

## Family planning impact evaluation: the evolution of techniques

*Albert I. Hermalin<sup>1</sup>*

### **Acknowledgment:**

Support for the seminar paper on which this report is based was provided in part through The EVALUATION PROJECT, USAID sponsored contract at the Carolina Population Center.

### **BACKGROUND AND ABSTRACT**

This paper is a slightly revised version of a paper prepared for the seminar on methods for impact evaluation of family planning programs held in Jaco, Costa Rica, May 14-16, 1997. The seminar was sponsored by the International Union for the Scientific Study of Population (IUSSP), the United States Agency for International Development (USAID), the Carolina Population Center of the University of North Carolina, and the Central American Population Program of the University of Costa Rica.

The goal of the seminar was to look at current methodological problems facing careful evaluation of the impact of programs, to examine some of the new methods that have been developed to address persistent issues, and to assess the methodological challenges posed by the expanded goals of many programs following the 1994 Cairo International Conference on Population and Development.

This paper was designed to serve as the background to discussions of current methodologies and issues by tracing the development and nature of methods for assessing impact that started soon after the first programs were initiated in the 1950s. The techniques discussed include standardization and trend analysis, the analyses of acceptor data, experimental designs, multivariate areal analysis, population-based surveys, and multilevel strategies.

The intent of the program sponsors and coordinators was to publish the collected papers but various contingencies intervened to make this infeasible. A description of the seminar and many of the papers are maintained on the University of Costa Rica website: <http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>. As a background chapter, the original version contained references to many of the other chapters planned for the volume. As many of these papers appear on the website, relevant references are given to the authors and this website throughout the paper.

---

<sup>1</sup> Research Scientist, Population Studies Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, [alberth@isr.umich.edu](mailto:alberth@isr.umich.edu)

## THE GROWTH OF FAMILY PLANNING PROGRAMS

In 2002, the Indian family planning program marked its 50<sup>th</sup> anniversary. Its initiation in 1952 is generally regarded as the date of the first family planning program in the sense of a government sponsored set of social and health interventions designed to provide contraceptive services and supplies in order to promote lower fertility, assist couples to achieve their desired family size, and/or improve the health of childbearing age women and their children. The historical noteworthiness of this action lies in the intersection of government policy along with government action toward the implementation of these goals. The individual components have long existed: governmental concerns about population size or other dimensions can be traced far back (United Nations, 1973); contraception in some form was practiced in ancient times (Himes, 1963); and even organized efforts to affect family size were under way in late 19th century Europe and gained momentum in the early decades of the 20th century through voluntary associations in many parts of the world (Peterson, 1975). The national program in India and in many of the countries which followed were prompted in large measure by concerns that fertility levels existing in much of the developing world, coupled with decreasing mortality levels, would lead to unprecedented high rates of population growth which would undermine social and economic development in these countries, many of which were newly created political entities (Harkavy, 1995, Donaldson and Tsui, 1990, Caldwell and Caldwell, 1986).

The creation of the program in India did not immediately lead to a high level of family planning activity. The first few years of the Indian program were devoted to fact-finding, training, and provision of advice in health clinics. Few countries followed India's example during the 1950s (although China did temporarily alter regulations concerning abortion and contraceptive use in 1956), but the Indian program did generate considerable interest and attention and by 1960 a major conference on family planning programs was held, with proceedings appearing shortly thereafter (Kiser, 1961). In the 1960s, however, there was noticeable momentum. As Table 1 taken from Berelson (1974) shows, by the end of 1972, 31 countries had adopted official policies to reduce their population growth rates and either established national programs to provide services or gave support to private agencies involved in family planning. Another 29 gave support to family planning but for reasons other than reduction of fertility. Although this represented only about 50 percent of developing countries, the early adoption by most of the large countries meant that 87 percent of the developing country population was covered by a policy to make family planning available, but the level of implementation varied widely. Details on the nature of many of the early programs are given in Berelson et al., 1965.

Tables 2a and 2b present the evolution of programs overtime. Table 2a, developed from the valuable series of Factbooks produced by the Population Council, shows that as of 1973 Asia led the way in the number of countries adopting policies with demographic or fertility goals (see also Tsui, 1996, for early development in Asia); but Latin America had the highest proportion of countries with some kind of policy. There was relatively little change, however, in the number of countries with programs for the remainder of the 1970s. The Population Council series ends in 1984, but governmental views and policies concerning fertility levels have also been solicited over a long period by the United Nations. Brennan (<http://ccp.ucr.ac.cr>) displays trends between 1976 and 1993, with the data for the earliest period (1976-1983) paralleling the Population Council pattern shown in table 2a.

**Table 1. Developing countries by policy on family planning by population size, 1972-1973**

Population (in millions)	Official policy to reduce population growth rate	Official support of family planning activities for other reasons	Neither policy nor support
400 and over	China (1962) India (1952)		Brazil
100-400	Indonesia (1968)		
50-100	Bangladesh (1971) Pakistan (1960)	Nigeria (1970) Mexico (1972)	
25-50	Egypt (1965) Iran (1967) South Korea (1961) Philippines (1970) Thailand (1970) Turkey (1965)		Ethiopia Burma
15-25	Morocco (1968) Taiwan (1968) Colombia (1970)	Algeria (1971) South Africa (1966) Sudan (1970) Afghanistan (1970) North Vietnam (1962) South Vietnam (1971) Zaire (1973)	North Korea
10-15	Kenya (1966) West Malaysia (1966) Nepal (1966) Sri Lanka (Ceylon) (1965)	Tanzania (1970) Uganda (1972) Chile (1966) Venezuela (1968)	Peru
10 and under	Botswana (1970) Ghana (1969) Mauritius (1965) Tunisia (1964) Laos (1972) Singapore (1965) Barbados (1967) Dominican Republic (1968) Jamaica (1966) Puerto Rico (1970) Trinidad & Tobago (1967) Fiji (1962) Gilbert & Ellice Isl. (1970)	Dahomey (1969) Gambia (1969) Hong Kong (1956) Bolivia (1968) Costa Rica (1968) Cuba (early 1960s) Ecuador (1968) El Salvador (1968) Guatemala (1969) Haiti (1971) Honduras (1966) Nicaragua (1967) Panama (1969) Paraguay (1972)	Asia: 8 countries; North Africa and the Middle East: 12; Sub-Saharan Africa: 28; Latin America: 5.

Source: Bernard Berelson, (1974), "World Population: Status Report 1974, A Guide for the Concerned Citizen." *Reports on Population/Family Planning*, No. 15, Jan. 1974, based on Dorothy Nortman, "Population and Family Planning Programs: A Factbook." *Reports on Population/Family Planning*, No. 2, September 1973.

This pattern of slow growth during the 1970s is also confirmed by the Ross and Mauldin survey of the strength of program effort scores based on judgments of knowledgeable informants who rate programs on many dimensions (Ross and Mauldin, 1996). Table 2b shows that for all the countries surveyed, the proportion without programs or very weak programs increased between 1972 and 1982, reflecting in part the increasing base of newly established countries. There was, however, substantial movement between 1982 and 1989 and continued development to 1994. Summarizing these trends in terms of scores, Ross and Mauldin (1996, Table 3) find that the average score (with each country weighted equally) advanced from 22 in 1972, to 32 in 1982, 47 in 1989, and 50 in 1994. In addition to the material cited above, statistical tabulations of various program characteristics appear in Mauldin and Lapham, 1987; Ross, Mauldin, and Miller, 1993; and EVALUATION, 1996.

**Table 2a Number of developing countries by government position on family planning by region for selected years**

	1970	1973	1979	1984
All developing countries				
- All positions	102	118	131	134
- Official antinatalist policy and a family planning program	25	31	35	37
- Support of family planning activities and no official policy	17	28	31	33
- Little or no support of family planning activities and no official antinatalist policy	60	59	65	64
Africa				
- All positions	42	47	52	52
- Official antinatalist policy and a family planning program	6	7	8	8
- Support of family planning activities and no official policy	5	9	17	19
- Little or no support of family planning activities and no official antinatalist policy	31	31	27	25
Asia				
- All positions	37	42	46	50
- Official antinatalist policy and a family planning program	14	18	18	19
- Support of family planning activities and no official policy	1	5	3	3
- Little or no support of family planning activities and no official antinatalist policy	22	19	25	28
Latin America				
- All positions	23	29	33	32
- Official antinatalist policy and a family planning program	5	6	9	10
- Support of family planning activities and no official policy	11	14	11	11
- Little or no support of family planning activities and no official antinatalist policy	7	9	13	11

Source: Nortman (1970), Nortman and Hofstatter (1973, and 1980), and Nortman (1985)

**Table 2b. Number of countries surveyed by strength of family planning program, 1972,1982, 1989, 1994**

	1972	1982	1989	1994
Same countries surveyed in each time period				
- Number of countries	77	77	77	77
- Strong	8	8	13	14
- Moderate	11	15	30	32
- Weak	9	22	23	30
- Very weak or none	49	32	11	1
All countries included in survey				
- Number of countries	93	97	98	94
- Strong	10	9	14	14
- Moderate	15	16	29	38
- Weak	31	27	37	40
- Very weak or none	37	45	18	2

Source: Ross and Mauldin (1996)

The development of large-scale social interventions involves societal values that speak to the worth of the goals pursued and to agreement about the means for achieving them. Differences in the priorities and resources accorded to one program versus another, and disagreements about the strategies best employed to achieve desired goals can lead to considerable controversy (Mauldin and Acsadi, 1975). Social action programs also have to meet a set of "validity assumptions" which are more narrow than values and speak to the reasonableness of the assumptions which guide program objectives and the specific mechanisms that are being implemented to achieve them (Suchman, 1967). All family planning programs depend implicitly on a underlying theory of what determines fertility, the appropriate interventions for a given time and place, and the ability to mount these interventions successfully.

Family planning programs have been controversial on these accounts almost from the start, with some critics questioning whether programs in the absence of sufficient social or economic development could have much impact, and others questioning the apparent evidence that many couples were having more children than they wanted and thus might adopt contraception if made available to them in an appropriate manner (Davis, 1967, Hauser, 1967). A recent review of the controversies and issues surrounding these programs is given by Seltzer (2002).

