPP workers (Bean and Seltzer, 1968, p. 959). This is not necessarily so. Most of these estimates do not account for a variety of secular trends and intervening variables that could result in a true decline in fertility, for which the FPP would take the credit. Also, the assumption is that FPP workers are highly motivated to bring down the birth rate, whereas it is more likely that they are highly motivated to retain their jobs, regardless of what happens to the birth rate.

Nevertheless, these methods have a great deal of popular appeal because of their apparent simplicity (once the formulas and weights have been worked out), because of their claim to convert service statistics into estimates of program impact, and, we might suggest, because they can be used to set program output goals that are much easier to measure than program effects.

There is a real need for more work in this field, since present FPPE methods are subjected to so many data constraints that accurate measurement of program effects on fertility is a long way off. The CYP concept has found increasing acceptance (e.g., Robinson, 1969) because it deals with all contraceptive methods. It may yet prove to be a useful evaluative tool, especially if field tests can be mounted to refine the formulas and weights. A revised set of calculations and a manual of instruction is being prepared by Wishik and Chen (1971) for this purpose.

Another indirect approach, also devised by Wishik (1968) for Pakistan has been reported and is being revised for field testing. This involves surveys of women in a FPP target area to determine changes in "Age/Parenthood Status" (age and number of living children) over time. The concept is that changes within high fertility groups will reflect program impact.

Some researchers have tried to show that changes in KAP can be used to demonstrate the impact of the program on fertility. For example, Bogue (1970, p. 79) argues that modernization is less likely to affect the lower socioeconomic classes than the higher classes. Thus, "If the family planning clients are disproportionately drawn from the lower socioeconomic strata of society . . .", then one can argue that fertility decline is not due to modernization. Freedman and Takeshita have made the same argument (1969, p. 309). All we can say to that line of reasoning is that these are measures of program acceptance, not fertility.

As a final note we should make some reference to the evaluation of undesirable and unexpected effects. The major focus here has not been demographic or sociological, but medical. Most programs take pains to measure contraceptive side effects. Little attention has been paid to the negative effects of poorly run programs. Such things as excessive waiting time in clinics, poor staff attitudes, clinic inaccessibility and a multitude of other obvious factors may have an important bearing on client recruitment, retention and, eventually, fertility (Reynolds, 1970a).

Undesirable effects are less likely to be uncovered by hypothesis testing than by broader process analysis. This should stand as a recommendation for more systematic studies to determine how people react to family planning programs. Such studies may uncover desirable but unexpected effects, such as increased confidence in the government, increased use of medical services, the development of planning concepts, or even improved sexual relations (Westoff and Ryder, 1968).

Evaluation of Program Effectiveness

There has been much less evaluation of program effectiveness and efficiency. Effectiveness measures compare achievement to some goal, to a standard, to a recognized need, or to some competitive

program. For example, a program goal may be to open 100 clinics, to enroll 100,000 women, or to reduce the birth rate by 50 percent.

The most common effectiveness measures are related to goals, many of which have been set arbitrarily on the basis of educated guesses of what might be politically acceptable or practically attainable rather than what is needed. Few goals have been set for primary effects (changes in knowledge, attitudes, etc.), but a good deal of effort has gone into setting goals for behavioral and status changes. Many of these are based on computed estimates of "need".

One of the more serious problems here is the jump that is made from needed fertility changes to the numbers of acceptors or users required to attain those changes. Assuming that one can determine at what point a population should be stabilized (which is another controversial point), the next step is to define a target population that is in need of service.

The most sophisticated estimate to date is based on the Dryfoos-Polgar-Varky formula (summarized in Office of Economic Opportunity, 1968, p. 238). This formula attempts to define a target group for a segment of the United States population (low-income women who are fecund, exposed, and not pregnant or seeking a pregnancy and who are not currently provided with contraceptive care). Although this estimate has not been validated, it has been used for program planning purposes (Jaffe, et al., 1969; Dryfoos, 1971) and conceivably could be used for evaluation of effectiveness.

