

Chapter 7

Impact Evaluation

CHOICE OF EVALUATION METHODOLOGY

- Lesson 1. Various research designs are suitable for evaluating the impact of health communication programs.
- Lesson 2. The randomized control group design has limited value for communication program evaluation because it is usually not feasible and because by itself it does not provide information about the underlying processes of change.
- Lesson 3. Because most communication programs now use mass media to try to reach the entire population, the most rigorous design available is the one-group, before-after longitudinal design with multivariate statistical controls.
- Lesson 4. An interrupted time-series analysis of service statistics provides additional evidence of communication impact, especially when one of the objectives of the program is to promote service providers.
- Lesson 5. Evaluations can benefit from asking audience members for their own opinions about the impact of communication.

THEORY-DRIVEN EVALUATION

- Lesson 6. Evaluations that include subobjectives based on theoretical models such as the Steps to Behavior Change provide evidence of impact on behavior and useful feedback to improve future programs.
- Lesson 7. A thorough evaluation of the impact of communication includes an analysis of its indirect effects as well as its direct effects on health behavior.
- Lesson 8. Exposure to communication and the steps to behavior change have cumulative effects on behavior much like cumulative risk and dose response in epidemiologic research.
- Lesson 9. Aggregating survey data by local units of analysis such as villages or sample clusters reveals the impact of communication on local culture in addition to its impact on individuals.
- Lesson 10. Path analysis can depict and test the direction and strength of all the relationships measured in a communication program evaluation.

EVALUATING CLIENT-PROVIDER COMMUNICATION

- Lesson 11. Because of the interactive nature of interpersonal communication and counseling, innovative research methods are needed for effective evaluation.

COLLABORATION AND DISSEMINATION OF FINDINGS

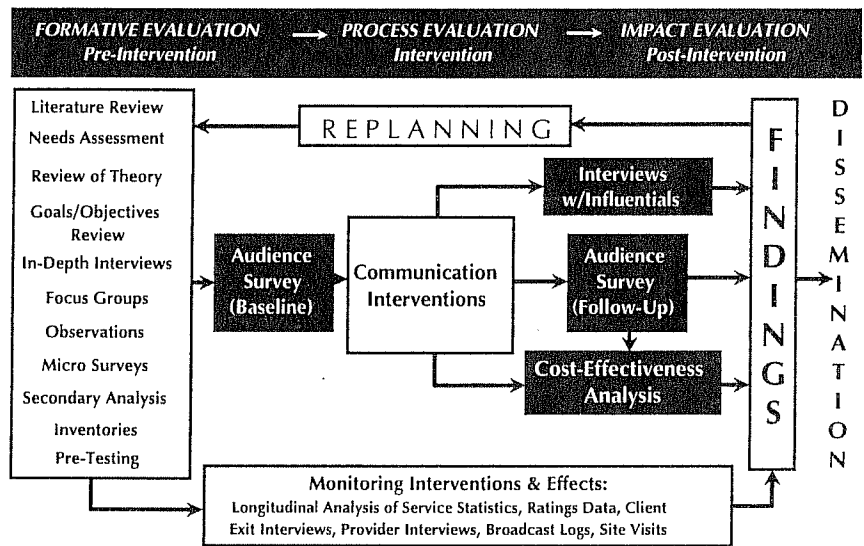
- Lesson 12. Collaborating with other organizations for program evaluation can provide substantial benefits.
- Lesson 13. Evaluation results can be presented at various levels of complexity to suit different audiences.

Evaluation is the systematic application of scientific procedures to assess the conceptualization, design, implementation, impact, and cost-effectiveness of social interventions (Bertrand & Kincaid, 1996). The purpose of evaluation is to measure the process and the impact of a program against the objectives established in the strategic design in order to contribute to decision-making. *Measurement* requires methods for collecting information and assessing the changes specified in the

program objectives. *Process* refers to what takes place as the program is implemented. *Impact* refers to program outcomes. *Decision-making* means applying the findings of an evaluation to improve ongoing or future programs.

Evaluation is easy to ignore until a project nears the end but impossible to carry out effectively unless it is planned from the start. Parallel with the P Process itself, evaluation in fact begins along with the preliminary analysis that provides background information and baseline data. It takes further shape in the strategic design phase, where specific, measurable (SMART) objectives and supporting subobjectives are established. Evaluation issues also arise in the pretest, development, and implementation stages, where evaluators determine whether the process described in the design document is actually taking place and whether the outputs are understood by the audience. Eventually, at the close of the program or at some milestone in an ongoing program, evaluation assesses impact on the intended audience or the entire population, measuring and analyzing changes specified by the original objective. This chapter focuses on that impact evaluation. Ultimately, evaluation findings, disseminated both within a program and externally, provide the basis for decision-making to correct deficiencies, build on program successes, and plan for continuity (see Chapter 8). The entire evaluation process is illustrated in the Communication Design and Evaluation System (CODES) diagram (see Figure 7.1).

Figure 7.1
Communication Design and Evaluation System (CODES)



Note: CODES has evolved out of the Media Impact Research System (MIRS) originally developed by D. Lawrence Kincaid for JHU/CCP/PCS in 1990.

Source: JHU/CCP/PCS, 1995.

Evaluating any social program is difficult. To make the most convincing case that the program was responsible for the observed outcome, eight criteria are important (Bertrand & Kincaid, 1996; Hill, 1971; Mohr, 1992; Schlesselman, 1982). It is difficult to satisfy all eight criteria in a single study. Fortunately, they are not all required in order to make a valid conclusion about impact. The more that apply, however, the greater the confidence in the conclusion.

- Measurement of a change or difference in the population of interest (variation in outcome).
- Correlation between exposure to the program and the intended outcome (covariation).
- Evidence that exposure to the program occurred *before* the observed change in the outcome (time-order).
- Control of the effects of confounding variables that might also influence the outcome (spuriousness).
- Observation of an abrupt, large impact in the absence of other major influences (magnitude).
- A direct and close causal connection specified by theory between a program and its outcome (proximity).
- Impact that increases in proportion to the level or duration of communication exposure (dose response).
- Consistency with the evidence from previous communication program evaluations (replication with variation.)

Communication programs pose special problems for evaluators. It is a daunting task to demonstrate that communication programs themselves make a difference to health behavior—both apart from other program functions, such as availability of services, adequate supplies, and trained providers, and apart from secular trends that may influence health behavior. First, communication is not only an independent function but also a part of other program functions from counseling to training. Second, most communication has both short-term and long-term effects. Many short-term campaigns in the United States—for example, promoting smoking cessation or use of automobile seat belts—have been evaluated as failures (McGuire, 1986), and yet over the last decade smoking has declined and seat belt use has grown. In fact, both practices are now social norms. People often consider information for a long time before they take action, making impact evaluations over a short time span invalid predictors of the long term. Third, communication may have both intended and unintended effects. Since program communication never takes place in a vacuum but always in a context, and often in competition with other messages, communication evaluation may need to identify unintended effects, not just those anticipated in program objectives, and also to track competing messages or activities in order to assess the net impact of the program.

The experience of Population Communication Services since 1982 suggests that the eight basic criteria can be most easily met and the special problems of communication evaluation resolved when the programs under evaluation are intensive, multimedia programs that take place over relatively short time periods, albeit repeatedly. The reason is that, when communication suddenly increases from

a very low to a very high level and then is suddenly eliminated, while very few, if any, changes occur in the other family planning program components, it is easier to attribute the outcome to the communication component. The converse is also true: When a communication program consists of a very diffuse set of activities conducted over a long period of time (more than one year), its impact is difficult to separate from that of other programs or social influences. Therefore long-term multimedia programs are presumably likely to have broad, cumulative impact that will require an extra effort to detect. The impact of intensive short-term campaigns will be easier to detect, but similar campaigns will have to be repeated periodically to have lasting effects.

Population Communication Services tried initially to evaluate specific print materials or selected community meetings and to differentiate between the impacts of one product or channel and another, but these efforts have gradually been abandoned in favor of comprehensive evaluation of major communication programs that will be repeated over several years. Although some pilot projects on interpersonal communication within the family or training in client counseling are still being evaluated separately, major evaluation resources now go to large-scale, continuing programs. Within these larger evaluations, it is often possible to compare the differential impacts of specific activities in a fairly simple and inexpensive way.

A useful rule of thumb is that an evaluation should cost about 10 percent of the total cost of the program being evaluated. There is little to be gained by spending any more than this on evaluation, since the resources would be better spent to expand the program and increase its (measurable) impact. This rule has an implication for small projects: If the project is very small—for example, US\$50,000—evaluation is not worthwhile because 10 percent (\$5,000) is not enough to pay for an adequate evaluation; at the same time, the program probably is not strong enough to have a measurable impact. Important exceptions to the 10 percent rule are new programs or pilot projects for which very little previous research has been done. In such cases a disproportionately large expenditure on research may be warranted, since the results may have long-term implications. Evaluation of client-provider interaction and counseling is an example of this (see Lesson 11).

Thus the lessons learned from 15 years of PCS program evaluation come primarily from major evaluations of several large country programs. The Bangladesh experience has been especially instructive because it covers a span of almost 10 years and includes innovations in communication linked directly to service delivery. It is less directly linked to mass media than other projects. Evaluations in Kenya, the Philippines, Tanzania, and Uganda show primarily how multimedia campaigns can be evaluated. In each case, the purpose of the evaluation was not only to determine whether the program achieved its preestablished objectives but also to learn what worked and what did not work within the program so that future activities could be improved.

Lessons about evaluation can be grouped under several summary lessons:

- **Choice of evaluation methodology.** No one evaluation design is ideal for all programs. The choice of design depends on the nature of the program, its objectives, and the data that can be collected. The most revealing evaluations combine multiple designs and measurements to address different issues and reinforce one another.
- **Theory-driven design.** Evaluation design that is theory-based and uses the same theory of behavior change that underlies the program itself can produce findings with powerful practical applications for improving future programs.
- **Collaboration and dissemination of findings.** Collaborating with other organizations saves time and money, introduces new ideas, and facilitates sharing information. The results of evaluation must be disseminated to important people at different decision-making levels, but they must be presented in ways appropriate to people's needs and knowledge at each level.

CHOICE OF EVALUATION METHODOLOGY

LESSON 1.

Various research designs are suitable for evaluating the impact of health communication programs.

Several major types of research design have been used by Population Communication Services to evaluate program impacts. They have been applied in evaluating mass media campaigns, community-based projects, counseling, and counseling training. They have been used primarily with individuals as the units for analysis but in a few cases with various geographic units or with groups. While every design has limitations, these designs, when adapted and applied in appropriate situations, can evaluate the impact of communication programs on behavior:

- **Randomized control group design**, sometimes called the "classical" experimental design, randomly assigns individuals, groups, or geographical areas either to an experimental treatment or to a control group. Random assignment helps to ensure that the two groups are equivalent. The differences that appear over time between the treatment group and the control group are then the measure of program impact. This design can be applied most readily in evaluating the training and performance of service providers, where a large number of providers, clinics, or clients can be assigned randomly to a treatment or control group.
- **Nonequivalent control group design** is similar to classic experimental design except that, instead of random assignment, the groups are matched as closely as possible on characteristics thought to affect the outcome.
- **One-group, before-after design with sample surveys** of the population of interest is appropriate to evaluate a communication program implemented for an entire population. A representative sample of that population is surveyed before and after the program. Differences between "before" and "after" findings measure how much change

occurred during the interval and whether people who were exposed to the program changed more than those who were not exposed. The survey data can be used to control statistically for the influence of confounding variables. This is the design most often used by PCS for mass media or large-scale, full-coverage interventions. For example, it was used to evaluate both of the Zimbabwe male motivation campaigns (Piotrow, Kincaid, et al., 1992; Y. M. Kim, Marangwanda & Kols, 1996) and the Bolivian Reproductive Health Campaign (Valente, Saba, et al., 1996), as well as Turkish, Philippine, Kenyan, and Ugandan campaigns and many others.

