Evaluating the Impact of Communication Programs

Summary of an Expert Meeting Organized by the MEASURE *Evaluation* Project and the Population Communication Services Project

October 4-5, 2001
Belmont Conference Center
Elkridge, MD

MEASURE Evaluation
Carolina Population Center
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 304
Chapel Hill, North Carolina 27516, USA

Population Communication Services
Johns Hopkins University
Center for Communication Programs
111 Market Place, Suite 310
Baltimore, Maryland 21202, USA

Prepared by: Maria Elena Figueroa, Jane T. Bertrand, and D. Lawrence Kincaid

Population Communication Services

Johns Hopkins University
Center for Communication Programs

August 2002

Other Workshop Summaries by MEASURE *Evaluation*

No. 1 HIV, STDs and Infertility: Key Issues, May 1999.

No. 2 Use of Birth Registers as a Data Source for Maternal and Perinatal Health Care, May 1999

No. 3 Review of Results of the Multi-Country Field Test of Quality of Care Indicators in Clinic-Based Family Planning Programs, July 1999

No. 4 Measuring Capacity Building in Health and Population Programs, March 2000.

No. 5 Biological and Clinical Data Collection in Population Surveys in Less Developed Countries and Annotated Bibliography, March 2000.

No. 6 Strengthening Monitoring and Evaluation of National AIDS Programmes in the Context of the Expanded Response, April 2001.

No. 7 Measuring Maternal Mortality from a Census: Potential Uses in Latin American Countries, October 2000.

The workshop summary series is made possible by support from USAID under the terms of Cooperative Agreement HRN-A-00-97-00018-00. The opinions expressed are those of the authors, and do not necessarily reflect the views of USAID.

Table of Contents

List of Boxes

List of Figures

Introduction

On October 4-5, 2001, the Population Communication Services Project (PCS), from the Johns Hopkins University Center for Communication Programs (JHU/CCP), and the MEASURE *Evaluation* Project (M2) convened an expert meeting to discuss methodological issues regarding the evaluation of large-scale communication programs. The overall objective of the meeting was to agree on the preferred methodologies for evaluating the impact of communication programs conducted as part of wide-scale health interventions. The specific objectives were as follows:

1. To identify the major methodological challenges to evaluating the impact of full-coverage communication programs for behavior change;

2. To review study designs that leading researchers have used to address the threats to validity in the context of evaluating communication programs;

3. To identify (a) means to further improve on existing methodological approaches, and (b) appropriate indicators for impact assessment; and

4. To strengthen the network of researchers and evaluators involved in evaluating communication programs.

The 45 participants fell in one of three categories: (1) program evaluation specialists, including communication researchers, (2) professional staff from donor agencies that fund the evaluation of Behavior Change Communication (BCC) programs, and (3) program specialists in the implementation of BCC programs. The organizers assumed that participants have a basic understanding of key methodological issues in communication research; thus, the program focused on relatively advanced evaluation methods and concepts.

The meeting consisted of a series of sessions in which researchers presented alternative study designs they had used to assess

the impact of BCC programs in the United States and in selected developing countries. Actual findings were secondary in importance to discussions of the strengths and limitations of each design. The informal atmosphere of the meeting allowed for an easy exchange of opinions.

This report synthesizes the evolution of BCC programs, the basic methodological issues involved in measuring their impact, and the discussions that emerged from the series of presentations of alternative study designs used in different settings. It concludes with several points of consensus emerging from the meeting, as well as recommendations for future research in this area. Appendix A lists all participants of the meeting, and Appendix B provides the meeting agenda.

Evolution of BCC Programs

Over the last 30 years, health communication programs have evolved substantially from largely ad hoc, isolated educational efforts to a more comprehensive approach that treats communities and individuals as participants and as consumers. Jose Rimon II described this evolution in relation to family planning and reproductive health:

> **The Clinic Era, characterized by clinic-based Information-Education-Communication (IEC).** During the 1960s and 1970s, many country programs assumed a philosophy of "build and they will come." Family planning services were often medically oriented, and communication played a limited role, largely as support to clinic-based providers. Communication was largely a monologue, in which the medical providers (experts) directed messages to clients (non-experts).
>
> **The Field Era, characterized by outreach initiatives to serve people outside the clinics.** In the 1970s and early 1980s, an approach analogous to agricultural extension attempted to reach beyond the clinics, either to attract clients to the services or to provide services where they lived. Programs produced large quantities of IEC materials, films, and

audio-visuals to support these initiatives. The dominant paradigm during the field era was SMCR (source, message, channel, receiver), though communication moved from monologue to dialogue.

The Strategic Era, characterized by multi-channel integration and increased use of electronic media (consistent with the expansion of radio and TV receivers throughout the developing world). Since the mid 1980s, program personnel and donor agencies have increased attention on evidence-based programming, audience segmentation, and program evaluation. The process of communication further evolved from dialogue to convergence, in which participants in the communication process share information, making the distinction between "senders" and "receivers" elusive.

Evolution in the Evaluation of Communication Programs

As programs grew in number and complexity, so did interest in knowing whether or not they effectively achieved behavioral change. The Johns Hopkins Center for Communication Programs (JHU/CCP) was one of the first organizations to systematically evaluate its communication programs, and Maria Elena Figueroa traced this history to allow participants to appreciate the current situation. Although Figueroa's presentation outlined the experience of CCP, it reflected the methodological challenges facing other organizations attempting to measure the effects of full-coverage health communication programs on behavior change. Following this overview, the rest of the meeting consisted of presentations illustrating the range of designs evaluators had used to measure the effects of communication programs in the U.S. and in developing countries.

Evaluation of communication programs was fairly limited during the 1970s and 1980s, with the primary effort directed toward program implementation. Evaluation became more widespread in the 1980s, and the design of choice was often the post-test only with no control group, especially for programs with a mass media component. Over time, frustration stemming from the

desire for more comprehensive results motivated a number of programs to incorporate evaluation from the start. By the 1990s, the pre- and post-test design (with or without a control or comparison group) came into widespread use. Although evaluators recognized that the experimental design (e.g., the pretest - post-test control group design with randomization of subjects) provides the most compelling evidence of program effectiveness, it was (and is) impractical for evaluating programs with a radio or TV component for several reasons:

(1) impossibility of randomly assigning individuals or other units, such as villages or regions, to the experimental groups,
(2) difficulty in getting a comparable population to serve as a control group,
(3) contamination of the communication intervention into the control areas when they are used, and
(4) intervening events ("history") unrelated to the intervention that altered the outcome, especially when programs last for several months or years.

In addition, the classic experimental design demonstrates change, but alone it provides little insight into the process by which communication influences behavior.