Another technique, calculation of "percent protected", has been described in the Population Council's *Handbook for Service Statistics* (Ross, et al., 1968, p. 59). A variety of other techniques is available, but their derivation is difficult to determine (see, e.g., Mauldin, 1967, pp. 73-75). All of these methods are based

on assumptions of fertility and other key factors that have been discussed above and are subject to the same methodological shortcomings. Many are based on assumptions of constant fertility rates and do not account for changes in fertility-related variables other than contraception.

There have not been enough studies of patterns of contraceptive use to permit the estimation of periods of service required to achieve desired fertility levels. The governing proposition is that reduction of unwanted fertility would contribute significantly to the stabilization of fertility rates (Bumpass and Westoff, 1970a; 1970b). Although this sounds reasonable, there is a serious question about the relationship between expressed ideal family size (unwanted fertility) and the market for family planning (Hauser, 1967; Kirk, 1969).

Many programs measure effectiveness by comparing numbers of acceptors enrolled to some target figure. Quite often the acceptor figures are simply accumulated and assumed to represent active users. This ignores program and contraceptive dropouts and results, therefore, in gross overestimates of numbers of women served (Reynolds, 1971a). With dropout rates ranging between 20-80 percent annually, this is a serious omission indeed.

Births averted also has been seen as a program goal. Often the target is based on dubious calculations, such as the assumption that "one birth would be prevented for each group of four users of effective methods during the year" (Gasie, et al., 1970, p. 5). The constants vary from place to place. The Taiwan figure has been four to five IUD insertions for one prevented birth. The figure in Korea appears to be five to seven IUDs or 400 condoms (Mauldin, 1967, p. 73).

Evaluation of Program Efficiency

These are measures of the relative

costs of providing services or of achieving effects. Often they are expressed in ratios, such as births averted per dollar, patients seen per physician-hour, etc. Like effectiveness measures they are usually compared to some goal, standard, competing program, etc.; for example, program A can serve a woman for one year for \$6.00 compared to \$9.50 for program B.

Efficiency measures have been most useful at the lower levels of evaluation (output/input calculations), such as cost per acceptor (\$5.25 in Korea, \$4.00 in Taiwan) (Population Council, 1970a, p. 15).

The measurement of effects compared to inputs is more complicated, mostly because of the difficulties outlined above in defining and measuring behavioral and status changes (Nortman, 1970, p. 15). Those who have taken the greatest interest in this field are economists, who have been particularly intrigued (and divided?) by the question of costs and benefits of population limitation, including the costs of preventing births (for a recent review see Robinson and Horlacher, 1971). Most of these studies have been theoretical or speculative and very few have focused on the costs and benefits of FPP. One exception is a "Cost-Benefit Analysis of Family Planning Programme" by Basu (1968) that is based on questionable assumptions and measures of both costs and benefits. Another approach was taken by Robinson (1969) in "A Cost-Effectiveness Analysis of Selected National Family Planning Programs". Robinson uses "assumed benefits" arising out of program "outputs" (here measured in CYPs). He then compares the outputs to program "inputs", which are defined as annual total expenditures. This is an output/input study rather than an effects/input study. Robinson explains why he did not select births averted as an effects measure and he describes many of the substantial limitations inherent in the meas-

ures he did select (Robinson, 1969, pp. 5-23). However, Berelson, in reference to the data, makes the jump: ". . . as a rough order of magnitude, assuming three to four CYPs for one averted birth—in which case the cost-benefit ratio for family planning ranges from 1:10 to over 1:30" (1970, p. 5). Evidently using the same report, USAID also makes the jump, listing figures for "births averted as a percentage of births expected in absence of the program" for seven countries (Office of Population, 1970, p. 29). How these conversions were made is not explained.

COMMON PROBLEMS IN PROGRAM EVALUATION

A large number of problems have been noted or implied above. In this section we will merely list some general problems that act as serious constraints on FPPE.