One-group, before-after designs can use one of at least three different types of surveys: (1) independent surveys, using a different random sample of respondents for each round of data collection, which is the most common approach; (2) longitudinal (panel) surveys, using the same random sample of respondents repeatedly for each round; and (3) another type of longitudinal survey using the same random sample of clusters, but interviewing a different random sample of individuals in each cluster at each round. Longitudinal surveys can show which came first—interest in and use of contraceptives or exposure to the communication program being evaluated. Assessment of the time-order relationship can either strengthen or undermine the inference that exposure to the program caused the intended change. Panel surveys have been used in PCS evaluations in the Philippines, Bangladesh, and Peru. A panel survey was also carried out by Westoff and Bankole to evaluate the impact of mass media family planning efforts in Nigeria (Westoff & Bankole, 1996).

- **Interrupted time-series design** measures behavior at many intervals over an extended period of time, usually with sales or service statistics. This design is useful in evaluating health communication programs, such as the mass media vasectomy campaigns in Brazil (Kincaid, Payne Merritt, et al., 1996), for two reasons: first, communication activities, unlike services, can be started and stopped at distinct intervals that can be compared with the time-series data; second, family planning and most health services already maintain service statistics. When unusually high demand for services coincides with the communication activities, as was the case in Brazil, the evidence for a causal link between the two is strengthened (see Lesson 4, below). Service statistics can, of course, also be used to evaluate other forms of communication such as counseling, interpersonal communication training, and community events, as was done in Nigeria (Piotrow, Rimon, et al., 1990). When providers are trained to ask clients directly what prompted them to come for services, these source-of-referral data can strengthen the conclusions drawn from the analysis of service statistics.

In theory each of these designs is distinct, but in practice evaluations sometimes can combine more than one methodology or measurement. The Bangladesh evaluation, discussed in Lessons 2 and 6 through 10, includes elements of a randomized control group design and one-group, before-after design with longitudinal sample surveys. Kenya evaluations involved one-group, before-after surveys and an interrupted time-series design using service statistics. The more that different designs and measurements can be used, the greater the confidence that the observed outcome can be attributed to the communication program.

 LESSON 2.

The randomized control group design has limited value for communication program evaluation because it is usually not feasible and because by itself it does not provide information about the underlying processes of change.

Effective evaluation accomplishes both goals of evaluation—measurement of the *process* and the *impact* of communication—in a manner that provides useful information to improve future programs. The classical experimental design can yield the most definite conclusion about the impact of a program, but unless it is supplemented by other methods, it provides no information about the underlying causal processes that lead to that impact (Campbell & Stanley, 1963; Cook & Campbell, 1979; Rossi & Freeman, 1989; Mohr, 1992; Bertrand & Kincaid, 1996). Theory plays an indispensable role in evaluation because it is used to identify the expected causal processes, to suggest ways to measure those processes, and to incorporate them into program design. This is why program evaluation is now described as either method-driven or theory-driven (Chen & Rossi, 1987; Chen, 1990). To accomplish both goals of evaluation, evaluation research needs to be driven by rigorous methods *and* by valid theories.

To understand this important lesson requires a clear understanding of the strengths and weaknesses of the classical experimental design. It was developed for use in agricultural research stations in the early part of the century, in a situation where theory was weak or nonexistent, and then adopted by psychologists, social scientists, and program evaluators for the same reasons (R. A. Fisher, 1956; Gigerenzer et al., 1989).

The three main weaknesses of the classic randomized control group design when it is applied to field experiments are not usually given enough emphasis in the textbooks.

1. The level of control implied by the logic of the design is rarely achieved in field settings.

In field settings the experimental treatment (in this case the program) may actually change over time and from place to place as it is being implemented. The longer the experiment lasts, the more likely these changes will occur. Ironically, flexibility in public health programs is both necessary and beneficial from the point of view of developing an effective program, but it negates the experimental design's assumption of a "uniform treatment" and interferes with the goal of assessing impact. This natural variation in treatment that occurs in field settings makes it difficult to say *what* the treatment was that led to the observed impact.

2. It is weak in external validity—the generalization of its results to future implementations on a full-scale basis.

A partial-coverage program tested under relatively controlled conditions will usually not be the same as the one that is later implemented full-scale. Nor should it be. The program should be improved first. Many policy-makers are aware of this lesson, and they say, "Don't give us another pilot experimental project. It takes too

long, costs too much, and, when you're done, you still don't know how it will work on a full scale." Less well understood is that, if a program is comprised of several diverse components, then, technically speaking, separate experiments or experimental treatment groups would have to be organized to test each one. This level of experimentation takes too much time and money for most social programs, such as those in public health.

3. Unless supplemented with other methods, the experimental design provides no information about the processes of change that are necessary to improve the program.

The classic experimental design provides a macro-level indication of impact without specifying what happens at the micro level inside the experimental group. To illustrate this strength and weakness, a field experiment from Bangladesh is described below. Lessons 6 through 10 describe supplemental methods that can be used to measure and test the underlying, micro-level processes in program evaluation.

Evaluation researchers must select a research design that satisfies as many of the eight criteria for a causal attribution (described above) as possible given the nature of the program. As noted, one of the key issues that the research design addresses is whether or not the outcome would have happened anyway without the program. The *strength* of the classic experimental design is that it readily satisfies this criterion. It includes a control group that does not receive the program. Each unit of interest (individual, village, school, etc.) is randomly assigned to either a control group or a treatment group. The assignment by means of a random process ensures that the two groups will be statistically equivalent before the program is implemented. If the outcome in the group that received the communication program is significantly different from the outcome in the one that did not receive it, then the difference can be attributed to the program, since all other influences should be the same for both groups.

This type of experimental design was used to evaluate the newly developed *jiggasha* approach to family planning/health promotion in Bangladesh (Kincaid, Das Gupta, et al., 1993). *Jiggasha*, a Bangladeshi term meaning "to inquire," was the name given to an innovative interpersonal communication intervention developed to improve the effectiveness of the government family welfare assistants (FWAs), who for over 15 years have been visiting village women in their homes in Bangladesh (see Box 3.2, in Chapter 3).

In the experiment 24 villages in Trishal thana¹ were rank ordered in terms of contraceptive prevalence (any method), and after a random starting point in the list every other village was assigned to the treatment group and the remaining ones were assigned to the control group. The 12 villages in the treatment group received the new *jiggasha* approach to family planning communication. Individual women or households were not assigned to experimental groups because the smallest unit in which the program could be conducted was a village.

Even when whole villages are assigned to groups, there is still a possibility of "contamination" between the experimental and control villages. Field workers or individuals from the experimental treatment villages may inadvertently tell those

living in the control villages what they are doing or what they have learned. The other villages may even try to emulate the intervention on their own. In many evaluation situations it is so difficult to find units that are appropriate or numerous enough for random assignment to groups that the comparative experimental design cannot be used.

In the *jiggasha* experiment the villages were far enough apart that contamination did not become a problem. There were no policy or ethical constraints, either. Women in the control villages were not prevented from receiving health information and services. The *jiggasha* approach was designed to *improve* the work of the existing family planning field workers, so the 12 villages assigned to the control group continued to receive their normal services during the experiment. The experimental test, then, was whether or not the new approach would have any *added* advantage over the regular government program. Until this could be demonstrated, there would be no reason to expand the approach to other villages.

Table 7.1 shows the outcome of the program in terms of increasing modern contraceptive use, the primary objective of the project. In the 12 *jiggasha* villages the mean rate of modern method use increased from 1989 to 1992 by 11.4 percentage points, from 13.5 percent to 24.9 percent, while the mean rate of modern contraceptive use in the control villages rose by only 5.3 percentage points, from 24.8 percent to 30.1 percent. The increase in the villages with the *jiggasha* approach was more than twice as great as in the control villages. The intervention succeeded in helping these villages "catch up" to the level of the control villages in terms of modern contraceptive use.

Table 7.1
Mean Percentage of Married Women Using Modern Contraceptives, by
Village Intervention Status: Trishal, Bangladesh, 1989 and 1992

Village Intervention Status	1989	1992	Change*
<i>Jiggasha villages (N=12)</i>	13.5	24.9	11.4
<i>Control villages (N=12)</i>	24.8	30.1	5.3

Note: A panel of 1,372 married women interviewed before and after the intervention (repeated measures) aggregated into 24 village units for statistical analysis.

*1992 minus 1989. The difference of 6.1 percentage points between the pre-post changes in *jiggasha* and control villages is statistically significant ($t = -2.95$ with 22 d.f., $p < 0.01$).

Source: Kincaid, Das Gupta, et al., 1993.

The substantial initial difference in modern contraceptive use between the *jiggasha* and control villages is an example of what is sometimes called "unhappy" randomization in experimental designs (Mohr, 1992). With a relatively small number of units, in this case just 24, there is greater likelihood that a random assignment will not produce equivalent groups. A close examination revealed that the median levels of contraceptive use of the two groups were very close. When the very high and very low prevalence villages were randomly assigned, however, by chance the mean level of the control group was skewed upward, and the mean level

of the *jiggasha* group downward. An examination of the initial levels of any method use (modern and nonmodern) illustrates the problem of having only 24 units. There was no statistically significant difference between the groups in terms of *any* method use, but the difference was still substantial (22.2% v. 33.3%; $F = 3.59$, $p = 0.0715$). For the *jiggasha* experiment it was not feasible to conduct the program in more than 12 villages with the available time and personnel.

The experimental design confirmed that the communication program was a success in terms of the primary objective, but this knowledge by itself offered nothing that could be used to improve the program in the future. The experimental design—which answers the success/failure question so well—by itself only provides an “outer shell” for evaluation, within which more precise measures of the processes related to the change must be measured and analyzed in greater depth (Cronbach, 1982). This lesson applies whether the project has been found to be a success or a failure. When the results of an evaluation indicate no significant impact, policy-makers want to know why and how to do it better next time. Even if the impact of an experiment can be successfully demonstrated, when it is implemented under changed conditions on a wider scale, the policy-maker still needs to know what conditions and changes would make it work better the next time (Cronbach, 1982).

The next set of lessons describes alternatives to the classic experimental design that are practical for full-coverage communication programs. Then Lessons 6 through 10 describe methods for incorporating theory and measures of causal process into communication program evaluation.

LESSON 3.

Because most communication programs now use mass media to try to reach the entire population, the most rigorous research design available is the one-group, before-after longitudinal design with multivariate statistical controls.

Today the broadcast mass media can reach almost everyone. When the program is designed to reach as many people as possible, no group can be used as a control. Instead, the only possible comparison is between those who were actually exposed to the communication and those who were not. Follow-up sample surveys of the population can determine who was actually exposed and what happened to them as a result. Because exposure to family planning communication usually ranges from 40 percent to 70 percent, a comparison group is still available in one-group research designs.

Unfortunately, two serious threats arise that make it difficult to estimate the impact of communication in a one-group design where only part of the population is exposed: confounding effects and selectivity bias. *Confounding effects* mean that behavior change may not have actually occurred because of exposure to communication but rather because of the influence of other variables that also happen to determine exposure. Exposed persons more often own or have access to radio and television, for example. They usually are of higher socioeconomic status, have more

education, live in cities where electricity is available rather than in villages, are younger, and so forth. Because these variables are often related to use of family planning, which is the outcome variable, any statistical finding that shows a relationship between exposure and family planning is confounded by these other differences. In other words, family planning practice and communication exposure may both be due to higher socioeconomic status or education.

Selectivity bias arises because those who are exposed to family planning messages may already practice family planning or may be predisposed to practice, so they are already more motivated to attend to and to recall family planning messages than those who are not exposed to the messages or who paid no attention because they were not interested in family planning. If so, prior family planning attitudes and behavior may influence exposure to communication rather than the other way around. The time-order criterion may be violated.

Both confounding effects and selectivity biases can be taken into account by the type of survey design and analysis that is used. If a randomized control group design is not feasible, then the best way to deal with confounding variables is to control or adjust for them later by means of multivariate statistical analysis, such as multiple linear regression or multiple logistic regression. An evaluation of the Uganda National Interpersonal Communication and Counseling Program illustrates this type of analysis (Kiragu, Galiwango, et al., 1996). A representative follow-up survey of 1,323 men and women ages 20–40 was conducted in the urban and peri-urban areas of four districts that were designated as the intended area for the family planning communication program. Six distinct communication media were used: a radio drama, posters, pamphlets, newspapers, a special yellow flower logo, and a yellow flower advertisement. A dose response (see Lesson 8) was found for exposure to the six communication media: The level of modern contraceptive use among men and women exposed to none or only one medium was 10.9 percent. This increased to 25.9 percent for those exposed to two to four media, and to 37.2 percent for those exposed to five or six media (Kiragu, Galiwango, et al., 1996). This relationship was statistically significant, but there was no control for the influence of confounding variables. What if both use and exposure are simply a result of living in the urban areas as opposed to the peri-urban areas or of being able to afford to own a radio and television?