Because the classic experimental design has proven impractical in most full-coverage programs (i.e., those designed to reach the entire population of a country or region), evaluators have used statistical controls during data analysis that allow them to reduce (though not eliminate) the influence of confounding factors.

Throughout the 1990s, evaluation of communication programs at CCP tended to rely on a pretest-post test design with statistical controls to: (1) assess the effects of the program and (2) identify the causal pathways through which the program worked, as a means of informing the design of subsequent programs, and to test and refine the theories used to explain behavior change.

Current evaluation designs improve on earlier ones because they seek to trace the intervening processes through which communication affects behavior, as well as the effect of mutually reinforcing messages from different media. Moreover, the designs measure the extent of self-reported exposure to different communication channels and messages to establish a "dose effect." Even with the use of statistical controls in the pretest - posttest one-group design, threats to validity, such as potential self-selection[1] and endogeneity,[2] persist. In an effort to control for these threats, CCP has used a variety of designs and methods, including interrupted time series analyses complemented by reports on sources of referral (see Kincaid et al., 1996), longitudinal designs using panel data with lagged variables, as well as simultaneous equations systems and path models. Using accumulated evidence and theory, CCP developed eight criteria to guide its designs and to strengthen the claim that the observed change is attributable to the communication intervention (see Box 1). The more of these criteria the evaluation of a program fulfills, the stronger the claim on the effect of the program.

In its continual learning about the processes through which communication affects behavior, CCP has used network analysis and found an "indirect exposure"[3] effect to the communication intervention. Network analysis provides a framework for identifying and accounting for indirect exposure, and therefore establishes the indirect effects of communication interventions that occur through interpersonal communication. For example, findings from the Nepal study by JHU/CCP (see Storey et al.,

[1] Self-selection: The phenomenon occurs when exposure to the messages of a communication program is not random among individuals; rather exposure to the program results from the individual's voluntary choice which is influenced by a variety of factors, some measured and some unobserved.

[2] Endogeneity: This problem generally refers to the direction of the causality between exposure to the program and the observed outcome. The problem arises when the researcher cannot determine whether exposure to the program preceded the observed outcome/behavior, or the expected outcome/behavior influenced exposure to the program.

[3] Indirect exposure: This concept refers to exposure through secondary sources, and not simply through program activities. The two-step flow hypothesis or multi-stage communication process in diffusion research guides this concept (Katz and Lazarsfeld, 1955; Rogers, 1995).

1999) show that individuals exposed through secondary sources were as likely as those directly exposed to perform the desired behavior. Lessons from this program suggest that, when evaluators ignore exposure through secondary sources, they may underestimate the effects of the program.

Evaluation Research Designs

The underlying criteria of true experimental designs rest in random assignment of units to the intervention and in the existence of a counterfactual condition, "what would have happened." These criteria continue to constitute the "ideal" against which program evaluators tend to assess their own designs. Yet efforts to conduct experimental studies at the field level have yielded few successes for the following reasons:

(1) the impossibility of establishing a comparison group similar to the treatment group on all relevant dimensions, which would require random assignment of units; and
(2) the difficulty, if not impossibility, of withholding the program from any segment of the population, especially when the program uses wide-coverage mass media channels.

Few evaluators have the luxury of designing an evaluation with the necessary resources and authorization to use multiple designs to control for all potential threats to validity (one possible exception among the presentations being the evaluation of the national drug abuse program, presented by Robert Hornik). Rather, depending on the interests of the program, the stage at which they are "brought in," and the resources available, evaluators have devised multiple approaches to assessing the impact of communication programs, given the constraints of different situations. The organizers of the meeting solicited some of the leading experts in the field and gave them the choice of study design on which to present, in anticipation that each would present a viable methodology for evaluating a full-coverage program, despite the inevitable constraints. This approach proved very useful, in that all presenters were very familiar with the specifics of their study and brought to light the reasons for using the design in question, its strengths and limitations. To reiterate, the

presentations focused on methodological issues; speakers presented results for the purpose of enriching the methodological discussion of design.

Cross-Sectional Surveys

Charles Westoff, beginning the presentation of specific designs, cited the potential of cross-sectional surveys for demonstrating impact of communication programs, as illustrated by his research on family planning communication using the Demographic and Health Surveys.[4] He used repeated independent cross-sectional sample surveys to demonstrate program effect. He found a consistent positive association between campaign exposure and family planning use in all countries, an association he could not remove by statistical controls (see conditions 4 and 8 in Box 1). Westoff concluded that the association could not be "accidental," but rather was the direct result of the communication intervention.

Robert Magnani presented findings of an evaluation of an adolescent reproductive health communication intervention in Paraguay. The evaluation measured impact on the basis of the observation of changes in key behavioral outcome indicators over the 2.5-year study period and significant dose-response relationships between measures of program exposure and outcome indicators in the follow-up survey data. The most prominent threat to validity in this study was potential self-selection of participants. To address this threat, Magnani and colleagues used a Heckman-like selection test; specifically, they used a bivariate probit model to jointly estimate the selection and outcome equations and to calculate the ρ (rho) statistic, which tests the null hypothesis of no selection ($\rho=0$). Magnani also clearly illustrated the difficulties of finding appropriate "identification variables,"[5]

[4] Demographic and Health Surveys are nationally representative household surveys with large sample sizes. DHS surveys provide data for a wide range of evaluation indicators in the areas of population, health, and nutrition (MEASURE, *DHS*+, ORC Macro).

[5] The identification problem is a mathematical (as opposed to statistical) problem associated with simultaneous equation systems. It is concerned with the possibility of distinguishing each equation in the system from the rest, in the

which are needed to estimate a system of equations such as the bivariate model. In a later panel session, Guilkey, however, expressed concerns about the appropriateness of the identification variables used in the study.

Ruth Bessinger used a community aggregate index of exposure to evaluate the effect of mass media exposure on knowledge and use of condoms in the Delivery of Improved Services for Health (DISH) project in Uganda. To compute this community measure, the research team aggregated the individual level exposure responses at the cluster level. They proposed this measure as an alternative that would solve the potential bias for endogeneity that may occur when one uses cross-sectional data and self-reported measures of exposure to the program. This approach, however, came under question from Moffitt, because it relied on the same self-reported individual data of exposure.

Longitudinal Panel Designs

Robert Hornik revealed how rich an evaluation design can be when planned from the beginning and supported with adequate funding. Although the design did not include the randomization of individuals to the intervention, it provided different mechanisms to strengthen the validity of the results. The ongoing evaluation of the National Youth Anti-drug Media Campaign (NYAMC) consists of a cross-sectional study, which then becomes a cohort sample survey.