Conceptual Problems

There is a good deal of honest confusion about the meaning of the term "evaluation", especially as it applies to social action programs. Part of this confusion is due to a lack of agreement as to the purpose of program evaluation. Three schools of thought have been identified: judgmental, adaptive, and control. The first emphasizes the determination of the worth or value of programs (Scriven, 1967). Programs are described and *judged* on the basis of some criteria. The selection of the criteria is critical. In some cases the administrator selects the criteria that will make his program appear successful; in others the evaluator makes the choice. The second school emphasizes the provision of information for adaptive decision-making (Nelson, 1970, pp. 5-29). Programs are described in terms of performance and the administrator is presented with a choice of several courses of action. The third school emphasizes the rapid feedback of management control information

so that alterations from a planned course of action can be corrected. This type of evaluation is largely monitoring of activities coupled with corrective action when something goes wrong (see, e.g., Allen, 1970).

There is a good deal of overlap, since all of the above approaches can be called "decision-making" models. However, the subtleties of the differences are what seem to lead to the confusion between evaluators and administrators. The latter often see evaluation as fault-finding or muckraking. They complain about never seeing useful results or of not being told how they can apply the results they receive. In cases like these the evaluator probably belongs to the first school and is more interested in judging the program than in feeding programming alternatives to the administrator. Those who favor "outside evaluation" over evaluation conducted by someone within the organization are also more concerned with judgmental evaluation than with adaptive evaluation.

This might explain why it is that planning and evaluation are more closely linked in theory than in practice. Judgmental forms of evaluation are often threatening to administrators. Their resistance may mean that the program operates without evaluation. Decision-making forms of evaluation, such as those offered by the second and third schools, are tied to planning, since an inherent part of the evaluation is the development of alternative courses of action. Yet this type of evaluation is rarely found in FPP.

A good deal of the misunderstanding about the purpose of evaluation could be eliminated if the administrators and evaluators more clearly understood the distinctions outlined above (see Saslow, 1970, pp. 1-55).

Another conceptual problem is that most evaluators and administrators think only of the "goal-attainment" model of evaluation, although there are

alternatives, for example, the "systems model" (Etzioni, 1960). One of the major problems with the former is that it requires the stating of objectives, which in many programs is extremely difficult. Furthermore, program goals tend to change over time so that consistent evaluation of performance is not always possible. Evaluators and administrators need to be aware of the different models available so that they can select the type of evaluative approach most suited to their needs. No single form of evaluation is best for all family planning programs (see Seltzer, 1970; Reynolds, 1970b).

A third concept that has caused difficulty is "family planning". It is not at all clear what the term means. The controversy over whether FPP will succeed (Davis, 1967; Blake, 1969; Harkavy et al., 1969) depends, to a great extent, on what they are supposed to succeed in doing. Stated objectives vary widely (see, e.g., Hauser, 1967; White, 1971; Polgar and Kessler, 1966; Berelson, 1970; Davis, 1967; Blake, 1969; Harkavy et al., 1969). We would submit that there are at least three types of FPP: 1) those that admit they are in business to bring down birth rates; 2) those that deny it; and 3) those that don't want to say. It is obvious that the first type can, and should, be evaluated in terms of their success in reducing fertility. The second type should not be since they are not necessarily population control (or fertility reduction) programs. Programs that emphasize voluntarism, birth spacing, or having the number of children desired should not be measured only on fertility reduction criteria (see Polgar and Kessler, 1966). Indeed, the concept of birth spacing may preclude such simplistic measurement, since a success (for example, in reducing *subfertility*, or in successfully *planning* a pregnancy) can only be counted as a success, not as a program, method, or client failure. The term "family planning" has served as a euphemism for the third type in two

senses: for those who see family planning as a way to sneak in population control programs; and for those who see family planning as a way of getting funds and placating population control proponents without having to mount a program with clear-cut fertility objectives. These latter programs are the ones that will remain at the center of the controversy and probably should be reviewed and reclassified.