A multiple logistic regression analysis was conducted to control (or adjust) for the effects of eight potential confounding variables: urban residence, district, education, age, television or car ownership, radio ownership, number of children, and family planning attitude. The results confirmed that a dose response to the six media was still statistically significant *after* controlling for the effects of the other eight variables. Women exposed to two to four media were 2.4 times more likely to use modern contraceptives than women exposed to one or none (the reference group). (See Table 7.2.) Women exposed to five or six media were 4.2 times more likely to use modern contraceptives than those in the reference group.

This type of multivariate regression analysis is the most rigorous kind of statistical analysis that can be done with survey data. It can eliminate the possibility of confounding variables, but only for variables that are measured in the survey and

Table 7.2
Multiple Logistic Regression of Modern Contraceptive Use Among
Ugandan Women on Campaign Exposure and Eight Control Variables

Variable	Odds Ratio	95% Confidence Interval
<u>Campaign Exposure</u>		
<i>0-1 materials</i>	1.00	
<i>2-4 materials</i>	2.40	1.24-4.63*
<i>5-6 materials</i>	4.24	1.95-9.22*
<u>Location</u>		
<i>Urban</i>	1.00	
<i>Peri-urban</i>	0.91	0.59-1.39
<u>District</u>		
<i>Kampala</i>	1.00	
<i>Jinja</i>	1.00	0.58-1.72
<i>Mbarara</i>	1.42	0.80-2.51
<i>Masaka</i>	1.52	0.89-2.59
<u>Education Level</u>		
<i>None/primary</i>	1.00	
<i>Secondary</i>	1.74	1.11-2.72*
<i>Beyond secondary</i>	1.65	0.91-3.03
<u>Age</u>		
<i>20-24</i>	1.00	
<i>25-29</i>	1.37	0.85-2.20
<i>30-34</i>	1.56	0.89-2.73
<i>35-40</i>	2.97	1.20-7.41*
<u>Owns TV/Car^a</u>		
<i>Owns neither</i>	1.00	
<i>Owns one of the two</i>	1.60	1.03-2.47*
<i>Owns both</i>	1.72	0.90-3.30
<u>Owns a Radio</u>		
<i>No</i>	1.00	
<i>Yes</i>	1.11	0.66-1.89
<u>Number of Children</u>		
<i>Less than 3</i>	1.00	
<i>3 or more</i>	1.74	1.09-2.78*
<u>Attitude Toward Family Planning</u>		
<i>Opposed</i>	1.00	
<i>Favorable</i>	1.90	1.22-2.95*

*p ≤ 0.05. N = 643; unadjusted data.

^aAn indicator of socioeconomic status.

Source: Johns Hopkins University/Center for Communication Programs and Family Planning Association of Uganda, Uganda Family Planning Promotion Project (Kiragu, Galiwango, et al., 1996, p. 85).

included in the regression. Other confounding variables not identified may still account for some of the observed relationship between communication exposure and family planning behavior. Knowing that these eight are not confounding variables, however, considerably increases confidence that the campaign did have an impact on modern contraceptive use among Ugandan women.

Regression analysis of follow-up survey data only still suffers from the threat of selectivity bias. Those who already used modern methods may simply have paid more attention to the campaign than those who did not, and therefore they reported higher exposure. The best way to overcome selectivity bias in a full-coverage program is with longitudinal analysis—to collect panel data on the same individuals before and after some have been exposed to the communication program. Then it is possible to control for prior family planning behavior (Moffitt, 1991).

Population Communication Services used a longitudinal survey design to evaluate the 1993 National Communication Campaign in the Philippines. Three waves of data were collected from the same men and women before, during, and after the campaign (Kincaid, Coleman, & Rimon, 1995). This allows for multiple regression analysis to control for the prior status of any outcome variable, including contraceptive use (see Table 7.3). The dependent variable, current contraceptive use, assumed three values—0 for nonuse, 1 for traditional contraceptive use, and 2 for modern contraceptive use—so multinomial logistic regression was used for the statistical analysis. Both traditional and modern contraceptive users were compared with nonusers. The relative risk ratios indicate how much greater the proportion of method use is among those who are high on a predictor variable, such as campaign exposure, compared with those who are low on that variable.

Because the longitudinal design consisted of three waves of panel survey data collected from the same men and women, it was possible to estimate the effect of campaign exposure on contraceptive use at wave 2 and wave 3 while controlling for the influence of contraceptive use at each previous point in time as well as for the effect of other confounding variables.

Before the campaign began (wave 1), the best predictor of traditional method use was a positive attitude toward the rhythm/NFP (natural family planning) method, as indicated by perceptions that the method was healthy, effective, safe, moral, and easy to use. Those with a positive attitude toward rhythm/NFP had a probability of using a traditional method 1.63 times greater than those who had a negative attitude, after the effects of all of the other variables are controlled. The best predictors of modern method use in wave 1 (before the programs) were a positive attitude toward oral pills and a negative attitude toward rhythm/NFP.

Analysis of the survey data at wave 2, after the first phase of the campaign, showed that exposure to communication had a statistically significant impact on both traditional and modern method use. At wave 2, respondents were 1.45 times more likely to use traditional methods and 2.23 times more likely to use modern methods if they had been exposed to the campaign, after all other variables including prior method use are controlled. At wave 3, respondents exposed to the campaign during the second and third phases were 1.57 times more likely to use a traditional method and 1.9 times more likely to use a modern method.

Table 7.3
Multinomial Logistic Regression of Contraceptive Use on Campaign Exposure, Attitudes, and Sociodemographic Characteristics, Philippines, 1994

Variable	Wave 1 (N=1,421)		Wave 2 (N=1,486)		Wave 3 (N=1,105)	
	RR Ratio	95% Confidence Interval	RR Ratio	95% Confidence Interval	RR Ratio	95% Confidence Interval
Traditional Method Use						
<i>Campaign exposure</i>	—	—	1.45	1.05–2.01	1.57	1.09–2.25
<i>Pill attitude</i>	NS	—	0.64	0.47–0.86	NS	—
<i>Rhythm attitude</i>	1.63	(1.26–2.11)	2.16	1.61–2.89	1.38	1.00–1.92
<i>Prior modern use</i>	—	—	5.61	3.29–9.58	2.01	1.15–3.51
<i>Prior traditional use</i>	—	—	6.94	5.12–9.41	4.44	3.14–6.29
Modern Method Use						
<i>Campaign exposure</i>	—	—	2.23	1.44–3.45	1.90	1.22–2.95
<i>Pill attitude</i>	2.15	(1.62–2.86)	1.50	1.03–2.20	2.35	1.30–3.45
<i>Rhythm attitude</i>	0.70	(0.53–0.94)	NS	—	0.62	0.42–0.91
<i>Prior modern use</i>	—	—	86.38	50.1–149.0	25.54	15.25–42.76
<i>Prior traditional use</i>	—	—	2.98	1.86–4.70	2.01	1.28–3.16

RR Ratio = relative risk ratio.

NS = not significant.

Note: All RR ratios are also controlled for age, sex, education, rural-urban residence, and socioeconomic status. All RR ratios shown are statistically significant beyond the .05 level.

Source: Kincaid, Coleman, & Rimon, 1995.

It is clear from this analysis that previous use of either type of method was the best predictor of subsequent method use. In other words, past behavior was the best predictor of future behavior. At wave 3, for example, respondents were 4.44 times more likely to use a traditional method if they already used one at wave 2, but also 2.01 times more likely to use a traditional method if earlier they had used a modern method. This indicated that some Filipinos were switching methods over time. Respondents who used a modern method at wave 1 were 86 times more likely to use one at wave 2, and those who used a modern method at wave 2 were 25.5 times more likely to use a modern method at wave 3. These findings have major

implications for program design since they suggest that more attention should be paid to prior users and their reasons for discontinuation.

LESSON 4.

An interrupted time-series analysis of service statistics provides additional evidence of communication impact, especially when one of the objectives of the program is to promote service providers.

Service statistics as an evaluation tool were discredited early in the history of national family planning programs because the data were unreliable. Program statistics on new and continuing family planning clients were often inflated by the pressure of national targets or wishful thinking. Definitions of both new and continuing clients differed. Even more often, records were inconsistently and erratically maintained and not correctly aggregated from different facilities.

Nevertheless, in some cases where model clinics and fee-for-service providers keep service or sales statistics current, an interrupted time-series analysis based on these statistics can be a useful extension of the one-group research design. This design is especially appropriate when one of the objectives of the program is to increase use of family planning services at a defined number of facilities that have reliable statistics. In an interrupted time-series design the same group in effect serves as its own control group at an earlier point in time. Any changes in the direction or rate of change in service use after a communication program—especially abrupt changes—can be attributed to that program if no other known events or influences could have caused a change of the size observed.

Evaluation of the impact of three different communication activities over 12 years in the history of the Pro-Pater vasectomy clinic in São Paulo, Brazil, is an example of how this type of longitudinal analysis can be used to gauge impact (Kincaid, Payne Merrit, et al., 1996) (see Figure 7.2; see also Box 5.1 in Chapter 5). Each of the abrupt and substantial increases in vasectomies performed after the three mass media campaigns was statistically significant.

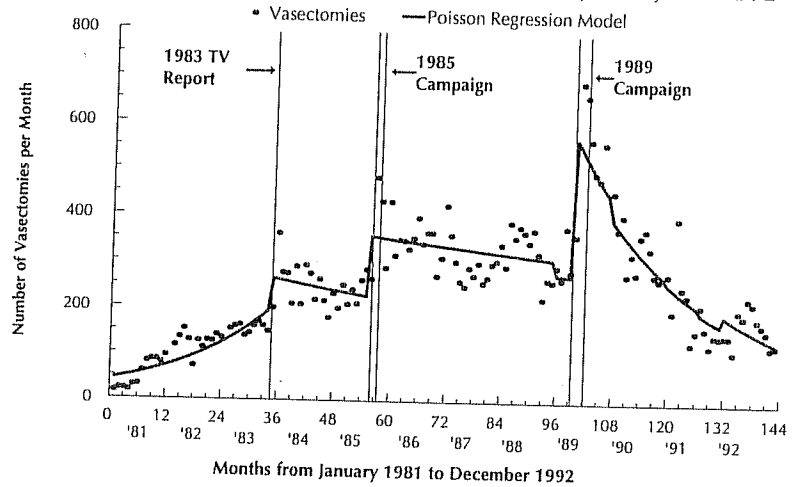
LESSON 5.

Evaluations can benefit from asking audience members for their own opinions about the impact of communication.

Rigorous design, theory, and analysis are necessary to demonstrate the impact of communication, but the opinions of the audience also are important. Most evaluation researchers discount this type of data because it is based merely on opinion. Opinion can be biased, especially in the case of survey respondents who may know that the purpose of the study is to evaluate the intervention and may want to please the interviewer.

Some of these problems with opinion data can be eliminated or reduced by a careful wording of questions, however. Open-ended questions that do not lead the

Figure 7.2
Effect of Media Events on Number of Vasectomies Performed per Month, and Poisson Regression, Pro-Pater Clinic, São Paulo, Brazil, 1981–1992



Source: Kincaid, Payne Merritt, et al., 1996.

Reproduced with the permission of The Alan Guttmacher Institute from D. Lawrence Kincaid, et al. "Impact of a Mass Media Vasectomy Promotion Campaign in Brazil." *International Family Planning Perspectives*. Volume 22. Number 4. December 1996.

respondent or indicate the desired answer are better than closed or precoded questions that do. Once exposure to a mass media campaign has been established, it is better to ask what, if anything, the survey respondent did as a result of hearing the messages rather than to ask directly, "Did you visit the health clinic as a result of hearing the message?" Later questions may confirm that the same respondent visited a clinic during the campaign, but only the respondent is in a position to say whether she thinks that the campaign was one of the reasons that led her to do so.