For the cross-sectional analysis, Hornik used the propensity score[6] method to resolve the potential problem of self-selection.

---

sense that each equation includes and excludes the same variables. To obtain consistent estimates, one must identify each equation in the system. One way to identify each equation is to exclude, on theoretical grounds, at least one variable from each equation. (See Kennedy, 1996; Bollen et al., 1995.)

[6] The propensity score is defined as the conditional probability of participating in the program (being exposed) given a series of observable variables (Rosenbaum and Rubin, 1983). Analysts have used the propensity score in at least two different ways to assess program impact. One application has been through the computation of the actual score, as Hornik used it in his presentation. The second application, "propensity score matching," uses statistical methods to

To apply the method, Hornik used regression techniques to provide the actual estimation of the probability of exposure to the program (the propensity score). Using all available socio-demographic and other causally prior variables in the data set, the propensity score is the predicted value obtained from the regression estimation. In one approach to the analysis, the propensity score is divided into quintiles. Within each quintile, statistical associations between the outcome of interest and the reported exposure to the program are computed and assessed for statistical significance. In this way, the effect of exposure on the outcome is controlled by the propensity of exposure to the program. An additional advantage of this method compared to more typical multiple regression approaches is that it eliminates the need to fit separate equations for each outcome of interest.

The cohort sample study, currently in the data collection stage, will include three types of analysis: (1) a change analysis, (2) a lagged individual analysis, and (3) an analysis using a multi-wave average exposure score as the predictor. If appropriate, additional analyses will be undertaken at the geographic sampling unit level of aggregation. The contribution of the cohort analysis to the validity of the evaluation is that the lagged analysis method addresses the problem of time-order of causal influences (see condition 3 of 8 criteria in Box 1), providing evidence as to whether exposure preceded attitude or behavior changes. In addition, the lagged aggregate analysis will allow testing for other routes of effect besides the individual.

John P. Pierce presented the evaluation design of the California Tobacco Control Program, which includes large representative surveys of the non-institutionalized civilian population every three years.[7] These random digit dialed California Tobacco Surveys (CTS) provide excellent data for trends in smoking attitudes and behaviors. Results indicate that there is consensus on many

approximate a control group. (Rosenbaum and Rubin, 1985; Heckman et al., 1996, Lu et al., 2001.)

[7] These surveys have the power to detect change in smoking prevalence for 18 mutually exclusive regions of the state. Surveys and data are available on the website http://ssdc.ucsd.edu/tobacco

of the predictors of adolescent smoking. Accordingly, these surveys provide trend data for changes in the prevalence of these "sentinel indicators"[8] of future behavior. However, the Tobacco Control Program does not only focus on changing the prevalence of the "sentinel indicators" (for example, the number of best friends who smoke). The program also tries to reduce the influence of this predictor on adolescent smoking. Cross-sectional surveys cannot address trends in the relative influence of these predictors. Rolling longitudinal surveys (re-interview surveys following each cross-sectional survey) are needed to identify trends in the experimentation rate for different levels of a predictor.

There are a number of threats to validity of telephone surveys. For example, the attrition rate, and the impact of increasing use of call screening and blocking across the United States, create a problem of self-selection. To control for these threats, the research team has used data on tobacco use collected as a supplement to the Current Population Survey[9] (CPS), on the same three-year rotation as the CTS. The CPS is a national household based survey and thus allows an estimate of any bias in adult smoking measures associated with change in attrition on telephone surveys. The researchers used weighted data to account for non-response and selection probability, as well as to adjust the sample to be representative of the population on major socio-economic variables.

Experimental and Quasi-Experimental Research Designs

The organizers had hoped to identify speakers who could present examples of evaluations of mass media programs based on an experimental design with random allocation of subjects. They were unsuccessful in identifying such researchers or studies. Rather, the presenters spoke on evaluations that used quasi-

[8] Sentinel indicators are intermediate variables that predict behavior, in this case adolescent smoking.

[9] The Current Population Survey, a national survey conducted by the Bureau of the Census, includes over 10,000 Californians and is household based.

experimental designs (pretest - posttest with a comparison group).

In Everett Rogers' evaluation of a Tanzania radio drama entitled *Twende na Wakati,* he raised a key evaluation design question: Are randomized experimental designs "the ideal" for evaluating the effects of health communication interventions? He summarized the characteristics of this type of design (see Box 2) and presented the problems encountered in the Tanzania study. The primary difficulties included a non-comparable control group, contamination of the control group, lack of a full understanding of how effects occur, and the fact that results cannot be scaled up. This latter problem refers to the well-known threat of "external validity"[10] of all small-scale experiments. Rogers concluded that whereas field experiments are strong in internal validity,[11] and therefore useful in evaluating intervention effects, they carry their own set of limitations.

Steve Booth-Butterfield presented a second example of a controlled field experiment, the "1% or less milk" program in Wheeling, West Virginia, which used a quasi-experimental design with one intervention city and one comparison city. The researchers assessed the effects of the program through pre- and post-telephone surveys, as well as sales data on milk purchases from supermarkets in both communities. The two sources of data yielded consistent results: self-reported consumption of 1 percent or less milk during telephone interviews and supermarket sales of the product were significantly higher in the intervention city than in the comparison city after the campaign. The evaluators also found a statistically significant association between the different combinations of media delivery and self-reported exposure levels that weakened the potential problem of self-selection. Likewise, they found that higher exposure was associated with greater changes in attitudes and intention, which led to behavior change (see conditions 6 and 7 of eight criteria in Box 1).

[10] External validity refers to the degree to which research methods allow the results of a study to be generalized from a sample to the population at large.

[11] Internal validity can be defined as the degree to which research methods yield conclusions that are free of alternative explanations.

As with the Rogers study, the primary difficulty in this study was a non-comparable comparison group. The researchers found that respondents in the intervention community consumed more low-fat milk at baseline than did those in the comparison community. Another finding of the study was that respondents lost to follow-up were significantly more likely to live in households that used high-fat milk than those who completed both the pre- and post-intervention surveys. With respect to sales data, meeting participants suggested a potential overstated effect in the observed rise in 1 percent or less milk sales due to seasonality of milk sales (i.e., differential consumption of milk based on weather or other factors). Whereas this study was easier to implement than the Tanzania evaluation, the researchers still had to deal with potential threats due to differences in the intervention and comparison communities and the inability to follow-up on some respondents.

Meta-Analysis Studies

Leslie Snyder conducted a meta-analysis of 49 US-based campaigns and examined the potential influence of the evaluation design on the average effect size. Meta-analysis studies have several advantages over individual analyses: (1) they have higher power to detect small effect sizes across studies, (2) the methodological problems of any one study have minimal effect in biasing the overall conclusions, and (3) conclusions from a meta-analysis are more generalizable than the results from a single study.