Several other confused terms have been mentioned above, e.g., acceptors, users, continuers. All such key terms need to be more precisely defined if FPPE is to become more precise.

Socio-Political Problems

We mentioned above that the selection of evaluation criteria and topics is often a problem. Weckwerth has stated that "evaluation will boil down to who has the right to decide and who renders judgement" (1969, p. 4). Who decides is often a political question. Evaluators have complained that they are often asked to *prove* that a program is successful after it has started to run into some trouble (see, e.g., Williams and Evans, 1969). Sometimes a particular evaluative approach is in vogue and is adopted because it is well known, popular, visible, or simple; sometimes a high-level official strongly suggests that it be used (witness the life-cycle of PPBS); sometimes a funding agency offers financial incentives for the use of a particular method. Selection of topics for evaluation is often done by the same procedure.

A related problem that makes experimentation difficult is that of ethics. Social action programs are usually subjected to severe criticism if they try to set up a controlled experiment. This is seen as unnecessarily withholding a service from a group in need. One of the fallacies of this attitude is that it assumes that the program is effective and that people are better off with than without the service. The experiences from the

evaluation of OEO programs are instructive (Weiss and Rein, 1969; Ferman, 1969; Williams and Evans, 1969).

These problems can be solved only if evaluation is considered well in advance so that evaluation approaches and topics can be carefully selected and systems and studies carefully designed. The experimentation problem can be alleviated partly by using quasi-experiments and by taking advantage of natural experiments.

Technical Problems

The inadequacy of broad theories relating social action programs to society is particularly troublesome for FPPE. Development theory is still a new area and the prospects for major breakthroughs in the near future are not encouraging given the long history of disagreements about economic theories of growth. There is also a good deal of resistance to many of these theories in the developing areas (see *Concerned Demography*, 1971). This makes it difficult to fit FPP into an overall framework of economic, social, and planned development. This in turn makes evaluation of success difficult.

Organization theory is also relatively new and most of the work that has been done in this area has focused on commercial enterprises and psychology. Very little has been done on public organizations outside of studies of bureaucracies. One of the more difficult problems here is that of relating organizations (and their products) to their environments. The difficulty is not restricted to the measurement of program impact but also includes accounting for political and sociological interaction. Are FPP causal variables, for example? This is an important conceptual question that has not been answered adequately. Most evaluations take the FPP as the independent causal variable of fertility decline. This seems to be an unwarranted and naive perspective. Many of the above findings

indicate that FPP are not always necessary, nor particularly sufficient. It is more likely that FPP are one of many intervening variables that contribute to fertility regulation *under certain conditions*. Instead of phrasing our evaluation in terms of direct causes, we might do better to study the conditions under which FPP contribute to fertility regulation. (Think of evaluating the impact of a supermarket on nutritional levels.)

A good many technical problems have been mentioned above and in the literature on FPPE. Some of the more obvious are: unknowledgeable and untrained field staff; use of indices and survey techniques of unknown reliability and validity; inadequate census, vital registration, and survey data; poor sampling frames and techniques; falsification of program activity reports; incomplete or inaccurate case cards; and on and on. Most of these shortcomings are solvable eventually, given careful training and supervision of personnel.

Several other problems are not so easily solved. For example, there is a great lack of indices and scales for the measurement of health, motivation, contraceptive behavior, and other key variables. There are not enough analytical techniques for measuring small changes over short periods of time, nor is enough attention being paid to non-program, intercultural, and intracultural variations. Developments in these areas are likely to come from outside the areas of FPPE since they are concerns of all social scientists.