The opinion of the participants is worthwhile even if it is somewhat biased. It is especially valuable as a complement to other types of evidence. On one hand, if none of the audience members exposed to the program's communication say that it influenced them to do anything, then the other evidence regarding the effects of communication is suspect. On the other hand, if a significant portion of those exposed can confirm that the communication made a difference for them, then the other findings are reinforced.

The evaluation of the *jiggasha* approach in Trishal, Bangladesh, did ask the 266 participants for their opinions (Kincaid, Kapadia-Kundu, et al., 1994). Over half of the participants said that as a result of the *jiggasha* group meetings they discussed family planning and the *jiggasha* experience itself with their husbands and with other women. About 30 percent said that they adopted a modern contraceptive method; 22 percent said that they decided to continue using their current method; 17 percent said that they visited a family planning clinic; and 7 percent said that they switched to another method as a result of participating in *jiggasha* meetings. These opinions were consistent with the other conclusions of the

evaluation, thus confirming that the *jiggasha* intervention had an impact. If used in conjunction with other types of evidence, asking the opinion of participants—even if all bias cannot be eliminated—is better than not having their opinions at all.

THEORY-DRIVEN EVALUATION

LESSON 6.

Evaluations that include subobjectives based on theoretical models such as the steps to behavior change provide evidence of impact on behavior and useful feedback to improve future programs.

To measure the process of communication and behavior change, evaluation should be not only *method-driven* but also *theory-driven*—that is, based on theories of change that explain or predict why and how individuals will respond to the program (Chen & Rossi, 1983 & 1987; Chen, 1990; Shadish, Cook, & Leviton, 1991). In family planning and health communication, these theories specify the causal links between exposure to communication and the behavior change intended. For evaluation purposes, these causal links should be the same ones used to develop the communication strategy for the project.

For communication projects, the Steps to Behavior Change model describes a clear set of intervening steps that are expected theoretically to take place in response to communication and that are expected to lead to sustained behavior change. Achieving these intervening steps should be treated as *subobjectives* of the project. Appropriate indicators can then be devised to measure these steps as well as the primary objective of behavior change. If the subobjectives are achieved as predicted by theory, and if these steps can be shown to be related to the desired outcome (behavior change), then the evaluation not only has evidence of success or failure but also has a more precise explanation of *why* that outcome occurred and with whom. This is the kind of information that can be used to improve the strategic design of future communication projects. This basic approach to evaluation, of course, can be used with other theoretical models of change.

In the *jiggasha* experiment described above, longitudinal sample surveys of the same women in all 24 villages were conducted by Mitra and Associates before and after the program. Thus, in addition to identifying differences in outcome between the experimental and control areas, the evaluation also could identify many other differences among individual women. The surveys collected information from each woman that corresponded to several of the subobjectives specified by the Steps to Behavior Change model. Statistical analysis of these variables revealed which steps were most closely related to individual behavior.

As Table 7.4 shows, the women who rated high or positive on each of the five steps measured were more likely to be using a modern contraceptive method than

Table 7.4
Steps to Behavior Change and Their Relationship to Modern Family Planning Use Among Married Women, Trishal Thana, Bangladesh, 1992

Intervening Variable	% High or Yes	% Using Modern Method When Intervening Variable Is:		Crude Odds Ratio*
		Low/No	High/Yes	
<i>Unaided knowledge of 4-6 modern FP methods</i>	68.3	15.4	30.2	2.4
<i>Positive FP attitude</i>	57.5	18.2	30.9	2.0
<i>Discussed FP with other women</i>	43.9	17.0	36.3	2.8
<i>Discussed family size or contraceptives with husband</i>	67.5	12.4	31.9	3.3
<i>Husband approves of FP practice</i>	74.8	5.7	32.1	7.8

FP = family planning.

*N = 1,705 to 1,707; all differences are statistically significant beyond the .001 level of probability.

Source: Kincaid, Das Gupta, et al., 1993.

those who rated low. Of the five steps that were measured, husband's approval was their husbands approved of their practicing family planning, 32.1 percent were using a modern method, compared with only 5.7 percent of women whose husbands disapproved. In other words, women whose husbands approved were 7.8 times more likely to use a modern method. (The odds-ratio calculation is useful because it best indicates the relative strength of each variable compared with the others.)

All five variables—positive attitudes, high levels of knowledge, discussions with other women, discussions with husband, and husband's approval—come from the Steps to Behavior Change model. There is a theoretical basis for predicting that each of these variables would be related to the intended behavior. Indeed, that was the same theoretical basis for designing the *jiggasha* concept of group discussion and networking from the start. Because it helps to explain *why* the difference occurred, the findings that, in practice, the *jiggasha* approach was associated with the results predicted by theory, strengthen the conclusion that the program was in fact responsible for the differences measured in the experimental design. It also helps to dismiss the possibility that some other event may have caused the difference (spuriousness).

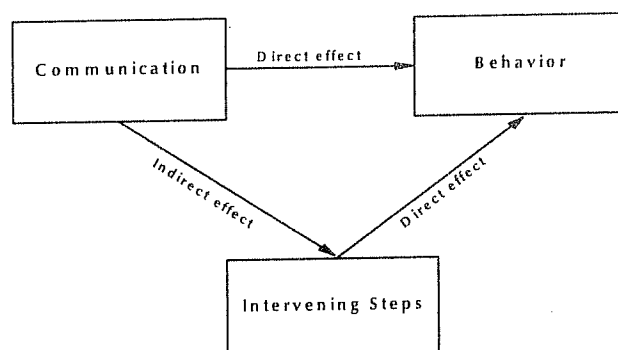
In addition, this theory-based evaluation provides useful conclusions for future programs: All of these variables should continue to be emphasized in the future, but, since husband's approval made more difference to contraceptive use than the other four, even more attention should be given to that aspect of the *jiggasha* approach. The men of Bangladesh, often neglected in maternal and child health programs, should become a more important focus for future communication activities.

LESSON 7.

A thorough evaluation of the impact of communication includes an analysis of its indirect effects as well as its direct effects on health behavior.

Evidence that these theoretical linkages are indeed related to the desired behavior outcome contributes substantially to the evaluation of impact, but evaluation must also show that exposure to the communication had an impact on these theoretically linked variables. In Bangladesh, did the *jiggasha* intervention itself make a difference in the attitudes and knowledge of participants? If so, then the communication had an indirect effect on behavior, through its effect on these causal links, as well as a direct effect. In terms of analysis, the five theoretical variables (steps) operate as *intervening variables* between exposure to communication and behavior (see Figure 7.3). The analysis of the *jiggasha* experiment in Lesson 6 showed that selected subobjectives (steps) were strongly related to modern

Figure 7.3
Schematic Diagram of the Direct and Indirect Effects of Communication



contraceptive use. Further analysis shows whether and to what extent the individual women who actually participated in *jiggasha* sessions differed, with respect to these five selected variables, from women who did not participate.

As Table 7.5 indicates, *jiggasha* participation had a statistically significant relationship with all but one intervening variable. *Jiggasha* participation appears to be most strongly associated with discussion of family planning with other women. Participants were 3.1 times more likely than nonparticipants to discuss family planning with other women. The *jiggasha* participants also were 2.1 times more likely to recall more than four modern methods, 1.7 times more likely to discuss family planning with their husbands, and 1.4 times more likely to have highly positive attitudes. Thus, four of the five subobjectives (steps) found to be related to modern contraceptive use also were significantly related to *jiggasha* participation. Evidence of indirect effects such as this increases confidence that communication had an impact on contraceptive behavior *in ways that were expected theoretically*.

Table 7.5
The Impact of *Jiggasha* Participation on the Intervening Steps to Behavior Change Among Married Women, Trishal Thana, Bangladesh, 1992

Intervening Variable	% Rated High on Intervening Variable by <i>Jiggasha</i> Participation		Crude Odds Ratio ^a
	Did Not Participate	Did Participate	
<i>Positive FP attitude</i>	56.5	63.8	1.4*
<i>Unaided knowledge of 4-6 modern FP methods</i>	66.0	80.3	2.1**
<i>Discussed FP with other women</i>	39.9	67.0	3.1***
<i>Discussed family size or contraceptives with husband</i>	65.7	76.6	1.7**
<i>Husband approves of FP practice</i>	73.7	79.4	1.4

FP = family planning.

Note: The 229 *jiggasha* participants represented 12.2 percent of the 1,877 women surveyed in all 24 villages in the follow-up survey in 1992, and 24.0 percent of the women surveyed in the 12 villages where the *jiggasha* intervention was implemented.

^aLevel of statistical significance: *p < 0.05; **p < 0.01; ***p < 0.001.

Source: Kincaid, Das Gupta, et al., 1993.

This information has immediate implications for improving the *jiggasha* intervention because it provides evidence on *why* and *how* the intervention influenced modern contraceptive use. For example, the data show that the strongest intervening variable for contraceptive use, husband's approval, was *not* significantly related to *jiggasha* participation at the individual level. Perhaps this was because participants and nonparticipants both had fairly high husbands' approval, 79.4 percent versus 73.7 percent. Still, it suggests that the *jiggasha* did not help women win their husbands' approval for family planning. For the 25 percent or so of these Bangladeshi women whose husbands disapproved, this disapproval appears to be an important constraint: 94 percent of these women do not currently use modern contraceptives.

These findings suggest that the *jiggasha* intervention could be improved to address this issue further. What can be done in *jiggasha* meetings for men, as well as for women, to influence the remaining opposing husbands to change their minds? Is it acceptable to talk openly about this during *jiggasha* meetings? To identify wives whose husbands still object, find out why they object, and what can be done to change their minds? Perhaps male opinion leaders in the village would be willing to talk to these men about it. Perhaps just informing the disapproving husbands that a majority of men already approve might be persuasive. Ideas such as these are now being incorporated into the revised field worker training curriculum for the next phase of the program.

LESSON 8.**Exposure to communication and the steps to behavior change have cumulative effects on behavior much like cumulative risk and dose response in epidemiologic research.**

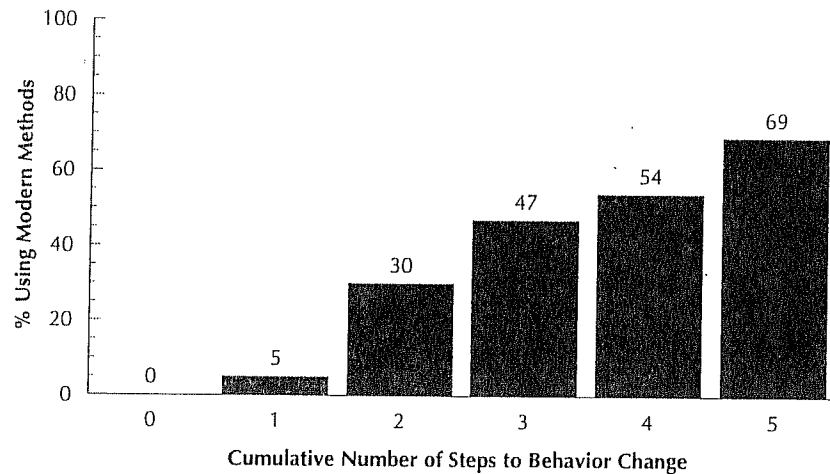
In research on communication, the concept of cumulative effects or impact is analogous to the concept of cumulative risk in epidemiological studies of disease and to a dose response in medicine (Schlesselman, 1982). In communication evaluation, cumulative impact means that the greater the number of steps to behavior change that apply to any individual, the greater the probability that the individual will adopt that health behavior. Thus an evaluation that shows this provides powerful evidence of impact. Such evaluation also is useful in determining the marginal utility of seeking to achieve each additional step.