Over the ensuing years, programs have evolved in a number of directions, their effectiveness continues to be widely debated, and they have been studied and evaluated from a number of perspectives. Expert and public opinion, the structure of programs, and the related research have been closely interrelated. Programs have responded by changes in structure and operations to challenges and opportunities posed by critics and by changing socioeconomic conditions; and evaluation techniques have evolved to reflect the changing shape of programs and the questions posed, as well as the internal dynamics of accumulated experience. This intensive scrutiny has led to a number of syntheses of program development and evaluation techniques over the years. Those particularly devoted to techniques of evaluation include: Chandrasekeran and Hermalin, 1975; Ross and Forrest, 1978; Hermalin, 1982a; Sherris et al., 1985; Lloyd and Ross, 1989;

United Nations, 1978, 1979, 1982, 1985, 1986; Ross and Lloyd, 1992; Buckner et al., 1995; Bertrand et al., 1996; Ahlburg and Diamond, 1996.

This chapter draws on this material to trace some of the changes in program development and program evaluation over the last 35 years as a backdrop to the more intensive analyses of specific techniques and issues pursued in subsequent chapters. In particular it identifies and describes the major techniques developed and used over several periods in terms of the questions posed, the data required, and the underlying structure. For the most part the presentation proceeds chronologically, to emphasize the relationship of the techniques to the changing shape of programs, and the accumulation of insights over time. The next section however briefly reviews some aspects of program structure and the nature of evaluation as guides to the range of techniques treated in some detail.

## **PROGRAM STRUCTURE AND THE NATURE OF EVALUATION**

Family planning programs may be thought of as nationally scaled interventions, with demographic or fertility goals and some allocated resources toward these goals. The goals may be implicit or explicit, the means well or poorly chosen. Similarly, the resources can be generous or deficient, efficiently used or not, evenly or unevenly spread. (For additional definitions see Freedman and Berelson, 1976, and Tsui, 1996.) As such, they are complex amalgams of policies, resources, services, supplies, personnel and infrastructure with multiple goals that may be aligned in terms of immediate, intermediate, and ultimate objectives. It is often useful to think of them from an organizational or systems perspective (Tsui, 1996; Simmons and Simmons, 1987; and Buckner et al. 1995). From this standpoint, they are systems that convert inputs (financial, management and policy resources) into processes (personnel recruitment and training, selection of delivery points, information programs and other activities) to yield program outputs (program structure and service utilization) that generate outcomes (i.e. population effects) at the intermediate (e.g. contraceptive prevalence) and ultimate level (fertility change). [See Tsui and Hermalin, <http://ccp.ucr.ac.cr>]

By the same token, evaluation is a multifaceted concept, which incorporates a range of activities. Evaluation is generally regarded as the process of assessing whether the goals of a program are being met but several different types of evaluation usually come into play in any complex program. Table 3 (from Buckner et al., 1995, p.17) presents one classification of types of evaluation and their uses. As shown, evaluating the relevance and adequacy of a program are a priori judgments tied to needs assessment and the validity assumptions referred to above, and often come into play in the launching of a program but, as noted, can also serve as critiques of ongoing efforts. Evaluation of progress is associated with the processes and activities of a program and usually entails keeping track of the quantity and quality of the various stages such as the number of personnel trained, their proficiency, the range and quantity of supplies, etc. Evaluating effectiveness and impact speak to assessing the intermediate and ultimate goals of a program and introduce the question of methodology by suggesting that adequate cause and effect techniques are desired at this stage.

**Table 3. Types/Potential Objectives of Program Evaluation**

Evaluation Component	Definition
Relevance	Evaluation of the appropriateness or equity of a program, or the correspondence between the program and the needs for the program, that is based specifically on a priori judgment.
Adequacy	An evaluation based specifically on an a priori judgment of the extent to which a program is likely to be able to address the entire range of a problem.
Progress	Evaluation of the extent to which scheduled activities occur on time, in the manner expected (according to professional standards, for example), at the budgeted cost, and producing expected outputs.
Effectiveness	Evaluation of the extent to which the program has produced expected intermediate outcomes (effects). Tests for a causal connection between the program and the impact.
Impact	Evaluation of the extent to which the program has produced expected ultimate outcomes (impacts). Tests for a causal connection between the program and the impact.
Efficiency	Evaluation that assesses appropriateness of the level of input (cost) for the level of outcome, whether intermediate (effects) or ultimate (impacts).
Sustainability	Evaluation of whether a program can capture the needed resources to sustain itself after the withdrawal of external support.

Source: Veney, James E. and Pamina Gorbach (1993), "Definitions for Program Evaluation Terms," Chapel Hill, NC: The EVALUATION Project (Working Paper Series No. WP-TR-01); and World Health Organization. (1981), *Health Programme Evaluation: Guiding Principles for its Application in the Managerial Process for National Health Development*. Geneva: World Health Organization.

This highlights an important distinction that has developed in the evaluation literature between monitoring and assessing impact. The former term refers to the development of indicators and measures for tracking the various components of a program to judge whether progress is being made according to plan, to detect shortcomings and identify possible remedies. As such, any phase of a program can be monitored and the generation of indicators and benchmarks relevant to each component often receives considerable attention from program managers. Impact assessment on the other hand analyzes the extent to which there is a cause and effect relationship between program inputs and program objectives. Impact assessments are therefore usually reserved for evaluating the intermediate and ultimate effects of a program but theoretically many of the processes can be studied in this way as well if proper design and data collection are provided for (Bertrand et al., 1996). In this review we will focus on the techniques that seek to assess the impact of programs on fertility as the ultimate outcome and, to a lesser extent, on contraceptive adoption and use as an intermediate outcome. It is worth noting that the demarcation between impact assessment and monitoring is somewhat fluid and subject to change over time as our data and techniques grow more sophisticated. For example, a number of the techniques introduced in the early phases of program development (like Couple Years of Protection-CYP), which were

treated as speaking to the impact of a program, were later seen to be more akin to an indicator of program progress because they were not adequate for establishing cause and effect relationships between program input and fertility change. These possible changes in interpretation should be kept in mind as we review the various techniques that have been utilized at different stages.

Two more potential objectives of program evaluation listed in Table 3 are efficiency and sustainability. The former speaks to the question of program costs in relation to effectiveness and impact. It is to be noted that none of the techniques described below explicitly addresses that issue, although the potential exists in experimental design, country level analyses, and several other approaches that can incorporate program inputs to analyze some aspects of the question. [Knowles, <http://ccp.ucr.ac.cr>, treats the issues of measuring costs in some detail.]

The issue of costs and efficiency is also a main impetus for operations research studies that have been carried out extensively in connection with family planning programs. These have as their focus how various facets of program operations can be redesigned to improve cost-effectiveness.) (See Fisher et al., 1991, Foreit et al. 1990, Ross et al. 1987, Wawer et al., 1991, Bertrand and Brown, 1997, Gallen and Rinehart, 1986.) Sustainability as defined in Table 3 also has a financial connotation, as it addresses the question of whether a program has reached a point where it can be self-sustaining, particularly in the absence of external funding. The concept of sustainability has received considerable attention in recent years and it has broadened beyond the purely financial to incorporate the sustainability of the “outcomes” that the program wishes to achieve. From this standpoint, the question is whether socio-economic and demographic changes ongoing in a society lead to such widespread acceptance and use of contraception that an explicit program to foster adoption is no longer needed or might be recast to serve only special groups within the population.

A dimension important to program objectives and their evaluation, not explicitly captured in Table 3, is the adequacy of underlying timetables. Realistic timetables for the achievement of primary and secondary goals are needed to utilize resources well and avoid premature detours, or costly persistence of an unproductive intervention. As illustration of the importance of proper correspondence between goals and actions and their relation to evaluation, we may divide, conceptually at least, programs into those that are mainly designed to affect desired family size and those whose goal is to assist couples achieve their desired number of children. In the first case one would expect that program operations in their early phase would concentrate heavily on mass media and forms of outreach that would encourage and legitimize smaller family sizes and in a second phase stress the service and accessibility that allow for a rapid implementation of newly established norms. One would not expect to see a reduction in fertility until there is both an overall reduction in desired family size and an opportunity to implement these new desires. A program that focuses on accessibility instead of mass media and other motivational outreach, where desired family size is high (and couples are not experiencing “excess” supply), is not likely to have much success. By the same token, efforts to evaluate programs before they have had an opportunity to promote lower fertility norms and practice may be premature in their judgment.

Programs that seek to assist couples achieve their desired family size are likely to focus their efforts on providing information and services about contraceptive use. These programs can only achieve a reduction in fertility if the perception that couples are having more children than desired is correct. Moreover, from an evaluation standpoint, since the demand for family planning precedes the program to some degree, program impact must be assessed in part by showing that a reduction in fertility was accelerated or broadened by the actions of the program, or that provisions of services hastened the adoption of contraception by those already motivated.



In several respects the foregoing scenarios are simplified in that measuring the demand for children or contraceptive services is not always straightforward. Couples and women are often conflicted about whether they wish additional children and the benefits and dangers of undertaking contraception. Several analysts view the major contribution of effective family planning program efforts as serving to activate and crystallize latent demand for contraception among couples who do not desire additional children, by relaxing the uncertainty and hesitation often associated with innovative practices (Freedman, 1997, p 2 ). (See also Westoff, <http://ccp.ucr.ac.cr>, on estimating demand.) Similarly, the measurement of exposure to program services and program inputs can be elusive and this greatly complicates the assessment of program effects since evaluation hinges on showing a relationship between the level of inputs and the outcomes associated with program goals. Weak and unreliable measures of program services to individuals or communities can confound estimates of program effects. These considerations will be addressed further in discussing several of the specific techniques of evaluation.

Table 4 lists the seven broad classes of techniques to be covered along with the specific methodology associated with each. The classes are roughly in chronological order and correspond with three distinct periods of family planning structure and with ideas about the underlying factors affecting fertility levels. They also tend to differ in the type of data employed, their underlying assumptions, and in the way they detect and measure program effects. The first three classes were widely used in the earliest period of family planning activity and we take them up first.