Finally, I wish to stress two major shortcomings in FPPE that have been mentioned in various parts of this review. The first is the general neglect to account for intervening variables other than those offered by the program. The other is related—the neglect to account for the efforts of the private sector. An ongoing study of the remarkable decline in fertility in Costa Rica has taken both of these factors into account (Tin-

MyaingThein, 1971). The conclusions are instructive. Between 1959 and 1965 the decline was due largely to a declining proportion of women entering into sexual unions. This had nothing to do with the program. From 1965 to 1969 this factor, coupled with increased use of contraceptives and sterilization, accounted for most of the fertility decline. The private sector contributed about 60 percent of the contraceptives, and sterilizations were not a part of the official program.

PROSPECTS FOR FAMILY PLANNING PROGRAM EVALUATION

Despite all of the problems and limitations outlined in the first two sections, we can end on an optimistic note. In the first place, there is a good deal of interest in FPPE, and an increasing amount of money (which is why there is a good deal of interest). Thus, we can expect not only an increase in the number of such programs, but also a broader interdisciplinary involvement in the field in the near future. This should be of benefit to FPPE in at least two ways: first, greater attention to the evaluation of programs; and second, an introduction of a variety of social science and research concepts that may prove quite valuable to FPPE. Already we see anthropologists, sociologists, psychologists, systems analysts, epidemiologists, economists, media experts, administrative scientists, ecologists, educators, nutritionists, marketing experts, political scientists, and even ethicists and a host of other specialists taking an interest in family planning research, at least a part of whose efforts should be of some evaluative use.

In the second place, there is an increased interest in formal program evaluation, both in the broad area of social action programs and specifically in FPPE. A number of universities are now engaged in operational program evaluation through consulting and advisory contracts. Our own Division is helping to establish Family Planning Program

Evaluation Units in El Salvador, Ecuador, and Costa Rica. One should not ignore the large amount of valuable work being conducted in such places as Korea, Taiwan, and Louisiana. Seminars and training courses in FPPE are being held with increasing frequency both in the United States and abroad. And several organizations are trying to produce manuals for program evaluation.

All of these developments are positive and augur well for the future of FPPE. Still we should not lose sight of the difficult problems noted in this review. There is a pressing need to develop clear concepts and to define key terms, to improve existing data collection and analysis procedures, to refine survey instruments, to refine and validate measurements, and especially to design, test, and install evaluative systems and to conduct evaluative studies that will provide information that can be used to improve family planning programs. Good intentions and dedication are not enough. As Hauser warned several years ago, ". . . enthusiasm without effective direction can produce great waste, and optimism which distorts vision can make much mischief" (Hauser, 1967, p. 136).