As indicated in Chapter 2, the Steps to Behavior Change model depicts a process. But all individuals do not necessarily pass through the steps in the same order, nor do all individuals necessarily pass through all of the steps before they change their behavior. For example, a woman may talk to her husband and friends about family planning before she knows much about contraceptive methods or without necessarily having a positive attitude about practicing family planning. She may even try the first method she hears about, discontinue it, and then try others until eventually she finds one that is suitable. Meanwhile, her knowledge increases and her attitudes toward contraceptive use become more positive as she learns the benefits first hand or through talking with other women. Such complex adaptive behavior is not fully captured by staged models of communication and behavior. What matters, however, is that the important steps are identified and that in combination they make a difference to behavior.

Several Johns Hopkins evaluations in different cultures have demonstrated that each of the steps to behavior change *by itself* usually has a statistically significant association with contraceptive use (as already shown for Bangladesh in Table 7.1). It has now been demonstrated, in addition, that these steps, regardless of their order, have a cumulative impact on behavior (Kincaid, 1995a; Kincaid, 1996). The measures of the steps to behavior change are usually so closely correlated with one another that they meet the statistical criterion to be combined into a single, underlying factor. If the number of steps (variables) that apply to an individual are simply summed, the resulting measure of *how many* intervening steps apply represents the strength of that factor. Analysis of the cumulative measure of the steps to behavior change can show how much impact each additional intervening variable has on behavior. Evaluation of family planning communication is the first area of communication research where this phenomenon has been demonstrated (Kincaid, 1995a; Kincaid, 1996).

In Bangladesh, when the *jiggasha* project was replicated in three new thanas, and a one-group, before-after longitudinal panel survey was used for evaluation (Kincaid, 1996), there were 1,479 women in the 1994 baseline survey and 1,058 of these women in the 1996 follow-up plus a supplemental sample. The same five

Figure 7.4
Modern Contraceptive Use Among Eligible Married Women by the Cumulative Number of Steps to Behavior Change, Bangladesh, 1994



N = 1,101.

Chi² = 170.0.

p < .0001.

Source: Kincaid, 1996.

intervening steps or variables were used as described in Lessons 6 and 7—women's knowledge, approval, discussion with other women, discussion with husband, and husband's approval.

As the bar graph in Figure 7.4 above illustrates, the percentage of eligible married women using modern methods increased cumulatively and significantly with each additional step to behavior change that applied. Most strikingly, the level of use of modern contraceptives among women who rated low on all five steps is *zero*—no users whatsoever. Among women who have taken just one step, the prevalence rate is 5 percent. With two steps, prevalence jumps to 30 percent, and then up to 54 percent with four steps. Women to whom all five steps apply have a modern contraceptive use rate of 69 percent, which is about the level of contraceptive use in many developed countries.

A comprehensive evaluation can identify a cumulative impact of communication similar to a dose response. As Table 7.6 shows, women who were exposed to a strong communication intervention—*jiggasha* discussion groups led by a family welfare assistant—were most likely to use modern contraceptive methods: almost 75 percent did so. These are the cumulative direct effects of communication activities on the behavior outcome. Some 56 percent of those who were visited by a family welfare assistant at home but did not attend *jiggashas* were modern method users, but only 17 percent of women exposed to neither *jiggasha* nor home visits.

A similar dose response effect is evident in the indirect impact of exposure on the intervening steps to behavior change. Of the women who participated in *jiggasha* discussion groups, 88 percent ranked high (with four or five steps) on the

steps to behavior change. In comparison, only 35 percent of women exposed to neither intervention and about 68 percent of those visited by a field worker at home only ranked high on the five steps.

Table 7.6
Modern Contraceptive Use and Level of Steps to Behavior Change Among Eligible Married Women in Bangladesh by Level of Family Planning Communication¹

Variable	Level of Family Planning Communication			Total (N)
	None (%)	Home Visit by FWA (%)	Jiggasha with FWA (%)	
Direct Effect: Modern Contraceptive Use²				
<i>No</i>	83.0	44.0	25.5	556
<i>Yes</i>	17.0	56.0	74.5	545
Indirect Effect: Number of Steps to Behavior Change³				
<i>Low (0-3)</i>	64.8	32.3	11.7	408
<i>High (4-5)</i>	35.2	67.7	88.3	693
<i>Number of Cases</i>	253	703	145	1,101

FWA = family welfare assistant.

¹Sample consists of married women ages 12-49 who are eligible for family planning: not pregnant, breastfeeding, or infertile.

²Chi² = 155.2; p < .0001.

³Chi² = 130.4; p < .0001.

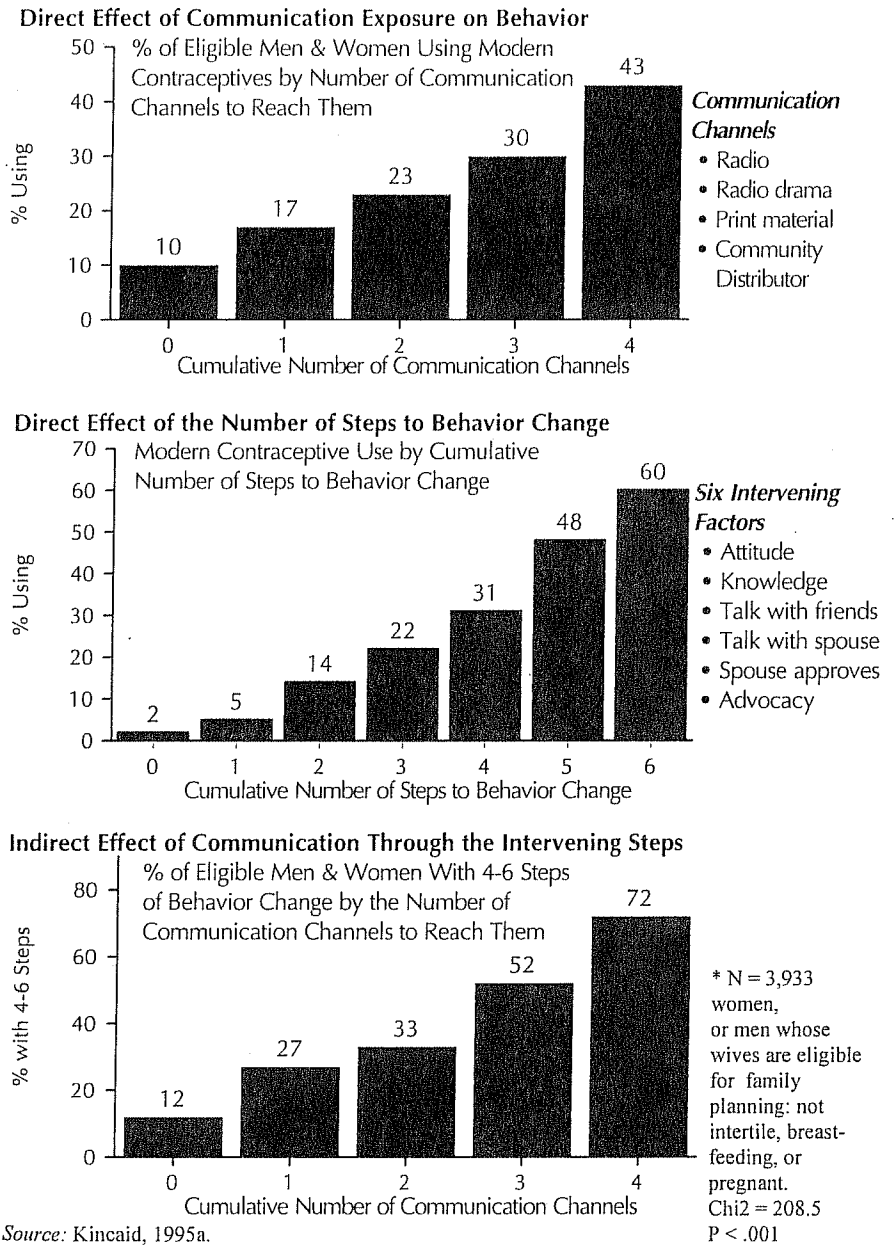
Source: Kincaid, 1996.

The same type of analysis can be used to evaluate predominantly mass media communication programs. In Kenya the *Haki Yako* family planning campaign was evaluated through a 1994 national sample survey of 4,459 men and women. Communication activities were divided into four principal channels: a radio serial drama, radio in general, print materials, and visits by community-based workers. The evaluation showed a similar dose response to levels of communication exposure and a similar cumulative impact—both direct and indirect—of the steps to behavior change (Kincaid, 1995a). The results are presented in the form of three related bar graphs in Figure 7.5.

The first bar graph shows the direct cumulative effect of exposure to family planning messages through the four communication channels. There is a clear dose response: The greater the number of channels of exposure to family planning messages, the higher the rate of modern contraceptive use. The prevalence rate is only 10 percent for men and women with exposure to none of the four channels. The rate increases steadily with each channel added, reaching a high of 43 percent using modern methods among those exposed to all four channels.

The second bar graph shows the direct effect of six selected steps to behavior change on modern contraceptive use. Men and women for whom none of the six

Figure 7.5
Direct and Indirect Effects of Family Planning Communication Among Men and Women in Kenya, 1994 *



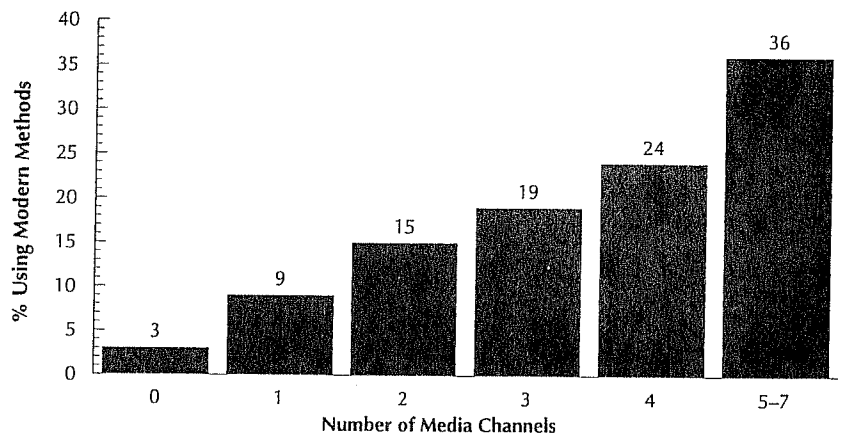
Source: Kincaid, 1995a.

intervening steps applies have a modern contraceptive use rate of just 2 percent. The rate increases dramatically with the addition of each new step. The use rate reaches 31 percent for those to whom four steps apply, which exceeds the national modern contraceptive prevalence rate of 20.4 percent. The men and women to whom all six factors apply have a modern prevalence rate of 60 percent.

The third bar graph shows the strong dose response of the number of channels on the six intervening steps related to contraceptive adoption. Only 12 percent of respondents with no communication exposure have high levels (4 to 6) on the steps to change. The percentage of respondents who are at high levels increases very rapidly with the addition of each channel of family planning communication: 72 percent of those exposed to all four channels of communication have high levels on the steps. Similar dose responses to communication exposure have been found in evaluations of family planning communication programs in Zimbabwe (Piotrow, Kincaid, et al., 1992), Nigeria (Kiragu, Krenn, et al., 1996), Bolivia (Valente, Saba, et al., 1996), and Uganda (Kiragu, Galiwango, et al., 1996).

Even where separate surveys are not fielded to evaluate communication impacts, other surveys, such as Demographic and Health Surveys, can be analyzed to identify a dose response or cumulative impact of communication. In Tanzania, for example, a 1994 national sample survey of 4,225 women ages 15 to 49 (Weinstein et al., 1994), conducted by the Tanzanian Bureau of Statistics and Planning Commission, showed that during the preceding two years the modern contraceptive prevalence rate had nearly doubled, rising from 5.9 percent to 11.2 percent. A statistical analysis of the cumulative effect of seven family planning channels used during those two years (the *Zinduka* radio serial drama, other radio programs, the family planning logo, newspapers, posters, brochures, and

Figure 7.6
Modern Contraceptive Use Among Women by Level of Exposure to Communication Channels, Tanzania, 1994



N = 4,225 women ages 15-49.

Source: Adapted from Jato, Simbakalia, et al., 1996.

television) showed the expected stairstep pattern with modern contraceptive use (Jato, Simbakalia, et al., 1996). (See Figure 7.6.) The contraceptive prevalence rate among women not exposed to any channel was only 3 percent; this increased to 24 percent for those exposed to four channels and peaked at 36 percent for those exposed to five or more channels. Thus the modern contraceptive use rate for women exposed to family planning messages through more than five channels was three times higher than the national rate of 11.2 percent.