The meta-analysis indicated that the average effect size of US-based campaigns is 9 percent, which can be interpreted as 9 percent more people performing the desired behavior as a result of the campaign. Findings from the bivariate analysis suggested statistical associations between evaluation design and effect size. Designs in which the sample is partitioned by exposure with no comparison group had the greatest effect size (.12); this finding thus emphasizes the importance of measuring exposure. Quasi-experimental designs examining persuasive campaigns (.05), pre-post designs with no comparison group (0.06), and post test only with comparison group (0.02) resulted in lower average

effect sizes. Other important factors related to campaign effect sizes were using a random rather than a self-selected sample, attrition, the time lag between the end of the campaign and the post test, equivalency between the comparison and intervention groups, and trends over time in the comparison group. Thus, the bivariate analysis suggests that the research design influences the ability to detect communication effects when they are present. Design limitations or problems detected when the research is underway can produce small but important differences in the effects attributed to campaigns. The meta-analysis is ongoing and will test the effects of different factors simultaneously when there are data on a sufficient number of campaigns.

Theory-Driven Outcome Evaluations

Two presenters addressed the role of theory in evaluating programs. Huey T. Chen, presenting a conceptual framework for Theory-Driven Outcome Evaluations, examined the traditional neglect of theory in the discipline of program evaluation.[12] He reasoned that the atheoretical approach to program evaluation is characterized by a primary focus on the relationship between the input (intervention) and output (outcome) of a program. This type of evaluation, which he called the Basic Model of Outcome Evaluation (see Figure 1a), is characteristic of the experimental paradigm. And although objective and rigorous, this approach is method-driven, it fails to identify the underlying causal mechanisms and contextual programmatic elements that generate the treatment effects, and thus renders the findings useless for future program improvement and development.

A theory-driven outcome evaluation explores the leverage or cause of a problem that mediates between an intervention and an outcome. Chen named this leverage the determinant, and further elaborated on the role the determinant plays in the framework. He defined the "mediating mechanism evaluation" (see Figure

[12] Theory is a frame of reference that helps humans to understand their world and to function in it. Theory provides not only guidelines for analyzing a phenomenon but also a scheme for understanding the significance of research findings (Chen, 1990).

1b) as an outcome evaluation that *assesses the relationships among intervention, determinants, and outcomes*. Sources of determinants are scientific theory and stakeholders' implicit theories. The usefulness of the mediating mechanism evaluation is that it provides information for program improvement, it is future-oriented, and it enhances both construct and internal validity. To conduct this type of (future-oriented) evaluation requires that the evaluation be an integral part of the program.

Finally, Chen also introduced the "implementation environment" into the model and defined it as the pertinent environmental factors that support implementation and have potential to interact with the intervention to produce effects. Thus, the overall framework of theory-driven outcome evaluations is characterized by five elements: (1) the intervention, (2) the outcomes, (3) disturbances that affect both the intervention and the outcomes, (4) determinants, and (5) an implementation environment (see Figure 1c).

Larry Kincaid presented a general model of communication, ideation, and behavior change, which can be used to conduct theory-based evaluation of communication programs (see Figure 2). He then illustrated this approach with empirical data from the evaluation of a mass media communication campaign to promote contraceptive use in the Philippines. To show how communication affects behavior, the model specifies the *multiple causal pathways* from exposure through ideation, skills, and environmental support and constraints to intention and behavior. Ideation is a composite factor comprising all of the cognitive, emotional, and social interaction intermediate effects of communication that may influence behavior. The model postulates that people do not ordinarily act until they have sufficient knowledge about the behavior and its consequences, until they have a positive attitude towards it, until they have talked to others about it, and until they feel right about doing it. The validity of the causal inference that exposure is responsible for behavior change is enhanced when the communication itself is designed to change these intervening (ideational) influences, and when research obtains evidence that exposure to communication is related to

these intervening outcomes and that these intervening outcomes are related to the observed change in behavior.

He demonstrated this approach with a national, cross-sectional survey of 1,516 married women conducted after a 5-month television campaign. Use of a 15-month, modified form of the birth/contraceptive history chart from DHS questionnaires made it possible to treat month-by-month measures of contraceptive use as retrospective panel data and to conduct an interrupted time-series type of analysis. A statistically significant increase in modern contraceptive use was observed after the communication campaign, which was even higher among women with a high level of recall of campaign messages. To control for the potential confounding effects of other variables, multiple regression analysis was used for structural equation modeling. As hypothesized, communication recall had a significant direct effect on contraceptive adoption, as well as a significant indirect effect through its effect on the composite ideational variable (comprising contraceptive knowledge, attitudes, partner discussion and support, discussion and encouragement from friends, self-efficacy, and personal advocacy). Statistical tests for endogeneity confirmed that campaign recall and ideation may be considered exogenous to contraceptive use, and that campaign recall may also be considered exogenous to ideation. The main threats to validity in this study were the potential self-selection of participants and the adequacy of the instruments used to identify the three equations. The threat relating to self-selection (prior contraceptive use) was controlled statistically by means of the lagged measure from the retrospective panel data. Adequate instruments were available for communication recall and contraceptive use, but an additional instrument would be necessary to identify adequately the equation for ideation (see Guilkey and Moffitt below).

Two other presentations in the meeting included a theoretical framework of the pathways to behavior change (i.e., what Chen labeled a mediating mechanism evaluation model). First, Hornik's evaluation of the NYAMC program included a theoretical specification of the pathways to drug use, considering

individual and community level variables (see Figure 3). Second, Booth-Butterfield's evaluation of the low-fat milk campaign also included the theoretical pathways to milk consumption (see Figure 4).

Overcoming Threats to Internal Validity: Panel Discussion

Researchers/evaluators concur on the threats to validity in the evaluation of communication programs. The challenge is how best to address them. A panel of three researchers (two economists, one strategic planning analyst from the private sector) addressed different aspects of the problem. The two econometricians focused on the corrections for endogeneity and self-selection. Solutions to these problems generally require the use of independent or exogenous variables called instruments or instrumental variables,[13] as well as tests of the validity of exclusion (identification) restrictions.[14] In general, each right-hand-side endogenous variable needs to be explained by at least one, but preferably more, strictly exogenous variables that do not have direct effects on the outcome of interest. For example, if one is interested in identifying the effect of a respondent having heard a family planning message on the use of contraception, the instrumental variables approach would require that one have at least one variable that is a determinant of having heard a family planning message that does not have a direct effect on the use of contraception.