REFERENCES

- Allen, David T. 1970. Systems Specifications: Tennessee Family Planning Record and Evaluation System. Atlanta, Georgia: Family Planning Evaluation Activity, Center for Disease Control.
- Basu, R. N. 1968. Cost-benefit analysis of family planning programme. *Family Planning News* (December), pp. 2-6.
- Bean, Lee L., and William Seltzer. 1968. Couple years of protection and births prevented: a methodological examination. *Demography* 5:947-959.
- Beasley, Joseph D. 1969. View from Louisiana. *Family Planning Perspectives* 1:2-27 (Spring).
- Berelson, Bernard. 1964. Sample surveys and population control: introduction. *Public Opinion Quarterly* 28(3):361-366.
- . 1966. KAP studies on fertility. In Bernard Berelson *et al.* (eds.), *Family Planning and Population Programs*. Chicago: University of Chicago Press.
- . 1970. The present state of family planning programs. *Studies in Family Planning* No. 57:1-11.
- Bogue, Donald J. (ed.). 1970. *Family Planning Improvement through Evaluation*. Chicago: Community and Family Study Center, University of Chicago.
- Bumpass, Larry, and Charles F. Westoff. 1970a. The 'perfect contraceptive' population. *Science* 169:1177-1182.
- . 1970b. Unwanted births and U.S. population growth. *Family Planning Perspectives* 2:9-11 (October).
- Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Chandrasekaran, C., and Moye W. Freymann. 1965. Evaluating community family planning programs. In Mindel C. Sheps and Jeanne C. Ridley (eds.), *Public Health and Population Change*. Pittsburgh: University of Pittsburgh Press.
- Chow, L. P. 1968. A study of the demographic impact of an IUD program. *Population Studies* 22:347-359 (November).
- , Ronald Freedman, Robert G. Potter, and A. K. Jain. 1968. Correlates of IUD termination in a mass family planning program. *Milbank Memorial Fund Quarterly* 46(2) Part 1:215-227.
- , M. C. Chang, and T. H. Liu. 1969. Taiwan: demographic impact of an IUD program. *Studies in Family Planning* No. 45:1-6.
- Concerned Demography. 1971. Vol. 2 No. 3.
- Davis, Kingsley, and Judith Blake. 1956. Social structure and fertility: an analytical framework. *Economic Development and Cultural Change* 4:211-235.
- Dryfoos, Joy G. 1971. Planning family planning in eight metropolitan areas. *Family Planning Perspectives* 3:11-15 (January).
- Etzioni, Amitai. 1960. Two approaches to organizational analysis: a critique and a suggestion. *Administrative Science Quarterly* 5:257-278.
- Ferman, Louis A. 1969. Some perspectives on evaluating social welfare programs. *Annals* 385:143-156.
- Freedman, Ronald, *et al.* 1969. Hong Kong: the continuing fertility decline, 1967. *Studies in Family Planning* No. 44:8-15.
- , and John Y. Takeshita. 1969. *Family Planning in Taiwan*. Princeton: Princeton University Press.
- Gaisie, S. K. *et al.* 1970. Ghana. County Profiles (October).
- Green, Lawrence W. 1969. Measuring the extent of contraceptive knowledge and use