LESSON 9.

Aggregating survey data by local units of analysis such as villages or sample clusters reveals the impact of communication on local culture in addition to its impact on individuals.

Most Population Communication Services evaluations involve surveys to measure individual behavior change. These, like most other surveys, are carried out with samples designed to select a certain number of independent individuals with a known probability. For economic and practical reasons, large surveys are conducted by multistage cluster sampling procedures. That is, through systematic or simple randomized sampling methods, certain geographical units within a country are selected first, followed by a second randomized sample of smaller units within the larger area, such as villages and urban neighborhoods or blocks. Within these units, maps are usually drawn of all of the households, and then a randomized sample of households is selected. Households are then selected so that they are geographically dispersed in order to reduce the interdependence that occurs due to communication and other shared cultural influences among individuals who live in close proximity. A final randomized selection procedure ensures that every eligible individual within each household has an equal probability of being selected. As a result, the sample is representative of the population from which it is drawn. The smallest geographical unit—the village or neighborhood—is usually referred to as a “cluster.”

The effects of local culture within these clusters on the individuals who are ultimately sampled cannot be entirely eliminated. Some villages and neighborhoods are wealthier than others, for example, or include more individuals with higher levels of education or who own radios, so that the average levels of income, education, or radio ownership for that cluster are higher than for other clusters. Also, some clusters are closer than others to roads, schools, and health/family planning service facilities, and hence are visited more often by outreach workers.

For randomized sampling and for analyzing individual behavior, these local cultural effects are a nuisance, and efforts are usually made to adjust the statistical analysis for such effects. For understanding social and cultural change, however, cluster effects are quite informative. Although it may seem obvious that the individual is the unit who adopts a new behavior, the evidence and many theories suggest that change also occurs at a higher level, the level of cultural units. Why should this be so? Unless a woman believes that those in her community—friends

and neighbors as well as local leaders—publicly endorse a new technology or behavior, it is difficult for her to adopt it. Stated another way, once the new behavior seems socially acceptable, it is easier for individuals to adopt it (Fishbein & Ajzen, 1975). This phenomenon is one of the rationales for conducting community mobilization programs in public health and for including discussion and advocacy of family planning in the Steps to Behavior Change model.

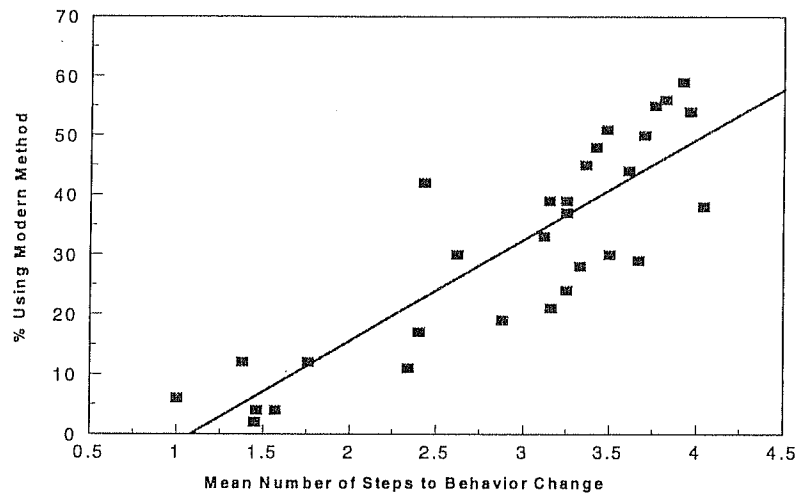
PCS evaluations have identified a strong village cluster effect on use of family planning in Bangladesh. In the 1989 baseline survey within the socioeconomically homogeneous thana of Trishal, the contraceptive prevalence rates of the 24 villages varied from 7 percent to 58 percent (Kincaid, Das Gupta, et al., 1993). At that time, the national rate and the overall trishal rate were about 32 percent (Kincaid, 1995b). When the *jiggasha* project was replicated in 1994, surveys in 30 villages of three thanas revealed the same wide range of contraceptive prevalence among villages, from 9 percent to 58 percent (averaging 40 percent) (Kincaid, 1996).

Analysis of the 1996 follow-up survey for the *jiggasha* replication found that there was also a village-level relationship between modern contraceptive prevalence and the steps to behavior change. When the individual-level data were aggregated at the cluster level, there was a high correlation ($r = 0.85$) between the modern contraceptive prevalence rate and the women's mean number of steps to change. At the individual level, the comparable correlation was $r = 0.39$, moderately high but much lower than the village-level correlation. In technical terms, this means that 72 percent of the variance in modern contraceptive use was explained by these cumulative steps to behavior change at the cluster or village level of analysis. This relationship, illustrated in Figure 7.7, is linear: the higher the cluster's average level on the intervening steps to behavior change, the higher the rate of modern method use.

Figure 7.7 supports the conclusion that contraceptive adoption varies among local cultural units according to ideational factors, as measured by the five steps to behavior change. This finding implies that family planning adoption occurs at the level of villages as well as individuals. It corresponds to findings on the fertility transitions in Europe described in Chapter 1—namely, that fertility decline spreads first through similar cultural and linguistic areas within countries, regardless of the levels of socioeconomic development (Coale & Watkins, 1986). This finding helped to give rise to the ideational theory of fertility transition, which states that fertility decline spreads primarily through the communication of new ideas (Cleland & Wilson, 1987).

The same cultural effect can be shown in national-level sample surveys of countries where mass media programs play a larger role. Following similar procedures, a 1994 Kenyan national sample survey of 4,459 men and women was aggregated at the cluster level for further analysis. This yielded a total of 268 clusters with an average size of 16.6 individuals per cluster. Figure 7.8a shows the scatterplot and regression line of the modern contraceptive prevalence rate on the average number of the five steps to behavior change among men and women in the 268 Kenyan clusters. Each solid square represents the level of a cluster on the two variables. The correlation at the aggregate level is 0.54. This represents a fairly

Figure 7.7
Regression of Modern Contraceptive Prevalence on Five Intervening Steps to Behavior Change in 30 Bangladesh Villages, 1996



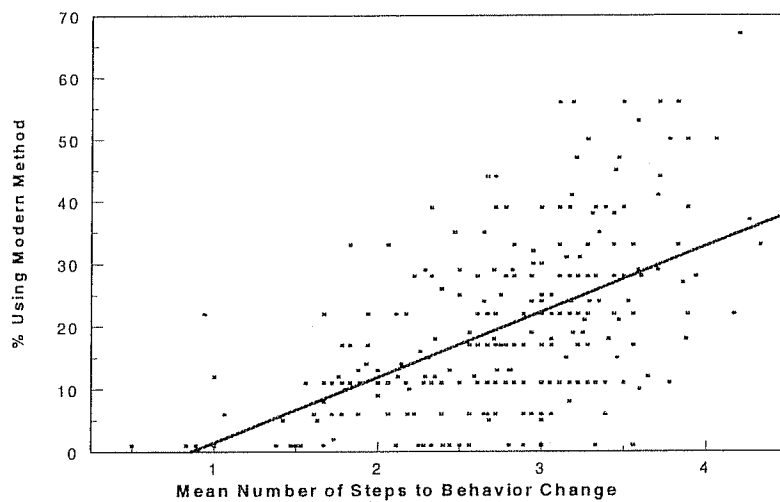
Note: $r = .85$.

strong linear relationship between local culture (ideation) and family planning practice in Kenya, but not as strong as the relationship reported for Bangladesh. This probably reflects the fact that Bangladesh is a more homogeneous society with less communication among villages.

In Kenya, exposure to family planning messages, via media or in person, also varies by local community, and the community level of media exposure is positively related to local ideational factors (as indicated by the intervening steps to behavior change). The scatterplot in Figure 7.8b shows the relationship between the average number of steps to behavior change and the average cumulative exposure to family planning media in Kenya by cluster. The correlation is 0.55.

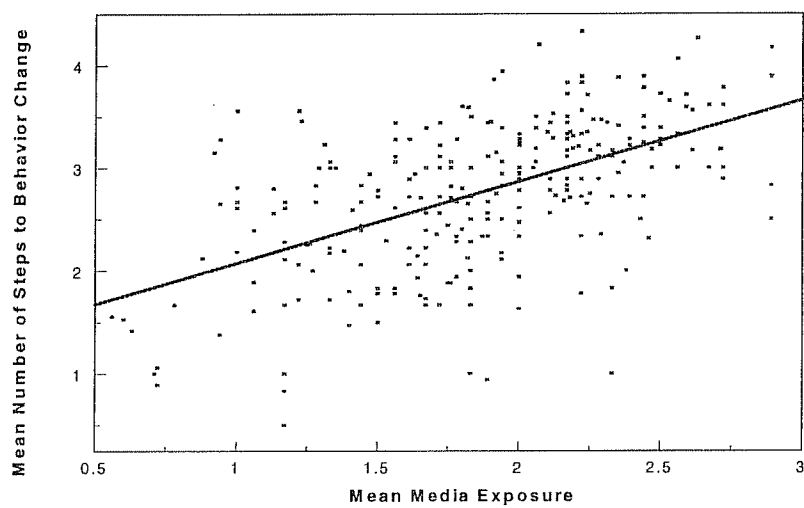
The ideational theory and the results from this type of evaluation and analysis in developing countries have important implications for communication programs: Health and family planning communication programs should be strategically designed for communities as well as for individuals. This, too, was one of the primary motives behind the *jiggasha* project in Bangladesh. Changing the field workers' pattern of work from individual home visits to local group meetings was expected to strengthen the community norm supporting family planning and thus stimulate individual practice. For evaluation, the lesson is that, when large national sample surveys are used for individual-level analysis, the results also should be aggregated at the cluster level to examine the impact of programs on social units. Then programs can be developed with the people and leaders of those units to address their concerns.

Figure 7.8a
Regression of Modern Contraceptive Prevalence on Six Cumulative Steps to Behavior Change in 268 Clusters in Kenya, 1994



Note: $r = .54$.

Figure 7.8b
Regression of Mean Number of Steps to Behavior Change on Media Exposure in 268 Clusters in Kenya, 1994



Note: $r = .55$.

Source: Kincaid, 1995a.

LESSON 10.

Path analysis can depict and test the direction and strength of all the relationships measured in a communication program evaluation.

One of the most advanced methodologies that can be used to represent and test the complex relationships described above is *path analysis*, also referred to as structural equation modeling (Hayduk, 1987; Glymour et al., 1987; Joreskog & Sorbom, 1979). When used with longitudinal survey data and when the theoretical relationships underlying the causal process are understood, path analysis offers a practical compromise. This compromise achieves the internal validity obtained from the randomized control group design and the external validity (generalizability) obtained from the sample surveys used in the one-group, before-after designs. The main disadvantage of using path analysis for evaluation research is that its use and interpretation require a sophisticated statistical background (Chen, 1990). If the path models are kept as simple as possible and can be clearly diagrammed, however, then the method offers many advantages for program evaluation.

In the case of the evaluation of the *jiggasha* approach in Bangladesh, for example, just three program variables and eight control variables were used:

- the outcome behavior (family planning practice),
- communication (*jiggashas* with family welfare assistants—FWAs), and
- the ideation factor (five steps to behavior change).

Communication and the ideation factor each were treated as having direct effects on family planning, and communication was also treated as having an indirect effect on family planning through its effect on the ideation factor. These same relationships were diagrammed as a set of paths in Figure 7.9. To test these relationships rigorously with data, however, it was necessary to control for any confounding variables that may be influencing each of the three variables and to assess time-order and the threat of selectivity (see Lesson 3). To do so required that a multiple regression analysis be conducted for each of the three variables and that the three resulting regression equations be evaluated as a whole set or system. Two computer programs have been developed to make this easier to do: LISREL (Joreskog & Sorbom, 1979) and EQS (Bentler, 1995). Also, a conventional method for diagramming the results has made them easier to interpret.