**Table 4. Methods for Evaluating the Impact of Family Planning Programs Classified by Type**

- A. Demographic and statistical techniques
  - 1. Standardization and decomposition
  - 2. Trend analysis
- B. Analysis of acceptor data
  - 1. (Standard) Couple years of protection
  - 2. Analysis of the reproductive process
  - 3. Component projection
  - 4. Simulation
- C. Experimental designs
  - 1. Random experiments
  - 2. Quasi-experimental designs
  - 3. Matching studies
- D. Multivariate areal analysis
  - 1. Subdivisions within a country
  - 2. Countries as units
- E. Population-based surveys
  - 1. Synthesis model
  - 2. Prevalence model
- F. Multi-level strategies
  - 1. Cross-sectional multi-level analysis
  - 2. Longitudinal multi-level designs
  - 3. Diffusion studies
- G. Other techniques
  - 1. Meta-analyses
  - 2. Weight-of-evidence assessments

## THE SUPPLY-ORIENTED PERIOD OF EVALUATION

In this section we focus on the early period of family planning programs, from the early 1960s to about the mid-1970s. During this period, programs were largely structured around fixed delivery points, there were a limited number of modern methods available in each program, and the interactions with the client often permitted the collection of data about the characteristics of each acceptor as well as their experience after acceptance. Accordingly it was possible to develop an accurate service statistics database based on acceptors of various types of services. It was also a period when several major experiments involving family planning strategies were undertaken and although experimental designs persisted beyond this period, the major contours of this approach are clearly visible early on.

We start however with the two techniques under category A in table 4 that do not directly employ any program data but which were used to make inferences about the possible impact of the program. The first of these, standardization and decomposition refers to standard techniques in demography for controlling the influence of one or more factors in order to gauge the effect of another, as in the development of standardized death rates to control for the effect of age structure on crude death rates. In many countries the measure of fertility most often available is the crude birth rate and it is well known that it can be influenced by age distributions (the proportion of the population who are reproductive age women as well as the internal age distribution of the women), and the proportion of women married as well as the age specific marital fertility rates. As it is the latter factor, which the family planning program seeks to change, standardization or decomposition was often undertaken to control the crude birth rates for the other factors to see if there was any decline in the marital fertility rates. If not, it was presumed that the program was not having any impact. If there were declines, the technique could not determine how much of the change to attribute to the program but the potential of a program effect had been established. As such the technique is not really a method of assessing impact but a tool for preliminary analysis to determine if a program effect may have occurred for the time period under analysis. (For additional information see United Nations, 1979, 7-33, and Buckner et al. 1995, 70-74.)

The other technique shown under category A, trend analysis, also relies on demographic data but utilizes a time series of observations of a fertility measure to make strong inferences about program effects. The basic methodology is to project forward the series of fertility rates in effect before a program is launched and compare the projected trend with that observed after the program is initiated. The difference between the fertility rate observed and that projected is taken as an estimate of the effect of the program, on the assumption that without the program the pre-program trend would have remained in force. This strong assumption, plus the difficulty of identifying the exact date of program initiation, and the sensitivity of the results to the number of observations employed in defining the trends have limited the utility of this approach. (For additional information see Mauldin, 1979; Hermalin, 1982a; Buckner et al., 1995.)

The techniques shown in category B have in common strong reliance on the acceptor data described above. In effect they take account of the number of acceptors, the expected length and effectiveness of use of the contraceptive adopted, and then translate these parameters into their effect on fertility. Differences arise in the exact methods used for the translation, the type of ancillary data employed and the time dimension for the effects estimated. As example, couple years of protection estimates the total years of protection or the births ever to be averted as a result of program acceptances in a single year, but does not allocate them to future calendar years. Similarly, the analysis of the reproductive process does the same thing for each segment of use by each acceptor of a specific method, employing additional information about the women's fecundity status among other characteristics. In effect, both methods, though they differ

considerably in detail, combine the effective period of use of a contraceptive method with the expected fertility over that period to estimate births averted. In contrast, component projection techniques employ the expected period of use to project forward the number of women protected at each point in time and then combine this number with the expected fertility of the acceptors to estimate the births averted in a specific period of calendar time. In contrast to these acceptor-based techniques, simulation is a general strategy like standardization that can be employed to investigate a large range of demographic phenomena. In the context of family planning evaluation, it is possible to simulate the reproductive process taking into account a number of biological characteristics like fecundity, length of post-partum amenorrhea, as well as effectiveness and use of a contraceptive method, and demographic contingencies such as mortality, and marital formation and dissolution. Simulations can vary considerably in approach from analytic to numerical and from macrosimulation to microsimulation methods. They can provide estimates of total births averted for a cohort of acceptors or generate projections for calendar years. An example of the latter type is CONVERSE (Nortman, 1979).

Consideration of the acceptor-based techniques brings to the fore two key conceptual issues. Although all the techniques produce an estimate of program effects in the sense of estimating births averted (or related measures) as a result of program activity, these estimates credit the program with all acceptors from program sources, including those who switched from private sources and, more hypothetically, those who might have adopted from a private source if program methods were not available. For this reason these estimates are termed gross program effects, in distinction from net program effects that would take into account these substitutions. The gross program effects, however, do not include those who were stimulated by program activity to adopt contraception but chose to do so from non-program sources--often termed the catalytic effect of programs. The net program effects would credit these acceptors to the family planning program. Data are rarely available to estimate these offsetting trends directly but the assumptions embedded in each technique lead to different types of estimates. (For additional information on gross and net program effects see United Nations, 1986, p. 35.)

Another important concept embedded in the acceptor based methods is that of potential fertility--the fertility that the acceptors would have experienced in the absence of the program. A number of different estimates have been employed for this key parameter, including: fertility rates of the whole population; the acceptors' own fertility in the period prior to entering the program; fertility of a matched group of non-acceptors; estimated natural fertility rates for the population. Analysis of these techniques shows that they are quite sensitive to the estimates of potential fertility employed and empirical studies of the application of different techniques to the same underlying data also reveal that potential fertility assumptions play a large role in the variance of the program effect estimates (Potter, 1981). (Additional details on these acceptor based methods may be found in Gorosh and Wolfers, 1979; Potter, 1979; Nortman, 1979; Menken, 1975.)

In later periods attention has shifted away from these acceptor based methods for two reasons. Changes in the structure of family planning programs in terms of community based distribution of contraceptives and social marketing strategies, described more fully below, undermined the acceptor database that are essential inputs into these techniques. At the same time there was increasing recognition that by generating only gross program effects and ignoring the socioeconomic changes ongoing in many countries that might be contributing to increased demand and use of contraception, these estimates were overstating the role of programs. Continued scrutiny of program impact required a more nuanced approach. Although these techniques are rarely employed now to estimate the impact of programs on fertility, it is worth noting that couples years of protection remains an important tool for monitoring program progress.

Experimental designs of various types were employed in the early stages of program development in a number of settings. Indeed, within a few years of the start of India's program experiments were underway there, and by the mid-1960s, major experiments were in operation in several other countries. Experiments of course are a general scientific tool for measuring the effects of one or more treatments and have been used extensively in agriculture, medicine, and a variety of social interventions. Theoretically, a well-designed experiment can provide unambiguous evidence of whether a program is having an impact on contraceptive adoption and/or fertility levels as well as a strong estimate of the magnitude of those effects. Experiments provide estimates of net program effect, because the changes in the control groups reflect the effects of socioeconomic and other forces except for the program, so the contrast with the treatment groups provides only the effect of the program.

As previously noted, although experiments seem like an ideal means for assessing the impacts of programs, they have been used relatively little, especially in recent years, and this is especially true of experiments designed to measure the impacts of programs on fertility. As shown in Table 5, in 1977 Cuca and Pierce reviewed 96 experiments and quasi-experiments, and considered only 12 of them to be true experiments; 41 of the 96 studies employed controls and 55 did not. A more recent assessment (by Bauman et al., 1994) identified only 16 true experiments after an extensive search. In the realm of operations research, a major review (Gallen and Rinehart, 1986) found only one true experiment out of 143 examined.

**Table 5. Experimental designs: true experiments and quasi-experiments**

Numbers identified and reviewed	
Cuca and Pierce (1977)	
Total number reviewed	96
Number with controls	41
With random assignment	12
Number without controls	55
Bauman, Viadro, Tsui (1993)	
Total number of true experiments in family planning identified after extensive search	15
Population Information Program (1986)	
Review of family planning operations research projects completed, 1947-1986	
Number reviewed	143
Number of true experiments	1

Note: Of the 12 “true” experiments described by Cuca and Pierce, Bauman et al. regard three as not meeting the necessary criteria, and they regard one classified by Cuca and Pierce as a quasi-experiment as a true experiment. In addition, Bauman et al. identify five true experiments occurring after the Cuca and Pierce analysis.

One of the most influential experiments was that carried out in the city of Taichung in Taiwan from 1961 to 1963, in which city neighborhoods were randomly assigned to different treatment areas which varied the level of contact from the program (home visits only, mailings only, or both, vs. None) and the target (wife only or husband and wife) in a systematic manner to gauge the level of acceptance of IUD's over a two-year period (Freedman and Takeshita, 1969). The range of treatments studied in the Taichung experiment, although highly focused, was broader than that employed in most experimental and quasi-experimental designs.

An important and closely watched quasi-experiment has been conducted in Matlab, Bangladesh. It has benefited from a long period of observation, a broad series of inputs that have varied over time, and high quality and frequent measures of fertility. The differences in the degree of contraceptive use and the fertility change associated with the earlier more limited community based distribution plan, and the later more extensive outreach efforts have been influential in suggesting how programs can achieve effects in rural, low income settings (Phillips et al., 1988).

Table 6 presents the contents of the experiments as classified by Bauman (1995) and Cuca and Pierce (1977). Bauman presents detailed categories of the true experiments, while Cuca and Pierce present broader classes for all the designs. Both approaches indicate that the emphasis has been on specific and rather local questions of program operation. No comprehensive plans have been evaluated through this approach, and it is not clear whether this reflects limitations in the questions experimental designs are best able to address, or failures on the part of those utilizing this technique to take full advantage of its potential.