Guilkey's discussion focused primarily on the correct selection of instruments. This selection can be extremely difficult when one wants to evaluate the impact of communication programs on fertility-related outcomes, because it may be difficult to find variables in standard DHS data sets that affect whether or not a respondent recalls hearing the communication message that do

[13] An instrumental variable is defined as an independent variable that is highly correlated with the equation where it appears as regressor, but is not correlated with the other equation(s) in the system.

[14] A researcher should always conduct tests for the validity of the exclusion restrictions to assess the validity of his or her theoretical assumption for excluding a variable from an equation in the system (see Bollen et al., 1995).

not also affect outcomes such as contraceptive use. One possibility is community-level expenditure data for communication programs, but expenditure data are potentially endogenous if, for example, a country targets its programs to high fertility areas. Another possibility for identification could be whether or not a community lies within the broadcast area of a radio station.

Robert Moffitt further elaborated on the evaluation problem of assessing the effect of a program when random assignment of individuals is not possible. He used the classical two-equation model to explain the statistical problem of endogeneity and the role of instrumental variables in the solution, which are expressed as "z" in Box 3. Moffitt emphasized that for the equation of program exposure (equation 2 in Box 3), the appropriate instrument "z" would be an exogenous variable, such as local television reception, that could capture random variation in the "opportunity" to be exposed for each individual in the sample. He stated that the "z" variable could reflect a "natural experiment" in which individuals are randomly assigned to the intervention.

Regarding methods, Moffit mentioned that economists conducting program evaluation were moving away from panel designs and toward independent cross-sectional data focusing on variation of the outcome of interest over time. The line of reasoning behind this approach is that primary interest should center on how population averages change from one point in time to another when some intervention occurs, and for this reason it is not necessary to follow the same individuals. Moffitt opined that the studies presented in the meeting used appropriate methodologies for assessing program effect. He judged propensity scores as practical ways to deal with self-selection and endogeneity. Likewise, he pointed out that sensitivity analysis[15] could also assess the degree of the bias that endogeneity and self-selection may cause.

[15] Sensitivity analysis allows the researcher to assess the results of the evaluation with respect to unmeasured confounders. (See Rosenbaum and Rubin, 1983; Lin D.Y., et al., 1998.)

Jim Crimmins, a strategy analyst who evaluates media campaigns in the private sector, provided an expert view on communication program evaluation from a commercial marketing perspective. He focused on three elements: (1) improving validity, (2) improving usefulness of the findings, and (3) improving chances for program success. To improve validity, Crimmins recommended increased use of "routinely collected data," rather than self-reported data, to measure both exposure to the communication program and practice of the desired behaviors. Commercial market researchers in the U.S. rarely use self-reported data for either purpose. Measures considered more reliable include (1) the dollar amount of expenditures on advertising, which is preferred to "media impressions" or GRPs,[16] and (2) data generated by scanning bar codes at the time of purchase to track actual behavior.[17] Crimmins suggested that evaluators working in developing countries assess what data on target behaviors were routinely collected, what could be done to increase the amount of such data, and how to improve its reliability.

To improve usefulness, Crimmins suggested moving from "grading" to guidance: asking not only whether the program worked or not, but also how the evaluation could enhance it. In this regard, he agreed with Chen that evaluators should focus on "future oriented" evaluations and to include the evaluation as part of the program.

Finally, he offered three specific suggestions to improve the chances for success in a communication program. First, one should worry about what the audience takes away from commu-

[16] GRPs are gross rating points. If we reach 10% of the population with a particular vehicle, we have 10 rating points. Reaching 15% with another vehicle is equivalent to 15 rating points. If we sum the rating points without regard to duplication of audience, the sum is gross rating points, a common measure of advertising pressure.

[17] Due to time limitations, the group did not examine the relevance of these suggestions for the developing country context. Dollar amounts expended on a given campaign represents a potentially feasible measure, although the data could be biased by the extensive discounting or give-away to social programming. The source of data analogous to continuously available sales data would be service statistics, but these statistics are not always reliable in the developing world.

nication more than what goes in the communication. A message can communicate the opposite of what it literally says, or it can communicate much more than what it literally says. A message creates an impression of the behavior we would like to encourage. The impression may be more important than any factual information about the behavior. In creating the message, one should aim for the right impression. Second, messages should address the real person in charge. The mind is in charge, just not the conscious part of it. Our mind has 100,000 times more information processing capacity than our consciousness. Our consciousness doesn't even know why we do what we do. We can find out "why" by analyzing the connection between perceptions and behavior. Our messages should address the real reasons (motivations, perceptions) "why." Third, programs should seek change in society while seeking change in individuals. If the culture frowns on the use of condoms, for example, our efforts at persuading individuals to use condoms are probably doomed. We have to change the culture's view of condoms at the same time we attempt to change the individual's view to be successful.

Conclusions

Several conclusions emerged from this meeting:

1) **No perfect evaluation design exists**. In practice, all research designs entail some degree of internal and external threat to the validity of their conclusions, and the "perfect design" does not exist. Designs are often opportunistic to fit the circumstances of the program.

2) **Randomized experimental designs (the ideal) are not the optimal design for evaluating full-coverage programs**. The studies that used these designs proved difficult to implement in the field, suffered from lack of a comparable control group, experienced some degree of contamination, or required additional data and control variables. In general, the "ideal" design cannot be used for full-coverage communication programs in the field.

3) **Consensus appeared to exist on the eight criteria for establishing impact**. As their comments over the two-day period revealed, participants seemingly endorsed the eight criteria needed to claim impact (see Box 1), presented on the first day of the meeting.

Recommendations

1) **Use cross-sectional surveys**. Evaluators can use repeated independent cross-sectional surveys to demonstrate program impact. Westoff's presentation showed that four of the eight criteria for claiming impact (conditions 1, 2, 4, and 8 in Box 1) held true in his analysis of the effect of communication on family planning behavior. Researchers can benefit from the use of repeated surveys if they are able to include specific questions related to exposure to the communication program. The strength of cross-sectional surveys seemed supported by one of the panelists who mentioned that economists are moving toward repeated cross-sectional studies over panel designs.

2) **Accumulate evidence**. Evaluators should work to build evidence from a variety of data sources (triangulation) rather than relying on a single source for "definitive proof." Most of the case studies presented used a variety of mechanisms to build cumulative evidence of the effects of the program. The different data sources validate one another; moreover, studies may provide evidence of change in intermediate variables related to the behavioral outcome (condition 6 of the criteria in Box 1).

3) **Use theory to strengthen validity**. Evaluators should base evaluations on sound theoretical models that demonstrate, by means of intermediate variables, the causal pathways through which communication affects behavior. This recommendation specifically relates to condition 6 of the eight criteria.