The longitudinal survey data from the evaluation of the *jiggasha* replication in Bangladesh were submitted to the EQS program to test a path model for family planning use and intention. Five degrees of intention to use (ranging from “definitely will not practice” to “definitely will practice in the future”) were combined with traditional and modern contraceptive use to create a seven-point, continuous scale required for the analysis. Ideation comprised the five cumulative steps to behavior change described above, while the measure of communication

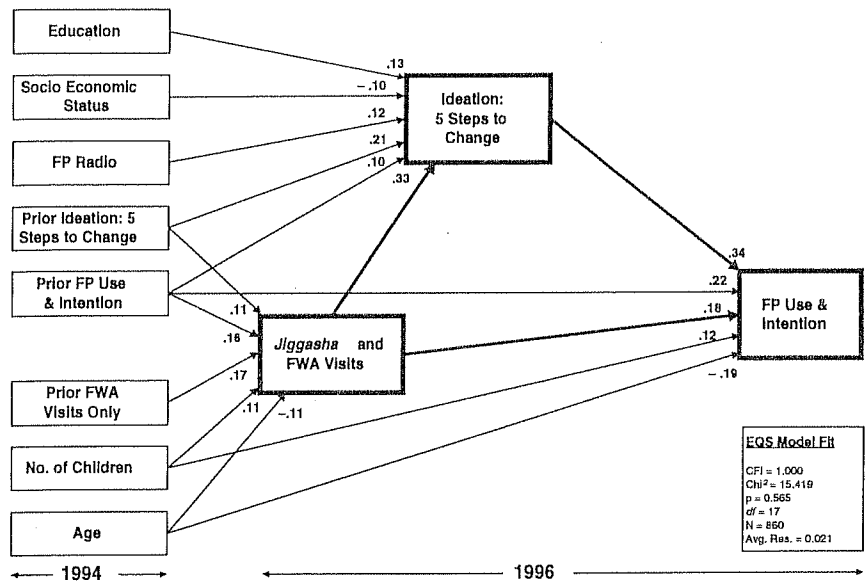
consisted of no contact, FWA visits at home, and participation in *jiggashas* led by FWAs.

The resulting path analysis is diagrammed in Figure 7.9. Each line in the graph represents a causal influence. The direction of influence is indicated by the arrowhead. The longitudinal (panel) design of the survey makes it possible to assess the efforts of the control variables and program variables measured in 1994 on the three program variables measured in 1996.² The paths of influence leading to each variable represent the results of the three multiple regression analyses. Only statistically significant relationships are shown. The three main relationships of interest for evaluation are represented by thicker lines: the direct and indirect effects of the *jiggashas* with FWAs and the direct effect of the ideation factor (five steps to behavior change [SBC]) on family planning use and intention. The numbers next to each arrowhead indicate the *relative size* of the causal influence. Technically, they are the standardized regression coefficients from the regression analysis for each variable. A negative sign next to a number indicates an inverse relationship. Thus, for example, younger women are more likely to participate in *jiggasha* meetings and to practice or intend to practice family planning (-.11 and -.19, respectively), controlling for the effects of number of children and other variables.

The diagram shows that only five variables have significant influences on family planning use and intention after controlling for all other variables. The strongest influence is the ideation/SBC factor (.34), followed by prior family planning use and intention (.22). The *jiggasha*/FWA intervention has the third strongest influence (.18), followed by the two control variables age (-.19) and number of children (.12). Education and socioeconomic status have no direct effect on family planning use and intention. They have influence only through their effects on the ideation factor. Also, socioeconomic status has a negative relationship with ideation. That is, lower-status women are higher on the five steps to behavior change. The strongest influence on ideation comes from *jiggasha*/FWA communication (.33), followed by a woman's prior (1994) ideation level (.21). *Jiggasha* and field worker visits in 1996 are determined mainly by prior FWA visits and prior family planning use and intention.

The path analysis illuminates the causal paths expected by the communication and steps to behavior change theories used to design the *jiggasha* program and to evaluate it. Path analysis controls statistically for the confounding influences of other variables. And because a longitudinal (panel) survey was used, the time-order of the relationships and the problem of selectivity are taken into account. Thus the analysis makes possible this conclusion: Prior family planning use and intention, as measured in 1994, did influence who participated in *jiggasha* meetings and who was visited by family welfare assistants in 1996, but *jiggasha*/FWA communication in 1996 still had a significant influence on family planning use and intention, even after controlling for this source of selectivity. Using path analysis with longitudinal data accomplishes the two goals of program evaluation discussed throughout this chapter: (1) measuring the impact of communication and (2) specifying the causal process that accounts for that impact.³

Figure 7.9
A Path Model of the Direct and Indirect Impact of *Jiggasha* Participation on Family Planning in Bangladesh, 1994–1996



FP = family planning.
 FWA = family welfare assistant.
 Source: Kincaid, 1996.

EVALUATING CLIENT-PROVIDER COMMUNICATION

LESSON 11.

Because of the interactive nature of interpersonal communication and counseling, innovative research methods are needed for effective evaluation.

Interpersonal communication and counseling (IPC/C) are often crucial components in the process of behavior change. Most people start to use family planning only after counseling sessions with service providers. They also talk with their spouse or partner, friends, and others to seek information, confirmation, or emotional support—usually even before seeing service providers.

It is universally assumed that the quality of client-provider interaction in counseling sessions can influence whether clients use contraceptive methods correctly, and for how long. Therefore much attention is now being given to efforts to improve the quality of counseling. These efforts have included setting standards

and guidelines for counseling, upgrading supervision of counseling, providing counseling aids, empowering clients to expect and ask for appropriate information, and, most frequently, training providers in interpersonal communication and counseling skills.

In the past, evaluation of training was often limited to output measures alone: counting the number of providers trained, comparing pre- and post-training test scores, and collecting trainees' reports on their satisfaction with training. More in-depth evaluation of the content and quality of interpersonal communication and counseling has been difficult until recent years because of the lack of standards, the variability in desired outcomes, and the lack of measurement tools. Therefore, the difficult task of evaluating how providers apply the knowledge and skills learned during training when they return to their jobs is rarely undertaken.

To assess limited aspects of client-provider interaction, such as waiting times, use of counseling aids, respect for privacy, and the number of contraceptive methods mentioned, either observation during the session or "exit" interviews with clients afterward are needed. More important, follow-up evaluation of the impact of training or other client-provider communication on the clients' subsequent family planning behavior has been seriously neglected.

To improve the quality and impact of counseling first requires improving understanding of the counseling process. Several innovative methods of studying family planning counseling have been developed over the last 10 years and applied by Population Communication Services. These include:

- refined indicators of counseling skills;
- assessment of clients' perspectives on counseling quality;
- studying the interaction between client and provider during counseling sessions; and
- follow-up of clients' contraceptive behavior after counseling sessions.

Refined indicators of counseling skills. New indicators and corresponding data collection instruments have been developed to assess counseling skills based on the widely used GATHER approach to counseling (greet, ask, tell, help choose, explain, and return/refer) (see Box 4.1, in Chapter 4). Several instruments for observation, client interviews, provider interviews, and study site observation have been used in Nigeria (Y. M. Kim, Rimon, et al., 1992), Ghana (Y. M. Kim, Amissah, et al., 1994), Kenya (Y. M. Kim, Lettenmaier, et al., 1996), Zimbabwe (Y. M. Kim, Marangwanda, & Kols, 1996), and Nepal (Heckert et al., 1996). A generic set of these instruments is available in a manual from Johns Hopkins Center for Communication Programs (Y. M. Kim & Lettenmaier, 1995). The indicators and instruments were also adapted for use by the Population Council in the Situation Analysis studies (A. Fisher et al., 1996).

Client-provider interaction and client behavior. Most efforts to improve and then evaluate counseling quality have focused on what the provider does. The GATHER process, for example, focuses mainly on the actions of the provider. But evaluation of counseling quality should also address the role of the client during the

counseling session and examine the dynamics of the interaction between the client and the provider.

Population Communication Services, in collaboration with the Quality Assurance Project, AVSC International, and the Population Council, has developed and continues to refine a technique for "interaction analysis" to better measure and understand the complexity of these interactions. The technique uses transcript analysis. It involves recording audio or videotapes and then coding and analyzing these dialogues using both qualitative and quantitative techniques (Roter & Hall, 1992). The interactions are then studied to determine how clients participate in counseling, how providers respond to clients, and how decisions are made. This technique has been applied in Kenya, Zimbabwe, and Indonesia (Y. M. Kim, Odallo, et al., 1997; Y. M. Kim, Marangwanda, & Kols, 1996; Y. M. Kim, 1997).

Interaction analysis offers several advantages over direct observations or client exit interviews. First, transcripts provide an accurate record and do not rely on human recall of what happened. Second, complex communication can be disaggregated to assess separately the behavior of both clients and providers in the same context. Third, evaluators can use and study the transcripts repeatedly to get a variety of points of view from clients, providers, counseling specialists, program managers, and trainers.

An example from Kenya based on audio transcripts of 178 counseling sessions with women provides the following insights into client-provider interaction and suggests guidelines for improving counseling in the future:

- Providers tend to dominate the sessions and do not allow clients to express or implement their own agendas.
- Providers do not elaborate on the risks or benefits of specific methods.
- Providers do not obtain adequate medical histories to be able to assess the appropriateness of method choices.
- Clients select contraceptive methods according to the conventional definition of "informed choice," but providers do not try to assess whether the decision is based on an accurate understanding of the information provided (Y. M. Kim, Odallo, et al., 1997; Y. M. Kim & Kols, 1996).
- Clients can be taught to take a more active role in the interaction and consequently increase the quality of information they receive and their own satisfaction.

Client perspectives. The client perspective on quality of counseling is important if the client is expected to change her behavior as a result. Exit interviews and focus-group discussions with clients have permitted Population Communication Services to incorporate client values into efforts to upgrade client-provider interaction and counseling. According to client interviews in Nigeria, nurses trained in counseling skills were more likely than untrained nurses to listen attentively to their clients, to make them feel comfortable, and to treat them politely (Y. M. Kim, Rimon, et al., 1992). In the Central Asian Republics, focus-group discussions with men and women revealed that service providers were perceived as unreliable sources of information and as lacking concern for the clients' welfare (Storey, Ilkhamov, & Saksvig, 1997). In Ghana, a study of service quality involved a follow-

up of contraceptive acceptors who did not return to the clinic, asking them directly why they did not return (Tweedie, 1995).

Impact evaluation. The most important question is: Does improved counseling have an impact on clients' behavior, and specifically on behavior that reduces unwanted fertility? Indicators now being tested in Kenya include decision to use a method, selection of an appropriate method for the client's reported characteristics, method continuation, compliance with effective use procedures, timely return visit for resupply, and method switching that avoids the risk of pregnancy. The study uses clinic records to follow clients over a 42-month period (Y. M. Kim & Kols, 1996). Because of the nature of the data, specialized statistical routines of longitudinal regression analysis will be used, including general equation estimation (Diggle, Liang, & Zeger, 1994). In Nigeria, the indicator of impact, collected from client records, was the timely return of new clients for the first follow-up appointment. The clients who had been attended by nurses trained in counseling were more likely to return to clinics than those attended by untrained nurses (Y. M. Kim, Rimon, et al., 1992). This was one of the first evaluations to link provider training with subsequent client behavior.

Figure 7.10a
Nature of Clients' Active Participation in Counseling, Kenya

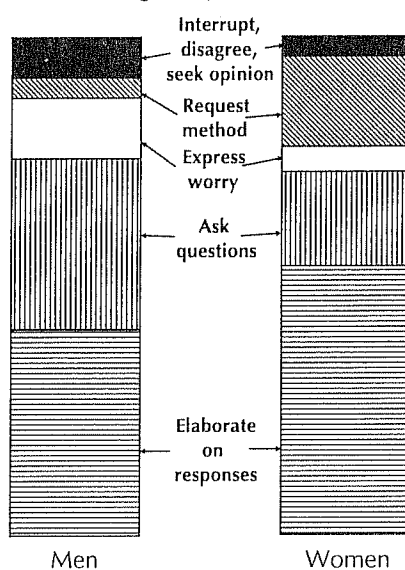
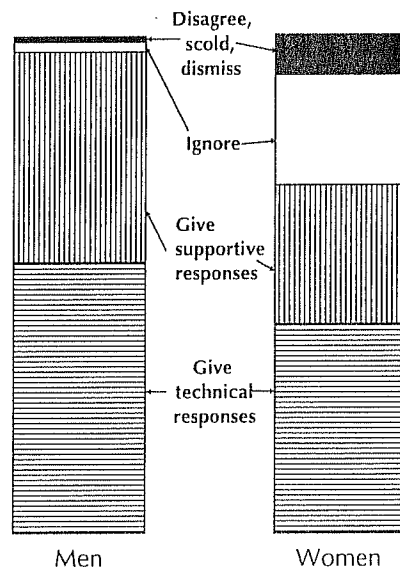


Figure 7.10b
Providers' Responses to Clients' Active Participation, Kenya



N = 65 men and 173 women.