**Table 6. Topics covered by experiments**

True experiments (Bauman, 1995)	True and quasi-experiments (Cuca and Pierce, 1977)
Home visit	Personnel: type; motivation; pay
Home visitor qualification	Mass media
Wife vs. couple involvement in home visit	Integration
Mailed pamphlet	Intensity
Single- vs. multi-purpose field worker	Incentives
Field worker incentive	Inundation-distribution
Field worker affiliation	
Satisfied user teamed with midwife	
Geographic distribution of health promoter	
Mother education class	
Prenatal education	
Early clinic return schedule	
Frequency of clinic supervisory visits	
Frequency of physician at clinic	
Pill prescription by midwife	
Location of condoms in supermarket	

Experimental designs also differ greatly in their durations. Table 7, taken from Cuca and Pierce (1977) displays the duration by types of design and reveals that generally the true experiments

had the shortest durations, while the quasi-experiments, particularly those without control groups, were among the longest. It is possible that the planned duration influenced the design employed, but to determine this would require a detailed assessment of each experiment's strategy. Nevertheless, insofar as longer experiments are more costly, there does appear to have been an undue allocation of resources to the weaker designs. Assessments of the outcomes of the experiments in terms of their effects are given by Freedman and Berelson, 1976; Mauldin, 1983; and Bauman, 1995, the latter via a meta-analysis of the 15 true experiments he identified.

**Table 7. Duration of experiments**  
**Duration of experiments and quasi-experiments, number by type and duration**

Duration	True experiments		Quasi-experiments		No
		Control/pretest pretest	Control only	Pretest only	
0-5 months	4	2	2	1	8
6-12 months <sup>a</sup>	2	3	1	1	1
1-2 years <sup>b</sup>	5	6	3	6	8
2-4 years	1	4	0	5	11
5 or more years	0	7	0	8	7
Total	12	22	6	21	35

Source: Derived from Cuca and Pierce, (1977) Table 4.

<sup>a</sup> Appears as "6-14" months in original; assumed to be a typographical error.

<sup>b</sup> Appears as "1 year" in original; assumed intent was 1-2 years. (See also Cuca and Pierce, Table 2.)

Note: Of the five true experiments which occurred after 1977, as described by Bauman et al., (1993), two had durations between 1-2 years and three had durations over 2 years, so that a distribution of all true experiments by duration would show a more equal distribution across categories.

Though experiments were more often employed early in the history of family planning programs, interest in this strategy has persisted, with calls for their greater use in current assessments of family planning program and reproductive health interventions. Bauman et al., (1994) discuss the merits of experimental designs as well as the problems often attributed to them and show how these can be minimized in appropriately designed interventions.

## THE DEMAND-ORIENTED PERIOD OF EVALUATION

By the mid-1970s, family planning programs were becoming increasingly complex in structure. Many programs had adopted some form of community based distribution in which contraceptives would be disseminated through depots in each village or door to door and a number of social marketing schemes were underway in which contraceptives were made available through commercial and other channels often under subsidy by the government program. An inventory by Foreit et al. (1978) identifies 83 non-clinical programs as of early 1978, almost all initiated after 1970. As noted, these developments undermined to a considerable extent the ability to employ acceptor-based methods of measuring impact since it became increasingly difficult to record acceptor characteristics and track method switching and length of use under these supply arrangements. In addition, programs were increasingly asked to provide more nuanced estimates

of program effects that took into account the levels of social and economic change underway in much of the developing world. In the absence of random experiments a number of alternate strategies were developed which made use of data on geographic divisions within countries, countries themselves as units of analysis, and the rapidly expanding portfolio of population based surveys generated by the World Fertility Survey and Contraceptive Population Survey projects.

The World Fertility Survey, which started in 1972, was a program to carry out fertility surveys in a comparable manner in a large number of developing countries. Among its goals were high-level estimates of recent and lifetime fertility, details on methods and duration of contraception used, and information on women's preferences as to number of children desired along with information on characteristics that might explain variation in fertility and contraceptive behaviors (Cleland and Hobcraft, 1985). Aside from a community module used in some countries that produced estimates from informants on time or distance to the nearest family planning facility, little attention was given to measuring family planning program inputs and this greatly limited their potential for assessing program impact. From an analytic standpoint the surveys were used mainly for studies that sought to understand couple's demands for children and family planning and for a range of comparative analyses. Many of the models and analyses from this perspective are represented in Bulatao and Lee, 1983.

The Contraceptive Prevalence Surveys were another set of surveys carried out during this period with the technical assistance of the Family Planning Evaluation Division of the US Centers for Disease Control (Morris and Anderson, 1981). These surveys did have a more explicit focus on measuring accessibility of family planning programs but since the information was collected almost exclusively from respondents, including non-users, with differing contact and knowledge of different facilities it proved difficult to use for this purpose. The challenge of utilizing surveys for the analysis of programs was the focus of a IUSSP-sponsored seminar in 1980 (Hermalin and Entwisle, 1982).

Despite their limitations with regard to evaluating the impact of programs, the survey data were rich enough to permit some assessment of program effects and two approaches to their use will be treated briefly in this section. As a major goal of measuring impact in this period was to establish the effects of programs independent of the changes in fertility and family planning that might have resulted from changes in the levels of economic development, almost all of the techniques to be discussed here are statistical regression strategies in which a measure of fertility or contraceptive use is regressed against a set of independent variables that include both family planning and socioeconomic characteristics. They differ mainly in the unit of analysis employed and this choice greatly circumscribes the range of measures generally available. Table 8 sets forth the general structure of the regression approach, along with illustrative variables associated with each strategy. The first two approaches rely on areas as the units of analysis and as category D of Table 4 stipulates, geographic divisions within a single country as well as entire countries have been employed. Surveys of individuals are shown in category E as population-based surveys. It is of course possible to combine analyses from more than one level of observation and these have become increasingly prominent in the study of family planning effects. For this reason, multilevel strategies are treated as a separate category and, since they have come to the fore after the mid-1980s, discussion is reserved for the next section.

**Table 8. Illustrative structure of regression analyses for assessing program impact by unit of analysis and types of variables employed**

Unit	Family Planning Inputs	Exogenous	Other independent variables		Impact measures
			“Intervening”	“Proximate”	
Communities within a country	Number of family planning clinics per 1000 pop. Number of clinics within time/distance parameters Types of services	School attendance Percent completing primary/secondary school Labor force part. rates (for women) Percent in tertiary occupations Population density Distance to cities	Infant or child mortality rate	Proportion women married (over all or age-specific)	Age-specific or total fertility rates Changes in fertility Child/woman ratio Contraceptive prevalence
Countries	Program effort scores Components of effort scores Amounts from donors	GDP per capita Female education Male education Percent in tertiary occupations Percent urban Religious composition Health measures Communication measures	Infant or child mortality rate	Proportion married Average age at marriage	Total fertility rate Wanted fertility rate Contraceptive prevalence
Individuals [Only or combined with other levels]	Perceived accessibility Actual accessibility <sup>1</sup> Worker inputs <sup>1</sup> Expenditures <sup>1</sup> Number of available methods <sup>1</sup> CBD services <sup>1</sup>	Age Education Residence Occupation Wages, economic level, income [Aggregate measures] <sup>2</sup>	Desired number of children Want more children	Fecundity Age at marriage (duration) Breast-feeding	Recent fertility CEB Children 0-4 Birth intervals Contraceptive use

Note: Variables in each column are illustrative of those found in the literature and do not represent any particular analysis.

<sup>1</sup> Insofar as these inputs are measured at a community (village, state, or province) level or country level, the analysis would be multi-level in design.

<sup>2</sup> Exogenous variables with individuals as units of analysis can also include community variables of an ecological, health, or communication nature, leading to a multi-level design.



A number of areal multivariate analyses using geographic subdivisions within a country were undertaken in the 1970s as a way of measuring family planning effects while taking into account socioeconomic factors. A search reveals relevant analyses in Chile, China, Colombia, Costa Rica, India, Indonesia, Korea, Pakistan, Taiwan, and Thailand, and some countries have been the subject of more than one analysis (Hermalin and Khadr, 1996). They derive from an ecological tradition of analysis within demography for the study of differential fertility or mortality at a time when data at the individual level was seldom available, and for the same reason they have played a prominent role in historical demography, as exemplified by the European fertility project (Coale and Watkins, 1986). From the standpoint of measuring the impact of family planning programs, areal multivariate analysis was widely used before the widespread availability of population-based surveys.

The basic structure of an areal multivariate level analysis is set forth in Table 8. One requires for some geographic division within a country (or other territorial unit), measures of program input (which often come from administrative records of the program), other independent variables that reflect the socioeconomic and demographic characteristics of the unit (which are often taken from censuses and national accounts) and as dependent variable an appropriate measure of fertility (taken from censuses and vital statistics, and/or registration systems) or contraceptive prevalence (often derived from program records).

The geographic level chosen depends on available data in large part and the needs of statistical analyses. One desires sufficient units to sustain the multivariate analysis, and to utilize geographic areas that are meaningful units of social interaction. In practice provinces, states, counties, townships, districts, municipalities and others have all been employed in these analyses, and are referred to generically as local areas. Extensive treatments of multivariate areal analyses as a technique of measuring the effect of programs, along with illustrative examples have been provided by Hermalin (1975, 1979) and it is also included in many of the overviews of methods of evaluating programs cited earlier.

Compared to using countries as the units of analysis, using areas within a country has the advantage that there is likely to be more uniformity within a country on the lags between program inputs and response, the nature of the inputs employed, and unmeasured cultural and program factors that can distort cross-national analyses. At the same time areal multivariate analysis has its own limitations, notable among them being the possible bias arising from lack of information about how resources are allocated, the difficulty of mounting strong theory at the areal level and properly controlling for individual differences that can affect behavior. Nevertheless the approach is not without appeal given that programs are characteristics of communities and interest centers on their effect on population based fertility outcomes, in competition with ongoing socioeconomic developments. Hermalin and Khadr's review (1996) of 14 analyses across 11 countries indicates that on balance they show a moderate negative effect of program inputs on fertility, but this conclusion is limited by our lack of knowledge of resource allocation, as discussed below.