4) **Make evaluations future-oriented.** Evaluators should improve usefulness of the evaluation by answering not only whether the program worked or not (grading), but also what can be done to improve it (guidance). In this sense, evaluators should focus on doing evaluations for future program improvement and development. For this purpose, it is important that the evaluation be part of the program.

5) **Use appropriate statistical techniques to counter threats to validity**. Evaluators should use statistical techniques to assess the extent to which self-selectivity and endogeneity threaten the validity of the evaluation results. Controlling for these problems when they are absent or minimal carries a severe penalty in the precision of the measurement of the effects of program variables.[18] When present, self-selectivity can be addressed through the use of lagged variables and statistical controls, as well as propensity score analysis.

[18] Controlling unnecessarily for endogeneity often leads to less reliable estimates than does ignoring small amounts of potential endogeneity. Even if the exogeneity test shows that endogeneity is a problem, one should not turn immediately to a two-step method. (For a broader discussion of the statistical requirements for a valid two-step approach see Bollen et al., 1995, pg. 128.)

Likewise, the evaluator can statistically assess the potential threat of endogeneity and can deal with it by instrumental variable analysis if the data contain adequate explanatory variables. In addition, the evaluator could conduct a sensitivity analysis to assess the results of the evaluation with respect to unmeasured confounders.

6) **Build in variables to use as instruments**. Before initiating the evaluation of a communication program, the evaluator should identify and include adequate variables in the data collection protocols to conduct instrumental variable analysis and propensity score analysis. These techniques, according to the econometricians in the panel, are good solutions for the problems of self-selection and endogeneity.

7) **Increase the use of routinely collected data**. When possible, the evaluator should make greater use of data that are regularly collected (such as clinic records) as another source to strengthen the claim of program effect. If the program includes the promotion of services, the evaluator can monitor the sources of referral for people attending the services and can use these data as additional evidence of the impact of the program.

8) **Measure "media intensity" directly, in addition to self-report.** The evaluator should explore and use, when possible, data related to local "media presence" (local program effort) to corroborate the self-reported measure of communication exposure (recall of the program's messages) as well as actual behavior. For example, in situations where broadcast frequency and reach vary by known geographical regions, one may use these data in the analyses of program effect. One should verify that the level of message recall by audience members corresponds to levels of media saturation planned by the program.

Relevant References to Subjects Covered in the Meeting

Evolution in the Evaluation of Communication Programs

Bertrand, J. T. and D.L. Kincaid, 1996. Evaluating Information, Education and Communication (IEC) Programs for Family Planning and Reproductive Health. Final Report of the IEC Working Group, the EVALUATION Project, Carolina Population Center, University of North Carolina at Chapel Hill.

Hill, A.B., ed. 1971. Statistical Evidence and Inference. *Principles of Medical Statistics*. New York: Oxford University Press.

Katz, Elihu and Paul Lazarsfeld, 1955. *Personal Influence: The Part Played by People in the Flow of Mass Communications*. New York: Free Press. (cited in Littlejohn, Stephen W. Theories of Human Communication, 7th edition, 2002. Belmont, CA: The Wadsworth Group.

Kincaid, D.L., A.P. Merritt, L. Nickerson, S.D. Buffington, M.P. de Castro, and B.M. de Castro, 1996. Impact of a mass media vasectomy promotion campaign in Brazil. *International Family Planning Perspectives* 22(4):169-75.

Piotrow, P.T., D.L. Kincaid, J.G. Rimon II, and W. Rinehart, 1997. *Health Communication: Lessons from Family Planning and Reproductive Health*. West Port, CT: Praeger Publishers.

Rogers, Everett M. 1995. *Diffusion of Innovations*. The Free Press.

Storey, J.D., M. Boulay, Y. Karki, K. Heckert, and D.M. Karmacharya, 1999. Impact of the integrated radio communication project in Nepal, 1994-1997. *Journal of Health Communication* 4(4):271-94.

Cross-Sectional Designs

Heckman, J.J., 1979. Sample selection bias as a specification error. *Econometrica* 47:153-161.

Kirby, D., 1999. Antecedents of Adolescent Sexual Risk-Taking, Pregnancy, and Childbearing: Implications for Research and Programs. Washington, DC: National Campaign to Prevent Teen Pregnancy.

Westoff, Charles F. and Akin Bankole, 1997. Mass Media and Reproductive Behavior in Africa. *Demographic and Health Surveys Analytical Reports No. 2*, Macro International Inc.

Westoff, Charles F. and Akin Bankole, 1999. Mass Media and Reproductive Behavior in Pakistan, India, and Bangladesh, *Demographic and Health Surveys Analytical Report No.10.* Macro International Inc.

Longitudinal Panel Designs

Bal, D.G., K.W. Kizer, P.G. Felton, H.N. Mozar, and D Neimeyer, 1990. Reducing tobacco consumption in California: development of a statewide antitobacco use campaign. *Journal of the American Medical Association* 264:1570-1574.

Gilpin, E.A., S.L. Emery, A.J. Farkas, J.M. Distefan, M.M. White, and J.P. Pierce, 2001. The California Tobacco Control Program: A Decade of Progress. *Results from the California Tobacco Survey, 1990-1999*. La Jolla, CA: University of California, San Diego (http://ssdc.ucsd.edu/tobacco/).

Heckman, J., H. Ichimura, and P. Todd, 1997. Matching as an Econometric Evaluation estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies* 64: 605-654.

Lu, B., E. Zanutto R. Hornik and P. Rosenbaum, 2001. Matching with Doses in an Observational Study of a Media Campaign

against Drug Abuse. *Journal of the American Statistical Association* 96:456.

Pierce, J.P., E.A. Gilpin, S.L. Emery, M.M. White, A.J. Farkas, B. Rosbrook, and C.C. Berry, 1998. Has the California Tobacco Control Program reduced smoking? *Journal of the American Medical Association* 280:893-899.

Rosenbaum, P. and D. Rubin, 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70:41-55.

Rosenbaum, P. and D. Rubin, 1985. Constructing a control group using Multivariate matched sampling methods that incorporate the Propensity Score. *The American Statistician*, 39:33-38.

Experimental and Quasi-Experimental Research Designs

Booth-Butterfield, S., and B. Reger, 2001. The message changes belief and the rest is theory: The 1% or less milk campaign and Reasoned Action. Paper presented at the International Communication Association meeting, Washington, D.C.

Reger, B., M.G. Wootan, and S. Booth-Butterfield, 2000. A comparison of different approaches to promote community-wide dietary change. *American Journal of Preventive Medicine* 18(4):271-5.