- not reported in interviews. *Pakistan Journal of Family Planning* 3(2):7-16.
- Han, Dae Woo. 1970. The Republic of Korea. *Country Profiles* (April).
- Harkavy, Oskar. n.d. Impact of family planning programs on the birth rate. Ford Foundation reprint.
- Hauser, Philip M. 1967. Family planning and population programs—a book review article. *Demography* 4:397-414.
- Hermalin, Albert I. Taiwan: an area analysis of the effect of acceptances on fertility. *Studies in Family Planning* No. 33:7-11.
- Hyman, Herbert, *et al.* 1962. *Applications of Methods of Evaluation*. Berkeley and Los Angeles: University of California Press.
- Jaffe, Frederick S., Joy G. Dryfoos, and Raymond C. Lerner. 1969. Planning for community-wide family planning services. *American Journal of Public Health* 59:1339-1354.
- Kapil, Krishan K., and Devendra N. Saksena. 1968. A biography of family planning knowledge, attitude and practice studies in India, 1951-1958. *Newsletter* No. 26 (October). Demographic Training and Research Centre, Chembur, Bombay, India.
- Keeny, S. M., and George Cernada. 1970. Taiwan. *Country Profiles* (February).
- Kirk, Dudley. 1969. Methods and results of surveys for evaluation of family planning. Unpublished paper prepared for the meeting of the International Union for the Scientific Study of Population, London, September 1969.
- Koya, Yoshio. 1963. *Pioneering in Family Planning*. Tokyo: Japan Medical Publishers.
- Lee, B. M., and J. Isbister. 1966. The impact of birth control programs on fertility. In Bernard Berelson *et al.* (eds.), *Family Planning and Population Programs*. Chicago: University of Chicago Press.
- Mauldin, W. Parker. 1965. Application of survey techniques to fertility studies. In Mindel Sheps and Jeanne C. Ridley (eds.), *Public Health and Population Change*. Pittsburgh: University of Pittsburgh Press.
- . 1967. Measurement and evaluation of national family planning programs. *Demography* 4:71-80.
- . 1968. Births averted by family planning programs. *Studies in Family Planning* No. 33:1-7.
- . 1969. Population surveys: an essential tool. In Bernard Berelson (ed.), *Family Planning Programs: An International Study*. New York: Basic Books.
- Nag, Moni. 1970. Patterns of mating behavior, emigration and contraceptives as factors affecting human fertility in Barbados. Unpublished paper.
- National Research Institute of Family Planning. 1969. *Inventory of Family Planning Research in Pakistan*. Karachi.
- Nelson, Frank G. 1970. *Models for Evaluation: An Introduction*. Monmouth, Oregon: Teaching Research, Oregon State System of Higher Education.
- Nortman, Dorothy. 1970. Population and family planning programs: a factbook. *Reports on Population/Family Planning* No. 2 (July).
- Office of Economic Opportunity. 1968. *Need for Subsidized Family Planning Services, United States, Each State and County 1968*. OEO Pamphlet 6130-6.
- Office of Population. 1970. *Population Program Assistance*. Washington: Agency for International Development.
- Okada, Louise M. n.d. (196?). The use of matched pairs in the evaluation of the District of Columbia Department of Health Birth Control Program. Mimeographed.
- Paulson, Casper F. 1970. *Introduction. A Strategy for Evaluation Design*. Monmouth, Oregon: Teaching Research, Oregon State System of Higher Education.
- Polgar, Steven. 1967. *Educational interviews*. Mimeograph No. 74/267. New York: Planned Parenthood-World Population.
- , and Frederick S. Jaffe. 1968. Evaluation and recordkeeping for U. S. family planning services. *Public Health Reports* 83:639-651.
- , and Alexander Kessler. 1966. *An introduction to family planning in the context of health services*. Geneva: World Health Organization. Mimeographed.
- Population Council. 1970a. *A Manual for Surveys of Fertility and Family Planning Knowledge, Attitudes and Practice*. New York: Demographic Division, The Population Council.
- . 1970b. Korea/Taiwan: report on the national family planning programs. *Studies in Family Planning* No. 54: 1-16.
- . 1970c. India: UN mission evaluation of the family planning program. *Studies in Family Planning* No. 56:4-18.
- Poti, S. J., B. Chakraborti, and C. R. Malaker. 1962. Reliability of data relating to contraceptive practice. In Clyde V. Kiser (ed.), *Research in Family Planning*. Princeton: Princeton University Press.
- Potter, Robert G. 1967. The multiple decrement life table as an approach to the measurement of use effectiveness and demographic effectiveness of contraception. *Contributed Papers, Sydney Conference, International Union for the Scientific Study of Population*.
- . 1969. Estimating births averted in a family planning program. In S. J. Behrman,