Multivariate analyses with countries as units of analysis got underway in the early 1970s with the availability of family planning program effort scores for large numbers of countries. These scores which attempt to represent the strength of programs and thus could be taken as a measure of program inputs, could be combined with a number of demographic and socioeconomic measures about each country usually available from censuses and other governmental statistical efforts to estimate the effect of program efforts on contraceptive use or fertility outcomes.

A number of multiple regression or path analytic models have been carried out across countries using a variety of socioeconomic indicators (sometimes separately and sometimes combined into a single index), program effort scores (in total and occasionally by component) and various measures of fertility (crude birth rates and total fertility rates, at a point in time or as declines over a period). These include Freedman and Berelson (1976), Mauldin and Berelson (1978), Tsui and Bogue (1978), Cutright (1983), Ness and Ando (1984), Lapham and Mauldin (1984), Tolnay and Christenson (1984), Mauldin and Ross (1991). These studies, which collectively cover the period from 1965 to 1990, although they differ to some degree in their approach, all point to the same finding--family planning program strength has a strong negative effect on fertility, even after taking into account a variety of socioeconomic factors. (These studies also tend to find that the socioeconomic dimensions strongly affect fertility and that socioeconomic factors and family planning strength are correlated, but not to a very high degree.)

Taken at face value, these eight studies provide strong evidence for a sizable effect of family planning programs on fertility. But they are not without controversy, centering around the themes of adequacy of theory and estimation.

Demeny (1979) in an influential article stated that "program effort may be a reflection of underlying fertility determinants not grasped by the available socioeconomic indicators." Accordingly, models that do not fully capture these indicators may be giving credit to family planning program efforts that properly belong elsewhere. Hernandez (1984) re-estimated several models and found lower effects from programs than the originals, though his strategy was also questioned (Tolnay, 1987).

Schultz (1997), like Demeny, also argues, on the basis of his statistical analyses, that family planning program effort scores cannot be used as an exogenous variable to study effects on fertility levels or change because of the strong likelihood that these scores are affected by many factors in society which relate to the determinants of fertility or to the degree of receptiveness to contraceptive use. He concludes that cross-national analyses are not satisfactory approaches to measuring the effect of programs or similar interventions. Pritchett's (1994) critique of previous research at the country level is from a different tack. He uses country level data to argue that observed actual fertility rates are mainly driven by desired fertility, and that unmet need is not an important determinant of fertility. Insofar as programs mainly address unmet need through facilitating access to services they can have only little impact on over-all fertility rates.

There have been several rebuttals of Pritchett's thesis and analyses (Knowles et al., 1994; Bongaarts, 1994), some on statistical grounds and some based on more substantive factors. Bongaarts, <http://ccp.ucr.ac.cr>, addresses many of Pritchett's criticisms, pointing up for example that programs have an effect that goes beyond simply providing access, through the provision of information and serving to legitimate the use of contraception. He also shows that the relationship between wanted and unwanted fertility depends very much on the strength of family planning efforts.

Where do these conflicting views of cross-country analysis leave us? It is important to note that in almost all cases the differences between critics and supporters of family planning programs concern estimates of the magnitude of program effects rather than whether programs have any effects at all. As Bongaarts notes, even Pritchett estimates that an increase of 50 points in the family planning effort score (which has a scale from 0 to 100) means a reduction of one birth in the total fertility rate. And although Pritchett regards this as a small program impact, Bongaarts shows that taking into account the degree of fertility transition, and the distribution of population

size across countries with weak and strong programs, even this lower bound effect is considerable.

The large number of population surveys which became available through the World Fertility Survey and Contraceptive Prevalence Survey projects led to many studies of the determinants of contraceptive use and fertility using the individual respondents as the unit of analysis but the limited amount of information on family planning program inputs restricted their ability to measure program impact. An influential model by Easterlin (1983) did address the issue by making use of reports by individuals on their degree of knowledge of contraceptive sources and statements about the location or the nearest facility. In this model, a couple's use of contraception is viewed as determined by the number of children they desire, the supply of children in the absence of contraception, and the costs of fertility regulation which incorporate financial as well as social and psychological dimensions. The family planning program factors enter this equation as measures of the costs of regulation. This equation, in turn, is one of a series of equations, including in particular an equation estimating children ever born as a function of the proximate determinants, including contraceptive use. The system of equations thus provides an estimate of program factors on fertility. Easterlin and Crimmins (1985) provide details of the methodology and apply this approach to several countries. Additional discussions of this approach can be found in Montgomery (1987) and in Ahlburg and Diamond (1996). Although the demand-supply cost framework has been very influential in framing the way analysts think about the factors affecting the adoption of contraception, the particular implementation has not been widely followed due in part to the limited data available on program characteristics. The advent of stronger program and community data described below led to alternate strategies of analysis but retained significant elements of the conceptual framework.

Another way of using the data from the population surveys was to take advantage of the strong estimates of fertility and contraceptive prevalence that they produced and the information on sources of contraception obtained (program vs. non-program), to obtain rather direct estimates of program and non-program effects. This approach known as the Prevalence Method (Bongaarts, 1986) makes use of the analytic relationships between natural fertility, observed fertility, contraceptive prevalence and the effectiveness of specific methods, to derive estimates of gross program effects. In effect, the method utilizes the population surveys to provide some of the data that was formerly available through the acceptor information system to generate similar estimates. Extensions of this approach which focus on the factors which lead countries to be outliers from expected levels of fertility, given their prevalence rates, have been proposed by Curtis, Diamond and Rutenberg (1989), and Curtis and Diamond (1995).

Despite the emphasis on the individual and couple in the population based surveys, a number of countries did use the community module of the World Fertility Survey which estimated the distance or travel time to the nearest outlet, among other measure of the community (Casterline, 1987). Several analysts incorporated this information into their models of the determinants of contraceptive along with relevant individual characteristics. These analyses drew attention to the potential of the multilevel strategy as an approach that could combine individual and couple socioeconomic factors impinging on fertility related behaviors, with aspects of program inputs in order to gauge the independent effect of the program. A number of these early efforts and a general review of community effects were discussed in a 1983 WFS conference (Casterline, 1985a). The seminar included a review of the studies carried out to that point which attempted to assess community and program influences on contraceptive behaviors (Tsui, 1985) and another on their effects on fertility (Casterline, 1985b). Both reviews indicated limitations in data, modeling, and estimation and noted that program results were mixed. At the same time they endorsed the underlying multilevel strategy and pointed up ways to strengthen its considerable

potential. These and other activities influenced the content of the Demographic and Health Surveys (DHS), which was the successor project to the WFS, and the types of analyses that might be most productive in understanding the dynamics of fertility behavior.

In summing up the regression-focused strategies that dominated this period, attention must be given to the question of resource allocation as a possible factor confounding many of the analyses. Just as the problem of substitution and estimating potential fertility limited the insights from the service oriented techniques described in the earlier section, the lack of knowledge about how family planning program resources are allocated and the failure to account for this deficiency greatly limit the inferences from cross-sectional regression analyses. The structure of a regression analysis assumes that the family planning program inputs are one of the causes of the outcomes of interest, that is, contraceptive use or fertility level. This assumption may not be warranted if the level of fertility or prevalence has been an influence on the way inputs are allocated across communities. If for example, administrators place resources where fertility is high, cross-sectional analysis is likely to show that family planning inputs are correlated with high fertility. Similarly, if resources are differentially allocated to areas where fertility is low, perhaps on the basis that this is where change is underway, then the regression estimates will overstate the effect of the program. As noted, this possible confounding of cause and effect can also occur with countries as the units of analysis, if policymakers base the decision to initiate a program on observed strong demand for services on the part of the population. Recognition that systematic methods of resource allocation unknown to the analyst could bias the estimates of program effects emerged in a series of papers (Hermalin, 1982b; Rosenzweig and Wolpin, 1986, Schultz, 1992; Gertler and Molyneaux, 1994) and led to the search for more robust techniques which are introduced in the next section and treated in more detail by Guilkey and Diamond, <http://ccp.ucr.ac.cr>.

## **THE FULL MARKETING STRATEGY**

The advent of the Demographic and Health Surveys (DHS) in 1984 marked the beginning of what will be termed a full marketing strategy in the study of family planning effects and the determinants of contraceptive and other fertility related behaviors. Whereas the earliest period stressed the services supplied by the program, and the second period focused on the characteristics of individuals and how these influenced demand, the most recent period has recognized the necessity of incorporating both supply and demand factors. As the relevant techniques are treated at some length in this volume, they are touched on only briefly here.

The general approach for melding supply and demand factors has been multilevel analysis, defined by Hermalin (1986) as "a strategy for combining information from more than one level of observation in studying the determinants of various forms of behavior." He presents a number of its advantages, general structure, and data needs. (An overview is also given in Bertrand et al., 1996, Chapter IV.)

The third tier of Table 8 illustrates the basic structure of a multilevel analysis. Generally, the family planning input measures will be measured at the community or country level along with other socioeconomic characteristics at the aggregate level thought to influence the behavior in question. These are combined with a model of the characteristics at the individual level that also affect the outcomes of interest. Because the regression framework combines variables at more than one level, special issues of estimation arise as noted by Hermalin (1986), Mason (1986). [See also Guilkey and Diamond, <http://ccp.ucr.ac.cr>.] An attractive feature of multilevel analysis is the ability to include interactions across levels. For example it is possible that the educational

level of respondents will interact with the degree of family planning availability in determining contraceptive use, in that communities with high availability may show less differential in use by education in comparison with communities with lower levels of availability (Entwisle et al., 1984). Another important feature of multilevel analysis is the flexibility in terms of the aggregation level employed. Analyses have been carried out combining individual and country level data, particularly in comparative studies from the WFS surveys (Entwisle and Mason, 1985) and within countries by combining the information about program inputs in each community (or cluster) with the individual responses (Mensch et al., 1995; Chamrarithirong et al., 1992). The latter strategy has been greatly aided by the large number of clusters employed in the DHS samples as well as their collection of fairly detailed accessibility data.

The DHS surveys [Vaessen, <http://ccp.ucr.ac.cr>] accelerated multilevel analysis of program efforts on contraceptive use and fertility by incorporating, in many of the countries, a Service Accessibility Module which collected information about the actual availability of services to couples in the clusters sampled. This was done either through a number of informed observers in the community who reported on the location of the nearest facility of various types or the nearest source of various contraceptives, or through actual visits to the facilities named, in order to obtain more details about the services provided. [Wilkinson, <http://ccp.ucr.ac.cr>, provides details of these modules.]