Reger, B., M.G. Wootan, and S. Booth-Butterfield, 1999. Using mass media to promote healthy eating: A community-based demonstration project. *Preventive Medicine* 29(5):414-21.

Rogers, Everett M., Peter W. Vaughan, Ramadhan M.A. Swalehe, Nagesh Rao, Peer Svenkerud, and Suruchi Sood, 1999. Effects of an entertainment-education radio soap opera on family planning behavior in Tanzania. *Studies in Family Planning*, 39(3):193-211.

Vaughan, Peter W. and Everett M. Rogers, 2000. A staged model of communication effects: Evidence from an entertainment-education radio soap opera in Tanzania. *Journal of Health Communication* 52(2):203-227.

Vaughan, Peter W., Everett M. Rogers, Arvind Singhal, and Ramadhan M.A. Swalehe, 2000. Entertainment-education and HIV/AIDS prevention: A field experiment in Tanzania. *Journal of Health Communication*, (Supplement), 81-100.

Meta-Analysis Studies

Snyder, L.B. and M.A Hamilton, 2001. Meta-analysis of U.S. health campaign effects on behavior: Emphasize enforcement, exposure, and new information, and beware the secular trend. In R. Hornik, ed. *Public health communication: Evidence for behavior change* 357-383. Hillsdale, NJ: Lawrence Erlbaum Associates.

Snyder, L.B. 2001. Development communication campaigns. In Gudykunst, W. & Mody, B., eds. *Handbook of International and Intercultural Communication*, 2nd ed. 457-478. Thousand Oaks, CA: Sage.

Snyder, L.B. 2000. How effective are mediated health campaigns? In R. Rice and C. Atkin, eds. *Public Information Campaigns,* 3rd ed. 181-190. Thousand Oaks, CA: Sage.

Theory-Driven Outcome Evaluations

Chen, Huey T. 2000. Development of a national evaluation system to evaluate CDC-funded health department HIV prevention programs. *American Journal of Evaluation* 22(1):55-70.

Chen, Huey T. 1997. Applying mixed methods under the framework of theory-driven evaluations in J. Greene and V. Carcacelli, eds. *Advancing Mixed Methods Evaluation. New Directions for Program Evaluation*. Jossey-Bass: San Francisco.

Chen, Huey T. 1997. Normative evaluation of an anti-smoking abuse program. *Evaluation and Program Planning* 20(2):195-204.

Chen, Huey T. 1990. *Theory-Driven Evaluations.* Sage Publications, Inc.

Kincaid, D. L. 2000. Mass Media, Ideation, and Contraceptive Behavior. *Communication Research*, 27(6):723-763.

Kincaid, D. L. 2000. Social Networks, Ideation, and Contraceptive Behavior in Bangladesh: a Longitudinal Analysis. *Social Science and Medicine*, 50(2):215-231.

Overcoming Threats to Internal Validity: Panel Discussion

Bollen, Kenneth.A, David K. Guilkey, and Thomas A. Mroz, 1995. Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive use in Tunisia. *Demography* 32(1):111-131.

Greene, William H. 2000. *Econometric Analysis*. 4th Edition, Prentice Hall.

Kennedy, Peter. 1996. *A Guide to Econometrics*, 3rd Edition. Cambridge, Massachusetts: The MIT Press.

Lin, D.Y., B.M. Psaty, and R.A. Kronmal, 1998. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *BIOMETRICS* 54: (3) 948-963.

Maddala, G.S. 1983. Limited Dependent and Qualitative Variables in Econometrics. *Econometric Society Monographs No.3*, Cambridge University Press.

Moffitt, Robert, 1991. Program Evaluation with Non-Experimental Data. *Evaluation Review*, 15:291-314.

Moffitt, Robert and Michele Ver Ploeg, 1999. *Evaluating Welfare Reform: A Framework and Review of Current Work*. Washington: National Academy Press.

Norretranders, Tor. 1998. *The User Illusion*. Penguin Books.

Robins J.M., A. Rotnitzky, and D.O. Scharfstein, 2000. Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. *Statistical Models for Epidemiology, the Environment, and Clinical Trials*. E Halloran and D Berry, eds. 1-95.

Rosenbaum, P.R. and D.B. Rubin, 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *The Journal of the Royal Statistical Society*. Series B. 45:212-8.

Sperber, Dan and Deirdre Wilson, 1998. *Relevance: Communication and Cognition*. Blackwell Publishers, Ltd.

Box 1. Eight Criteria for Claiming Impact

1. Observation of a change in the expected outcome
2. Correlation between that change and exposure to the program
3. Evidence that exposure occurred before the observed change (time-order)
4. No evidence of confounding variables that may have accounted for the change
5. Observation of a large, abrupt impact (magnitude): if promotion, easy and immediate/abrupt; if behavior that needs longer time-period, intermediate outcomes; maturity of program for some behaviors
6. Evidence of a causal connection (proximity and theoretical coherence)
7. Evidence that impact increases in proportion to level/duration of exposure (dose response)
8. Consistency with previous program research (replication with variation)

Notice that the first four criteria are the conventional parameters derived from the "experimental" approach.

Sources: Piotrow et al., 1997; Bertrand & Kincaid, 1996; Hill, 1971.

Box 2. Characteristics of Field Experiments

(a) Randomized Experimental Design

- An experiment carried out in the field rather than in the laboratory
- Randomly assigned set of individuals in the treatment group who receive the communication intervention
- Control or comparison group who do not receive the treatment
- Pretest/posttest measures of the effects variables

(b) Quasi-Experimental Design

- Random assignment is not used to create the comparisons from which the treatment-caused change is inferred (Cook & Campbell, 1979)

Source: Rogers' presentation, October 2001

Box 3. The Program Evaluation Problem

$$Y = a + bP + cX + \varepsilon \text{ - - - -} \quad (1)$$

$$P = c + dZ + \mu \text{ - - - - - - - -} \quad (2)$$

Where:

Y = Outcome
P = Exposure
Z = Instrument
X = Other variables and confounders

Z = Exogenous and random variation in the "opportunity" to be exposed
Z = "natural" experiment

$$Y = f + gZ + \upsilon \text{ - - - - - - - -} \quad (3)$$

(3) may be enough.