Note: Based on transcript analysis.

Source: Y. M. Kim, Odallo, et al., 1997.

Methods for assessing the quality of client-provider interaction have been developed primarily with women, but they can be modified and applied to men, couples, or adolescents of either sex. In Kenya, for example, both observation and transcript analysis indicated the following:

- Male clients participate more actively than women in counseling sessions when they come alone for counseling.
- When both partners are present in counseling sessions, women become more passive than when they are alone.
- Providers give more supportive responses to men than to women.
- Providers ignore and disagree more with women than with men (Y. M. Kim & Awasum, 1996; Mwarogo, Kim, & Kols, 1996). As Figures 7.10a and 7.10b illustrate, when clients take an active role in the counseling session, men are more likely than women to ask questions but less likely to request a family planning method and less likely to elaborate on their responses to a provider's questions. For their part, the providers' responses to their clients' active participation also varied by the sex of the client. Providers were more likely to ignore women's statements and more likely to give men supportive responses and technical information.

In Zimbabwe, youth counseling sessions were evaluated using expert observation and exit interviews. They found the following:

- Young people, ages 12 to 24, were very uncomfortable in dealing with sexuality and reproductive health issues. Younger clients were more anxious, embarrassed, or shy, and giggled nervously instead of answering questions during the session. Providers had to be more patient in dealing with young clients.
- Young people were more passive than adults in counseling sessions. Providers had to make considerable efforts to obtain essential information or to determine the client's agenda for the session.
- Despite being passive or even nonresponsive during the sessions, many young people wanted the sessions to last longer and to provide more information, according to their exit interviews.
- Young people were concerned about a number of issues that were not related to reproductive health, such as school, dating, friendships, and alcohol (Y. M. Kim & Marangwanda, 1996).

Innovative research such as this makes it possible to understand better the impact of provider interaction interventions on client behavior; to provide specific guidelines to improve the provider training curriculum, materials, and audiovisual aids; to develop client education at service facilities and campaigns that can increase clients' participation during counseling; to improve the methodologies for studying client-provider interaction; and, above all, to improve the services provided.

COLLABORATION AND DISSEMINATION OF FINDINGS

LESSON 12.

Collaborating with other organizations for program evaluation can provide substantial benefits.

Just as effective health programs require a broad base of collaboration, evaluation also benefits from collaboration in the following ways:

- reducing costs by avoiding duplication in design and data collection;
- reducing the time required for design and data collection;
- creating opportunities for multiple points of view and different analyses;
- increasing utilization of results by involving implementing organizations in the evaluation process; and
- increasing credibility, public scrutiny, and application of evaluation results to future programs.

Collaboration can vary considerably. Collaborating institutions can include a wide variety of agencies, from government statistical offices to market research firms to university faculty. Different mechanisms can be used to plan this collaboration, including workshops and seminars, technical advisory groups, periodic updates for collaborators, and circulation of materials for review and comment. While extra effort is usually involved in ensuring appropriate timing, comparable samples, or satisfactory questions, generally the advantages of collaboration outweigh the disadvantages. In practice, collaboration can take place at four stages of evaluation implementation: research design, data collection, analysis, and utilization of data.

Research design. Communication program evaluations should incorporate a broad base of institutional and expert participation in the research design and in the development of research instruments. This collaboration can help to ensure a common conceptualization of the problems, the theoretical model, the objectives, and the subobjectives and to win full support for the evaluation. In Zambia, for example, an Information, Education, and Communication Subcommittee of the Government's Intertechnical Committee on Population helped to review research designs and two sets of survey protocols. The committee provided high-level policy review of the evaluation. Later, based on the evaluation results, the committee helped to develop a set of program recommendations that became guidelines for future program implementation.

Data collection. Since data collection is often the most expensive, time-consuming, labor-intensive, and quality-sensitive component of evaluation, collaboration at this stage is especially useful. It can include sharing technical staff, field operations resources (vehicles, field workers, and area managers), organizational resources (sample frames, questionnaires, computers and software, accounting and financial systems), and, of course, costs. In Zimbabwe, the primary

survey for a new youth communication initiative was funded by the German Agency for Technical Development (GTZ). The design and analysis were guided by Population Communication Services. The Zimbabwe National Family Planning Council carried out the program.

A powerful and cost-effective way to collaborate in data collection is by participation in country-specific Demographic and Health Surveys (DHS). At the request of Population Communication Services, several countries have incorporated additional communication questions into their surveys to establish baselines for communication programs. In Tanzania, for example, the 1995 Demographic and Health Survey included questions on recognition and comprehension of the national family planning logo and a radio social drama with family planning messages (see Lesson 8). The 1997 Demographic and Health Survey in Jordan includes a survey of men, at the request of the government, to facilitate evaluation of a planned communication campaign to promote men's participation in family planning. Collaboration between Population Communication Services and the Demographic and Health Surveys was formalized when PCS developed a detailed communication module for optional inclusion in future surveys. PCS also collaborates with the Demographic and Health Surveys to improve the analysis of communication issues using survey data.

Communication researchers also can include questions on recurring consumer surveys, generally called omnibus surveys. Market research firms put together questions from a number of clients who then pay for the answers to their specific questions. Firms will perform simple bivariate analysis of clients' questions in relation to common background variables. Because such surveys usually are carried out quarterly, they can provide data periodically. While the audiences might not be precisely those intended and the survey quality might vary, omnibus survey questions are relatively inexpensive. They can be used to augment and help interpret data from other sources. In Kenya, Nigeria, and Zimbabwe, omnibus surveys have been used to measure national exposure to communication interventions, such as radio programs or logos.

Analysis. Collaborating with other agencies in data analysis can provide valuable opportunities for developing-country researchers as well as ensuring that evaluation results will become known to host-country counterparts. The Zimbabwe National Family Planning Council, one of the few family planning associations with a trained Research and Evaluation unit, often collaborates with Population Communication Services in analyzing Zimbabwe data.

Conversely, Population Communication Services also has collaborated in analyzing data collected by other organizations. For example, in Ghana the Population Council carried out a Situation Analysis, observing family planning services at different facilities. The data were then analyzed by the Population Council, the Ministry of Health, and Population Communication Services to assess clinic education and counseling activities, to improve training, and to support advocacy for an expanded role for communication in the Health Education Unit of the Ministry of Health.

Utilization. One of the most cost-effective forms of collaboration is to allow other researchers to use data for additional secondary analysis. This opens the way for new perspectives, peer review of work, and application in broader and perhaps unanticipated program areas. To facilitate additional secondary analysis and utilization, Population Communication Services has established an archive of quantitative and qualitative data sets from its program evaluations. The archive maintains copies of data collection protocols, reports, and analysis as well as the computer files of the actual data, searchable by region, country, study population, date of study, and the variables covered by the study. Approximately three dozen data sets are currently available for independent analysis by other researchers, students, donors, and program planners. Some of these have been used for doctoral dissertations at Johns Hopkins University and elsewhere.

LESSON 13.

Evaluation results can be presented at various levels of complexity to suit different audiences.

Like any other message, evaluation results are useless unless they are communicated, understood, appreciated, and applied by the intended audiences. The audiences for evaluation are usually policy-makers, donors, program managers, and eventually other communication and social science researchers and evaluators.

At the policy level, the challenge in presenting evaluation results is to make a convincing case that the communication program changed behavior without overwhelming officials with complex statistics. Policy-makers are usually very busy, interested mainly in “the big picture” and the implications for future programs, and expert in areas other than statistics. Moreover, they always want to see evaluation results as soon as possible.

Researchers and academic experts, in contrast, want to be sure that evaluation methodology, theory, and analysis are sound. They will apply various tests for causal inference. They will make comparisons with evaluations they have undertaken themselves. And, in the always lively spirit of intellectual debate, they will look for points to criticize.

To meet the demands of such diverse audiences, evaluation results need to be presented in at least three different ways.

Key findings need to report major program results in a timely way. These reports need to summarize briefly the program objectives and subobjectives, the basic program strategies, the messages and materials produced and distributed, and the nature of the evaluation, as well as its findings. Statistics should be simple and clear. Graphics might consist of a single line graph, a bar graph, or a pie chart to illustrate a main conclusion. The format should lend itself to presentations at official meetings or workshops, to press releases, and to notices on the Internet. The PCS Key Findings Series began in 1997.

Field reports need to describe background and program issues as well as reporting in more detail on findings and results. Ranging from about 20 to 60 pages,

field reports on major projects can be illustrated and printed. A field report should serve as a guide to program managers, highlighting what happened, as well as what was effective and what was not, and recommending improvements for the future. When time and space allow, a field report should provide a comprehensive record for future reference including relevant data tables. To date, PCS has issued 10 formal field reports and several dozen informal ones. These are listed in POPLINE, the computerized database, and single copies are usually available on request.

Journal articles, presentations at professional meetings, or book chapters usually require the presentation of complex statistical analysis. Several articles may focus on different issues within a single comprehensive program evaluation. In fact, journal articles and presentations rarely can present a comprehensive picture of a major program and its impacts. Instead, they often present in-depth analysis of a particularly interesting aspect of a program.

During the process leading from key findings to journal articles, the type and complexity of statistical analysis presented usually increases. Within this chapter alone, for example, the type of statistical analysis has progressed from comparisons of mean village contraceptive rates in percentages, to simple bar graphs, crude odds ratios, and scatterplots with linear regression, to multiple logistical regression, conditional multinomial regression with panel data, Poisson regression with longitudinal service statistics, and path analysis. The simpler methods—linear regression, percentages, and crude odds ratios—are easier to present and to understand; the more complex methods are necessary to compensate for alternative sources of influence and selectivity that are inherent with one-group research designs and to elaborate the causal pathways.

A close look shows, however, that even complex analyses can be presented in simple ways—that is, through odds ratios and bar graphs. Figure 7.5 (see page 155) on Kenya, for example, illustrates the complex relationship among three different sets of variables—modern contraceptive use, exposure to communication, and intervening steps to behavior change—in three fairly simple bar graphs. Meeting the increasing demand for rigorous and relevant evaluation of communication programs requires presenting even the most complex statistical analysis in a form that is accessible to everyone.

NOTES

1. In Bangladesh, a *thana* is an administrative subdivision equivalent to a U.S. county.
2. The terms, “control variables” and “program variables” are used to remain consistent with the terminology used in the rest of this chapter. In structured equation modeling terminology the variables on the left-hand side measured in 1994 are referred to as “exogenous” variables, and the variables on the right-hand side measured in 1996 are referred to as “endogenous” variables.
3. The EQS analysis provides several indicators of the fit of the path model to the data (lower right-hand box). The comparative fit index (CFI) is 1.000, indicating an almost perfect fit of the model to the data, and the Chi² goodness of fit test indicates that there is no statistically significant difference between the model and the data. The average correlation

among the residuals of each variable is 0.021, indicating that there are no variables left out of the model that might still be determining any of the relationships shown in the model. The covariances among the exogenous variables measured in 1994 are omitted for ease of presentation, as are the residual terms, E , for the three endogenous variables. The amount of variance in family planning use and intention that is explained by the path model is 36 percent, so residual coefficient was .80. Thirty-one percent of the variance in ideation (5 SBC) is explained by the model, so the residual coefficient was .83, and 14 percent of the variance in *jiggasha*/FWA communication is explained by the model, so its residual coefficient is .93.