An alternate approach to collecting data about facilities providing family planning services is through a situation analysis, described by Fisher et al. (<http://ccp.ucr.ac.cr>). Here the goal is to inventory the supplies and services provided. Generally the facilities to be studied are drawn as a sample of all facilities of certain types, and therefore the location of the facilities may not coincide with the communities in which couples have been interviewed by DHS or other population surveys. Hence the detailed data about facilities ordinarily cannot be combined with characteristics of actual and potential clients.

Multilevel modeling has been aided by the advent of digital geographic mapping techniques which permit a clearer delineation of the catchment areas for family planning services as well as the transportation and travel patterns of actual and potential clients. It is thus possible to calibrate much more effectively the definition of community and community and program characteristics with the characteristics of the women and households. [See Rosero-Bixby et al. (<http://ccp.ucr.ac.cr>) for examples of the insights from this technology.]

The data collected by the SAM and these other strategies can generate a very large number of variables concerning the nature and structure of family planning facilities of different types, their locations, the size, characteristics, and training of personnel in attendance, and the presence or absence of various methods and services. But it is not clear that any particular measure or combination of measures adequately captures the degree of exposure to and relevant program influences on potential and current users of contraception. In the absence of strong theory, the analyst is faced with the difficult problem of data reduction in deciding how to incorporate the considerable amount of information generated into equations, which test the effect of programs. As previously noted, insofar as the measures employed are insufficient to represent the program environment, estimates of program effects can be misleading. In addition the availability of these data do not obviate certain underlying problems confronting the multilevel strategy. Chief among these is the problem alluded to above, the lack of knowledge about how inputs are allocated and the potential bias arising from this source in estimating program effects. In addition, there is often insufficient knowledge of the history of program operations in each community, which prevents proper alignment of program input with each woman's fertility history.

Two approaches for relieving these constraints are the use of multi-equation random effects models, in which observed and unobserved factors influencing program variables are introduced into a series of equations determining program locations, as well as the behavioral outcomes of interest. Another approach has been to utilize fixed effects panel designs, in which the changes in outcomes are regressed against changes in program variables and changes in other factors. By focusing on modeling the changes overtime, invariant factors particular to each geographic area, which might be confounding the estimation of program effects, are largely controlled. These approaches are described and illustrated in Bertrand et al. (1996) who also cite the relevant literature. (Guilkey and Diamond, <http://ccp.ucr.ac.cr>, review the issues and provides examples).

Another important feature of the DHS surveys is the use of a calendar to collect detailed data, usually over the last five years, of contraceptive use by method for each month, reasons for discontinuance, as well as data on pregnancy, breastfeeding, amenorrhea and several other behaviors. [Vaessen, and Curtis (<http://ccp.ucr.ac.cr>.)] These data reconstruct in a sense the service statistics data that was traditionally available when women returned regularly to the same clinic for methods and service, and consequently permit detailed analysis on method acceptance and discontinuation over the period of observation. Having these data over time allows for the utilization of event history analysis, in which the probability of some behavior occurring in a specific period is modeled as a function of appropriate covariates at the start of the period. When the data about program inputs permit, it is also possible to include data about the program as covariates and thereby gain insights into program effects on the outcomes of interest. Due to lack of program data over time however, it is often difficult to align the program inputs with the actions of the women, limiting this aspect of event history analysis. Steele and Choe, <http://ccp.ucr.ac.cr>, describe and illustrate the methods employed and discuss some of these constraints.

Another form of multilevel analysis occurs in the study of diffusion effects. It is likely that the adoption of contraceptive use is influenced by one's reference groups and the nature and level of interpersonal communication in addition to personal characteristics, and the concept of diffusion has gained prominence in theories of the demographic transition. Diffusion may be viewed as providing an additional indirect effect of program accessibility. As Montgomery and Casterline (1993, p459) note, "the analysis of diffusion effects can be regarded as a special case of multi-level analysis, in which the structure of norms and the information channels embedded in reference groups set the parameters for decisions at the individual level." As they go on to illustrate, autoregressive models can capture this process of endogenous feedback in which changes in contraceptive use or fertility stimulate further change.

The large number of studies and experiments attempting to measure program effects over the years suggests the possibility of employing formal methods of meta-analysis to assess the weight of evidence. This technique, in which studies themselves become the unit of analysis, is widely used in medicine and epidemiology, and its potential for assessing family planning program effects is developed by Bauman and Suchindran, <http://ccp.ucr.ac.cr>. It is worth noting that less formal methods for assessing the weight of evidence have long been in use for gauging the impact of family planning programs, particularly within specific countries. In these studies, knowledgeable observers often muster a range of evidence about developments to judge whether a case can be made for net impact of a program in affecting the level or rate of change in fertility-related behaviors. Country-specific as well as broader assessments are given by Freedman and Berelson, 1976; Sherris et al., 1985; Knodel et al., 1987; Cleland et al., 1992; Freedman and Freedman, 1992; and Ahlburg and Diamond, 1996. Often the evidence combines survey data on attitudes and use, with demographic trends, and smaller qualitative studies that ascertain in some

depth past attitudes and forces contributing to change. (See also Bertrand, <http://ccp.ucr.ac.cr>, on qualitative approaches.)

## CONCLUSION

This overview of the main techniques employed to assess the impact of family planning programs demonstrates that evaluation of programs has never been a static exercise. The methods employed in each period have changed in response to the way programs themselves evolved; in response to critiques of what programs can achieve; and from accumulated knowledge and insight into the strengths and weaknesses of each technique. The growth of evaluation technology has not been smooth and continuous. There have been periods in which excellent opportunities for more definitive assessments have been missed, as in the case of the large number of quasi-experimental studies carried out in lieu of more adequate true experimental designs, and there have been periods in which evaluation has been hampered by lack of adequate data as in the period of the World Fertility Surveys, where little explicit attention was given to the question of generating information for measuring program impact.

Although the past record is mixed, there have been solid accomplishments and the long-term attention to the question and the contributions of analysts from different disciplines provide excellent opportunities for more definitive work in the future. Many of the elements are in place. These include sharper insights into the multilevel strategy that will circumvent the problem of possible bias arising from the way program inputs are allocated, and clearer knowledge of what should be measured at the community and facility level to generate the history and detail needed to execute strong conceptual frameworks. At the same time the evolving history of data collection at the household and facility level point to possible designs that will generate the needed information through retention of the sampling clusters in successive surveys (or the use of panel designs) and for techniques that meld special studies of facilities with population based surveys so that the service environment of the households interviewed is known in sufficient detail.

These advances will be challenged by new developments in program structure and goals. In recent years and particularly as a consequence of the International Conference on Population and Development in 1994, family planning programs are changing in several important ways. There has been less emphasis on demographic targets per se and more attention to enabling each woman to achieve her reproductive goals, greater concern with the quality of services provided, gender equity, and a mandate to move beyond provision of contraceptive services to broader concerns with reproductive health in general which incorporates such elements as sexually transmitted diseases and HIV/AIDS, safe pregnancy, breast-feeding, and nutrition (See McIntosh and Finkle, 1995; Jain, 1995). These developments have widespread implications for evaluation in terms of objectives, measurement, and techniques for estimating effects. As example, insofar as the objective of furthering individual reproductive goals is adhered to, it will be necessary to develop measures and techniques that speak to the proportion and number of successful couples rather than measures that reflect net or average outcomes. In many areas of reproductive health measures for monitoring program progress and as inputs into impact assessment still need to be developed and refined. (See Bertrand et al., 1996; Tsui et al., 1997.) Though the precise contours of a new evaluation strategy for dealing with these changing program objectives are still to be worked out, it is probable that they will incorporate many of the features that have emerged in estimating program effects on fertility. In particular, it would appear that in addition to the advances in multilevel modeling and measurement treated above, the coming period will see attention to the potential of longitudinal surveys and the use of surveillance areas to manage the widened scope of interest, more use of qualitative studies to assess quality of programs, and a

renewed interest in experimental designs as a mechanism for varying inputs of different kinds and carefully assessing both the costs involved and the effects of the interventions. As programs can rarely introduce new services and products throughout the country at one time, they should consider adopting an experimental strategy in dissemination so as gain the benefit of more precise knowledge of effects as the program matures. The broader set of goals will also put increased pressure on adequate mechanisms for utilization of research findings for policy and program development (See Khuda, <http://ccp.ucr.ac.cr>.)

Careful attention to the past history of family planning program evaluation can prove fruitful in the years ahead as programs undergo changes in structure and goals and new demands are made on the techniques of evaluation to assess the impact of program activities.

## REFERENCES

- Ahlburg, D. and Diamond, I. (1996), 'Evaluating the Impact of Family Planning Programmes' in Ahlburg et al.(eds.), *The Impact of Population Growth on Well-Being in Developing Countries*. New York: Springer.
- Bauman, K., Viadro, C., and Tsui, A. (1994), 'Use Of True Experimental Designs For Family Planning Program Evaluation: Merits, Problems And Solutions.' *International Family Planning Perspective*, 20(3):108-113.
- Bauman, K. (1995), 'How Effective Are Family Planning Programs? Evidence From True Experiments.' Paper presented at the Annual Meeting of the Population Association of America, San Francisco, CA, April 5-7.
- Bauman, K. and C. Suchindran. 'Meta-analysis: an Overview with Application to Evaluations of Family Planning Programmes.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Berelson, B. (1974), "World Population: Status Report, 1974. A Guide for the Concerned Citizen." *Reports on Population/Family Planning*, No, 15, January 1974.
- Berelson, B., Anderson, R., Harkavy, O., Maier, J., Mauldin, P. and Segal, S. (1965), *Family Planning and Population Programs: A Review of World Developments*. Chicago: The University of Chicago Press.
- Bertrand, J., Magnani, R. and Rutenberg, N. (1996), *Evaluating Family Planning Programs, with Adaptations for Reproductive Health*. Chapel Hill, NC: The EVALUATION Project.
- Bertrand, J. and Lisanne F. Brown, (1997) 'The Working Group on the Evaluation of Family Planning Operations Research Final Report'. Chapel Hill, NC: Evaluation Project, Carolina Population Center, University of North Carolina.
- Bertrand, J. 'Qualitative Methods for Evaluating Family Planning Programs.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Bongaarts, J. (1986), 'The Prevalence Method.' In United Nations, *The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*. Manual IX. Department of International Economics and Social Affairs, ST/SER.A/66/Add.1 New York: United Nations.