- Leslie Corsa, and Ronald Freedman (eds.), *Fertility and Family Planning: A World View*. Ann Arbor: University of Michigan Press.
- Ravenholt, R., and Fredericksen, H. 1968. Numerator analysis of fertility patterns. *Public Health Reports* 83:449-457.
- Reynolds, Jack. 1970a. Delivering family planning services: autonomous vs. integrated clinics. *Family Planning Perspectives* 2:15-22.
- . 1970b. A framework for the selection of family planning program evaluation topics. *Manuals for Evaluation of Family Planning and Population Programs*, No. 1. New York: Division for Program Development and Evaluation, International Institute for the Study of Human Reproduction, Columbia University.
- . 1970c. A framework for the design of family planning program evaluation systems. *Manuals for Evaluation of Family Planning and Population Programs*, No. 2. New York: Division for Program Development and Evaluation, International Institute for the Study of Human Reproduction, Columbia University.
- . 1970d. Operational evaluation of family planning programs through process analysis. *Manuals for Evaluation of Family Planning and Population Programs*, No. 4. New York: Division for Program Development and Evaluation, International Institute for the Study of Human Reproduction, Columbia University.
- . 1970e. KAP revisited. *Family Planning Perspectives* 2:48-49 (October).
- . 1971a. Methods for estimating future caseloads of family planning programs. *Family Planning Perspectives* 3:56-61 (January).
- . 1971b. Family planning program evaluation: status, problems, prospects. Unpublished paper presented at the annual meeting of the Population Association of America, Washington, D. C.
- . 1971c. Utilización rendimiento de las clínicas: accesibilidad, disponibilidad, adecuación y aceptabilidad de servicios. Paper presented at A Laboratory for Executives and Administrators, Bogotá, Colombia (in press).
- , and Rukmani Ramaprasad. 1970. A method for estimating future caseloads of family planning programs. *Manuals for Evaluation of Family Planning and Population Programs*, No. 3. New York: Division for Program Development and Evaluation, International Institute for the Study of Human Reproduction, Columbia University.
- Robinson, Warren. 1969. A cost-effectiveness analysis of selected national family planning programs. Unpublished report for U.S.A.I.D. (December).
- , and David Horlacher. 1971. Population growth and economic welfare. Reports on Population/Family Planning No. 6 (February).
- Ross, John, Frederick F. Stephan, and Walter B. Watson. 1968 (revised 1969). *A Handbook for Service Statistics in Family Planning Programs*. New York: The Population Council.
- Saslow, Michael G. 1970. Establishing the purpose for evaluation. In Casper F. Paulson (ed.), *Strategy for Evaluation Design*. Monmouth, Oregon: Teaching Research, Oregon State System of Higher Education.
- Scriven, Michael. 1967. The methodology of evaluation. In Ralph Tyler, Robert Gagne, and Michael Scriven (eds.), *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally.
- Seltzer, William. 1970. Measurement of accomplishments: the evaluation of family planning efforts. *Studies in Family Planning* No. 53:9-16.
- Stoeckel, John, and Mogbul A. Choudhury. 1969. Pakistan: response validity in a KAP survey. *Studies in Family Planning* No. 47:5-9.
- Takeshita, John Y., J. Y. Peng, and Paul K. C. Liu. 1964. A study of the effectiveness of the pregnancy health program in Taiwan. *Eugenics Quarterly* 11(4):222-233.
- TinMyaingThein. 1969 (revised 1970). Some variables of fertility decline in Trinidad and Tobago. Unpublished paper presented at the annual meeting of the American Anthropological Association.
- . 1971. Contraception in Costa Rica: the role of the private and public sectors in fertility decline, a preliminary report. Unpublished.
- Weckwerth, Vernon E. 1969. On evaluation: a tool or a tyranny—II. Comment Series, No. 0-11 (22). System Development Project, University of Minnesota.
- Weiss, Robert S., and Martin Rein. 1969. The evaluation of broad-aim programs: a cautionary case and a moral. *Annals* 335:133-142.
- Westoff, C. F., and N. B. Ryder. 1968. Duration of use of oral contraception in the United States of America. *Public Health Reports* 83:277-287.
- White, Paul E. 1971. Ritualism in family planning programs. Unpublished paper presented at the annual meeting of the Society for Applied Anthropology.
- Williams, Walter, and John W. Evans. 1969. The politics of evaluation: the case of head start. *Annals* 335:118-132.

- Wishik, Samuel M. 1968a. Indexes for measurement of amount of contraceptive practice. Unpublished paper presented at meeting of Expert Group on Assessment of Acceptance and Use-Effectiveness of Family Planning Methods, ECAFE, Bangkok.
- . 1968b. Evaluation of national family planning programs through group-specific fertility patterns. Proceedings of the American Statistical Association meeting.
- , and Kwan Hwa Chen. In preparation. Manual for estimating couple-years of protection. Manuals for Evaluation of Population and Family Planning Programs. New York: Division for Program Development and Evaluation, International Institute for the Study of Human Reproduction, Columbia University.
- Wolfers, David. 1968. An evaluation criterion for a national family planning program. *American Journal of Public Health* 58:1447-1451.
- . 1969. The demographic effects of a contraceptive programme. *Population Studies* (March):111-140.
- World Health Organization. 1970. Health Aspects of Family Planning. WHO Technical Reports Series No. 442. Geneva.