- Bongaarts, J. (1994) 'The Impact Of Population Policies: Comment.' *Population and Development Review* 20(3): 616-20.
- Bongaarts, J. 'Role of Family Planning Programmes in Contemporary Fertility Transitions.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Brennan, E. 'The United Nations Population Inquiries.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Buckner, B., Tsui, A., Hermalin, A. and McKaig, C., eds. (1995), *A Guide to Methods of Family Planning Program Evaluation, 1965-1990*. Chapel Hill, NC: The Evaluation Project
- Bulatao, R. and Lee, R., eds. (1983), *Determinants of Fertility in Developing Countries*. (Two volumes.) New York: Academic Press.
- Caldwell, J. and Caldwell, P. (1986), *Limiting Population Growth and the Ford Foundation Contribution*. Dover, NH: Frances Pinter Ltd.
- Casterline, J., ed. (1985a), *The Collection and Analysis of Community Data*. Voorburg, Netherlands: Statistical Research Institute.
- Casterline, J. (1985b), 'Community Effects on Fertility,' in J. Casterline (ed.), *The Collection and Analysis of Community Data*. Voorburg, Netherlands: Statistical Research Institute, 65-75.
- Casterline, J. (1987), 'The Collection and Analysis of Community Data,' in J. Cleland and C. Scott (eds.), *The World Fertility Survey: An Assessment*. Oxford: Oxford University Press.
- Chamrathirong, A., Hirschman, C. and Guest, P. (1992) 'A Multilevel Analysis of the Determinants of Fertility in the Four Regions of Thailand.' *Asia Pacific Population Journal*, 7(1): 51-64.
- Chandrasekaran, C., and Hermalin, A. (1975), *Measuring the Effect of Family Planning Programs on Fertility*. Liege, Belgium: Ordina Editions.
- Cleland, J. and Hobcraft, J. (1985), *Reproductive Change in Developing Countries: Insights from the World Fertility Survey*. Oxford: Oxford University Press.
- Cleland, J., Phillips, J., Amin, S., and Kamal, G. (1994), *Determinants of Reproductive Change in Bangladesh: Success in a Challenging Environment*. Washington, DC: The World Bank.
- Coale, A. and Watkins, S., eds. (1986), *The Decline in Fertility in Europe*. Princeton, NJ: Princeton University Press.
- Cuca, R., and Pierce, C. (1977), *Experiments in Family Planning: Lessons from the Developing World*. Baltimore, MD: The Johns Hopkins University Press.

- Curtis, S. and Diamond, I. (1995), 'When Fertility Seems Too High for Contraceptive Prevalence: An Analysis of Northeast Brazil.' *International Family Planning Perspectives* 21(2): 58-62.
- Curtis, S. 'Using Calendar Data to Study Contraceptive Use Dynamics.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Cutright, P. (1983), 'The Ingredients Of Recent Fertility Decline In Developing Countries.' *International Family Planning Perspectives*, 9(4): 101-109.
- Davis, K. (1967), 'Population Policy: Will Current Programmes Succeed?' *Science*, 158: 730-739.
- Demeny, P. (1979) 'On The End Of The Population Explosion.' *Population and Development Review*, 5(1): 141-62.
- Donaldson, P. and Tsui, A. (1990), 'The International Family Planning Movement.' *Population Bulletin*, 45:3.
- Easterlin, R. (1983), 'Modernization and Fertility: A Critical Essay.' In Bulatao, R. and Lee, R. (eds.) *Determinants of Fertility in Developing Countries, Volume 2: Fertility Regulation and Institutional Influences*. New York: Academic Press.
- Easterlin, R. and Crimmins, E. (1985), *The Fertility Revolution: A Supply-Demand Analysis*. Chicago: The University of Chicago Press.
- Entwisle, B., Hermalin, A., Kamnuansilpa, P., and Chamratrithirong, A. (1984), 'A Multilevel Model of Family Planning Availability and Contraceptive Use in Rural Thailand.' *Demography* 21, No. 4: 559-573.
- Entwisle, B. and W.M. Mason (1985) 'Multilevel Effects Of Socioeconomic Development And Family Planning Programs On Children Ever (sic) Born.' *American Journal of Sociology* 91(3):616-49.
- EVALUATION Project (1996), 'Monitoring Family Planning Programs.' Wallchart. Chapel Hill, NC: The EVALUATION Project and the Population Reference Bureau.
- Fisher, A., J. E. Laing, and John E. Stoeckel (1991), *Handbook for Family Planning Operations Research Design. 2<sup>nd</sup> edition*. New York: The Population Council.
- Fisher, A., R. Miller, and K. Miller 'Situation Analysis: Assessing the Functioning and Quality of the Service Delivery Environment.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Foreit, J., Gorosh, M., Gillespie, D., and Merrit, C.G. (1978), 'Community-Based and Commercial Contraceptive Distribution: An Inventory and Appraisal.' *Population Reports: Family Planning Programs*. Series J, No. 19, March 1978.
- Foreit, J.R., J.E., Rosen, M. Ramos, E. Mostajo, and R. Monge. (1990), 'The Impact of Service Delivery Frequency on Family Planning Program Output and Efficiency.' *Studies in Family Planning* 21 (4): 209-215.

- Freedman, R. and Takeshita, J. (1969), *Family Planning in Taiwan: An Experiment in Social Change*. Princeton, NJ: Princeton University Press.
- Freedman, R., and Berelson, B. (1976), 'The Record Of Family Planning Programs.' *Studies in Family Planning* 7(1):1-40.
- Freedman, R. and Freedman, D. (1992), 'The Role Of Family Planning Programmes As A Fertility Determinant.' In J.F. Phillips and J. A. Ross, eds., *Family Planning Programmes and Fertility*. Oxford: Oxford University Press.
- Freedman, R. (1997), 'Do Family Planning Programs Affect Fertility Preferences? A Literature Review' *Studies in Family Planning*, 28/1: 1-13.
- Gallen, M., and Rinehart, W., eds. (1986), 'Operations Research: Lessons for Policy and Programs.' *Population Reports: Family Planning Programs*, 14(2).
- Gertler, P.J. and Molyneaux, J.W. (1994), 'How Economic Development And Family Planning Programs Combined To Reduce Indonesian Fertility.' *Demography* 31(1): 33-63
- Gorosh, M. and Wolfers, D. (1979) 'Standard Couple-years of Protection,' in *The Methodology of Measuring the Impact of Family Planning Programs on Fertility*, pp. 48-62. New York: United Nations.
- Guilkey, D. and I. Diamond, 'Structural Models for Programs Evaluation.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Harkavy, O. (1995), *Curbing Population Growth: An Insider's Perspectives on the Population Movement*. New York: Plenum Press.
- Hauser, P. (1967), 'Family Planning and Population Programs: A Book Review Article.' *Demography* 4 (1967): 397-414.
- Hermalin, A.I. (1975), 'Regression Analysis Of Areal Data.' In C. Chandrasekaran and A.I. Hermalin (eds.), *Measuring the Effect of Family Planning Programs on Fertility*. Liege, Belgium: Ordina Editions.
- Hermalin, A.I. (1979), 'Multivariate Areal Analysis.' In *Manual IX: The Methodology of Measuring the Impact of Family Planning Programs on Fertility* (Sales No. E.78.XIII.8). New York: United Nations Department of International Economic and Social Affairs.
- Hermalin, A.I. (1982a), 'Family Planning Programs: Effects On Fertility.' In J.A. Ross (ed.), *International Encyclopedia of Population*. New York: Free Press.
- Hermalin, A.I. (1982b), 'Some Cautions In The Use And Interpretations Of Regression Analysis For The Evaluation Of Family Planning Programs.' In *Evaluation of the Impact of Family Planning Programs on Fertility: Sources of Variance* (Sales No. E.84.XIII.9) New York: United Nations Department of International Economic and Social Affairs.

- Hermalin, A.I. (1986), 'The Multi-Level Approach: Theory And Concepts.' In *Manual IX: The Methodology of Measuring the Impact of Family Planning Programs on Fertility, Addendum* (Sales No. E.86.XIII.4) New York: United Nations Department of International Economic and Social Affairs.
- Hermalin, A. and Entwisle, B. (1982), *The Role of Surveys in the Analysis of Family Planning Programs*. (Proceedings of a seminar held in Bogota, Colombia, October 1980.) Liege, Belgium: Ordina Editions.
- Hermalin, A. and Khadr, Z. (1996), 'The Impact of Family Planning Programs on Fertility: A Selective Assessment of the Evidence.' Paper presented at the conference, 'Assessing the Past for the Future: Family Planning Policy, Programs, and Resources,' June 1996, Bellagio, Italy.
- Hernandez, D.J. (1984), *Success or Failure? Family Planning Programs in the Third World*. Westport, Conn.: Greenwood Press.
- Himes, N. (1963), *Medical History of Contraception*. New York: Gamut Press, Inc.
- Jain, A., (1995), 'Implementing the ICPD's Message,' *Studies in Family Planning*, 26/5, 296-298.
- Khuda, B. 'Assessing the Policy Impact of Operations Research on the Bangladesh Health and Family Planning Programme.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Kiser, C. (ed.) (1962), *Research in Family Planning*. Princeton, NJ: Princeton University Press.
- Knodel, J., Chamratrithirong, A., and Debavalya, N. (1987), *Thailand's Reproductive Revolution: Rapid Fertility Decline in a Third-World Setting*. Madison: University of Wisconsin Press.
- Knowles, J., Akin, J. and Guilkey, D. (1994), The Impact Of Population Policies: Comment. *Population and Development Review*, 20(3):611-15.  
Knowles, J. 'Cost-effectiveness and Cost-benefit Analysis.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Lapham, R.J. and Mauldin, W.P (1987) The Effects Of Family Planning On Fertility: Research Findings. In Lapham R.J. and G.B. Simmons (eds.), *Organizing for Effective Family Planning Programs*. Washington, D.C.: National Academy Press.
- Lloyd, C. and Ross, J. (1989), *Methods for Measuring the Fertility Impact of Family Planning Programs: The Experience of the Last Decade*. Population Council Research Division Working Paper No.7.
- Mason, W. (1986), 'The Multi-Level Approach: Illustrative Example.' In United Nations, *The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*. Manual IX. Department of International Economics and Social Affairs, ST/SER.A/66/Add.1 New York: United Nations.

- Mauldin, W.P. (1979), 'Fertility Projection/Trend Analysis.' In United Nations, *The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*. Manual IX. Department of International Economics and Social Affairs, ST/SER.A/66/Add.1 New York: United Nations.
- Mauldin, W.P. (1983), 'Population programs and fertility regulation.' In R.A. Bulatao and R.D. Lee, eds., *Determinants of Fertility in Developing Countries*. New York: Academic Press.
- Mauldin, W.P. and Berelson, B. (1978), 'Conditions Of Fertility Decline In Developing Countries, 1965-75.' *Studies in Family Planning*, 89-148.
- Mauldin, W.P and Johnson-Acsadi, G. (1975), 'Introduction,' in C. Chandrasekaran and A.I. Hermalin (eds.), *Measuring the Effect of Family Planning Programs on Fertility*. Liege, Belgium: Ordina Editions, 1-15.
- Mauldin, W.P. and Lapham, R. (1987), 'The Measurement of Family Planning Inputs.' In Lapham and Simmons (eds.) *Organizing for Effective Family Planning Programs*. Washington D.C.: National Academy Press.
- Mauldin, W.P. and Ross, J. (1991), 'Family Planning Programs: Efforts and Results, 1982-89.' *Studies in Family Planning* 22(6):350-367.
- McIntosh, C. A. and J. L. Finkle (1995), 'The Cairo Conference on Population and Development: A New Paradigm?' *Population and Development Review*, 21/2: 223-260.
- Menken, J. (1975), 'Simulation Studies.' In C. Chandrasekaran and A.I. Hermalin (eds.), *Measuring the Effect of Family Planning Programs on Fertility*. Liege, Belgium: Ordina Editions, 97.
- Mensch, B., Arends, M., Jain, A., and Garate, M. (1995), 'Meeting Reproductive Goals: The Impact of the Quality of Family Planning Services on Unintended Pregnancy in Peru.' Research Division, Working Paper No, 81. New York: The Population Council.
- Montgomery, M. (1987), 'A New Look at the Easterlin 'Synthesis' Framework.' *Demography*, 24:4.
- Montgomery, M. and Casterline, J. (1993), 'The Diffusion of Fertility Control in Taiwan: Evidence from Pooled Cross-Section Time-Series Models.' *Population Studies*, 47; 457-479.
- Morris, L. and Anderson, J. (1982), 'The Use of Contraceptive Prevalence Survey Data to Evaluate Family Planning Program Service Statistics.' In A. Hermalin and B. Entwisle (eds.), *The Role of Surveys in the Analysis of Family Planning Programs*. (Proceedings of a seminar held in Bogota, Colombia, October 1980.) Liege, Belgium: Ordina Editions.
- Ness, G.D. and Ando, H. (1984), *The Land is Shrinking: Population Planning in Asia*. Baltimore, MD: The Johns Hopkins University Press.
- Nortman, D. (1970), 'Population and Family Planning Programs: A Factbook.' *Reports on Population/Family Planning*, July 1970.

- Nortman, D. (1979), 'Component Projection Approach I: A Computerized Model.' In *The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*, pp. 48-62. New York: United Nations, 1979.
- Nortman, D. (1985), 'Population and Family Planning Programs: A Compendium of Data through 1983.' *A Population Council Factbook*, 1985.
- Nortman, D. and Hofstatter, E. (1973), 'Population and Family Planning Programs: A Factbook.' *Reports on Population/Family Planning*, No. 2, September 1973.
- Nortman, D. and Hofstatter, E. (1980), 'Population and Family Planning Programs: A Compendium of Data through 1978.' *A Population Council Factbook*, 1980.
- Peterson, W. (1975), *Population*. (Third Edition) New York: Macmillan.
- Phillips, J., Simmons, R., Koenig, M., and Chakraborty, J. (1988), 'Determinants of Reproductive Change in a Traditional Society: Evidence from Matlab, Bangladesh.' *Studies in Family Planning*, 19(6): 313-334.
- Potter, R. (1979), 'Analysis of Reproductive Process.' In United Nations *Manual IX: The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*, pp. 76-96. New York: United Nations.
- Potter, R. (1981), 'The Analysis of Cross-Method Variance in Assessing Family Planning Programme Effects on Fertility.' In International Union for the Scientific Study of Population (IUSSP) International Population Conference, Manila. Volume 1. Liege, Belgium: IUSSP.
- Pritchett, L. (1994), 'Desired Fertility and the Impact of Population Policies.' *Population and Development Review*, 20(1):1-56.
- Rosero-Bixby, B. Entwistle, R. Rindfuss, S.J. Walsh 'Geographic Information Systems (GIS) in Evaluation of Family Planning Programmes: Examples for Costa Rica and Thailand.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Ross, J., J. Donayre, and R. McNamara, (1987), 'Perspectives on Operation Research.' *International Family Planning Perspectives* 13/4, 128-135.
- Ross, J. and Forrest, J.D. (1978), 'The Demographic Assessment of Family Planning Programs: A Bibliographic Essay.' *Population Index*, 44(1).
- Ross, J. and Lloyd, C. (1992), 'Methods for Measuring the Fertility Impact of Family Planning Programs: The Experience of the Last Decade.' In J. Phillips and J. Ross (eds.), *Family Planning Programs and Fertility*. Oxford: Clarendon Press.
- Ross, J. and Mauldin, W.P. (1996), 'Family Planning Programs: Efforts and Results, 1972-94.' *Studies in Family Planning*, Vol. 27, No. 3: 137-147.
- Ross, J., Mauldin, W.P., and Miller, V. (1993), *Family Planning and Population: A Compendium of International Statistics*. New York: The Population Council.

- Rozenweig, M. and Wolpin, K. (1986), 'Evaluating The Effects Of Optimally Distributed Public Programs: Child Health And Family Planning Interventions.' *American Economic Review*, 76(3):470-82.
- Schultz, P. (1992), 'Assessing Family Planning Cost Effectiveness Applicability Of Individual Demand-Programme Supply Framework.' In J. Phillips and J. Ross (eds.) *Family Planning Programmes and Fertility*. Oxford: Clarendon Press.
- Schultz, P. (1997), 'Sources of Fertility Decline in Modern Economic Growth: Is Aggregate Evidence on the Demographic Transition Credible?' In M. Rosenzweig and O. Start (eds.) *Handbook of Population and Family Economics*. Amsterdam: Elsevier.
- Seltzer, J.R. (2002) *The Origins and Evolution of Family Planning Programs in Developing Countries*. Santa Monica: RAND.
- Sherris, J., ed. (1985), 'The Impact of Family Planning Programs on Fertility.' *Population Reports: Family Planning Programs*, XIII(1).
- Simmons, R. and Simmons, G. (1987), 'The Task Environment of Family Planning.' In R. Lapham and G. Simmons (eds.), *Organizing for Effective Family Planning Programs*. Washington, DC: National Academy Press.
- Steele, F. and M. Choe 'Event History Analysis for Family Planning Programme Evaluation.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Suchman, E. (1967), *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. New York: Russell Sage Foundation.
- Tolnay, S.E. and Christenson, R.L. (1984), 'The Effects Of Social Setting And Family Planning Programs On Recent Fertility Declines In Developing Countries: A Reassessment.' *Sociology and Social Research*, 69(1): 72-89.
- Tsui, A. (1985), 'Community Effects on Contraceptive Use,' in J. Casterline (ed.), *The Collection and Analysis of Community Data*. Voorburg, Netherlands: Statistical Research Institute, 77-99.
- Tsui, A. (1996), *Family Planning Programs in Asia: Approaching a Half-Century of Effort*. Asia-Pacific Population Research Reports, No.8; April 1996.
- Tsui, A.O. and Bogue, D.J. (1978), 'Declining World Fertility: Trends, Causes, Implications.' *Population Bulletin*, 33(4): 1-55.
- Tsui, A.O. and Hermalin, A.I. 'Improving the Effectiveness of Family Planning Programs: Measurement Needs and Issues.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Tsui, A.O., Wasserheit, J. N. and Haaga, J. G., eds. *Reproductive Health in Developing Countries*, Washington, DC: National Academy Press.

- United Nations (1973), *The Determinants and Consequences of Population Trends: New Summary of Findings on Interaction of Demographic, Economic and Social Factors*. Vol 1. New York: United Nations.
- United Nations (1978), *Methods of Measuring the Impact of Family Planning Programmes on Fertility: Problems and Issues*. (Sales No. E.78.XIII.2) New York: United Nations.
- United Nations (1979), *Manual IX: The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*. (Sales No. E.78.XIII.8) ST/ESA/Ser. A/66. New York: United Nations.
- United Nations (1982) *Evaluation of the Impact of Family Planning Programs on Fertility: Sources of Variance* (Sales No. E.81.XIII.9) New York: United Nations.
- United Nations (1985), *Studies to Enhance the Evaluation of Family Planning Program*. (Sales No. E.84.XIII.9) New York: United Nations.
- United Nations (1986), *Manual IX: The Methodology of Measuring the Impact of Family Planning Programs on Fertility, Addendum*. (Sales No. E.86.XIII.4) New York: United Nations.
- Vaessen, M. 'Demographic and Health Surveys.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Wawer, M.J., R. McNamara, T. McGinn and D. Lauro, (1991), 'Family Planning Operations Research in Africa: Reviewing a Decade of Experience.' *Studies in Family Planning*. 22/5: 279-293.
- Westoff, C. 'Unmet Need Revisited.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)
- Wilkinson, M. 'Using DHS Service Availability Data for Evaluation.' (<http://ccp.ucr.ac.cr/noticias/plani/iusspi.htm>)