Source: Moffitt's presentation, October 2001

Figure 1a. Basic Model of Outcome Evaluation

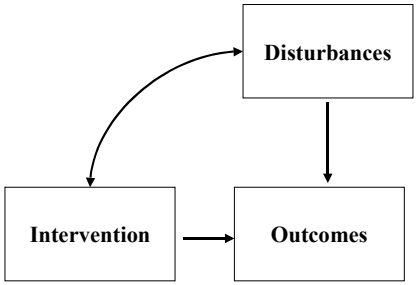


Figure 1b. Mediating Mechanism Evaluation

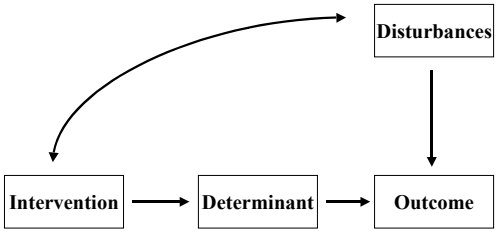


Figure 1c. Conceptual Framework of Theory-Driven Outcome Evaluation

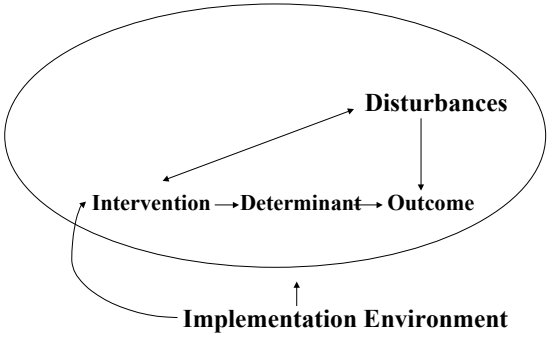
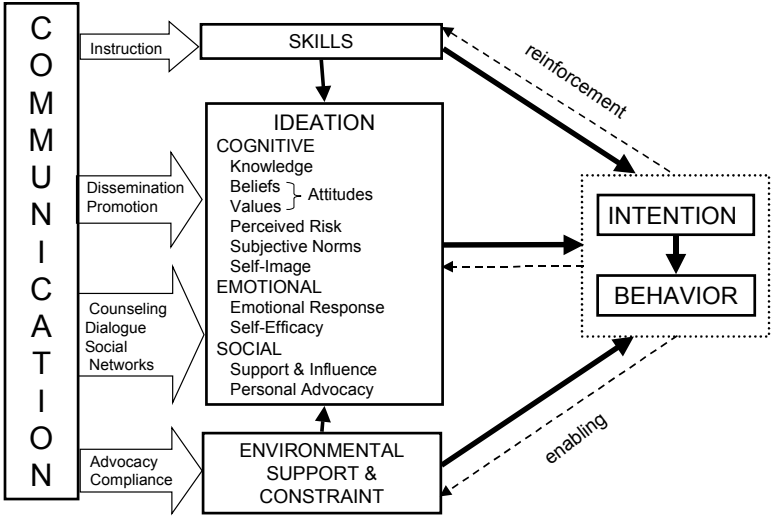


Figure 2. The Ideation Framework

A Model of Communication, Ideation, and Behavior Change



Kincaid, JHU/CCP, 2002

Figure 3. Theoretical Framework for the National Youth Anti-Drug Media Campaign

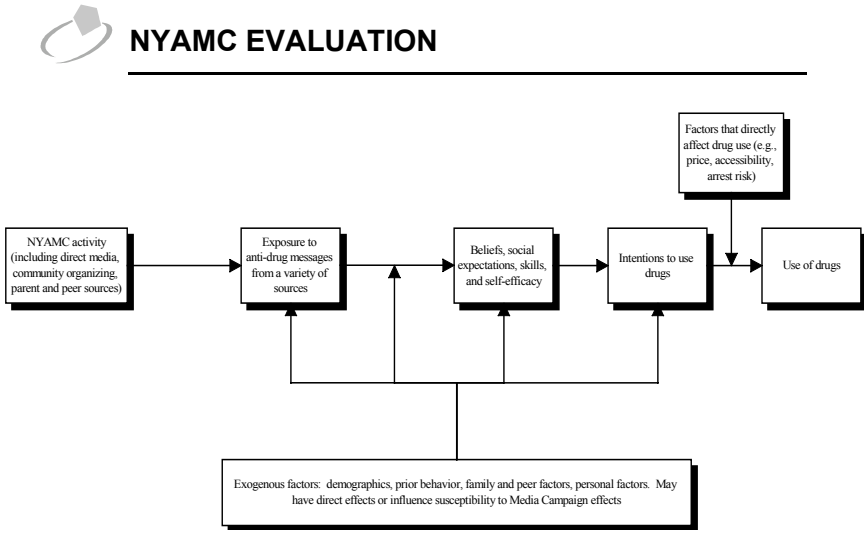
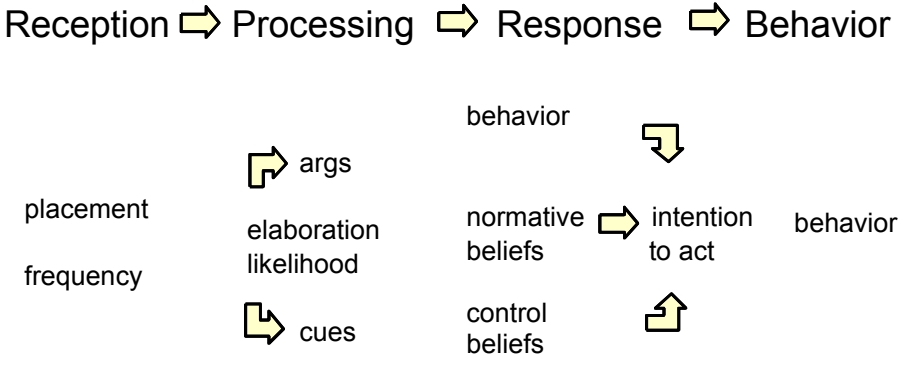


Figure 4. A Standard Model for Communication Based Interventions. The 1% or Less Milk Campaign



Source: Booth-Butterfield presentation, October 2001

Appendix A: List of Participants

NAME	INSTITUTION	TELE	EMAIL
Stella Babalola Sr. Program Evaluation Officer	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6232	Sbabalol@jhuccp.org
Jane Bertrand Professor & Director Designate	JHU/CCP Center for Communication Programs 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6304	Jbertran@jhuccp.org
Ruth Bessinger Evaluation Scientist MEASURE *Evaluation* Project	ORC Macro International 11785 Beltsville Drive Calverton, MD 20705	301-572-0811	Rbessing@macroint.com

NAME	INSTITUTION	TELE	EMAIL
Ties Boerma Project Director	The MEASURE *Evaluation* Project Carolina Population Center University of North Carolina at Chapel Hill University Square East 123 W. Franklin Street Chapel Hill, NC 27516-3997		Ties_boerma@unc.edu
Steve Booth-Butterfield Chief, Health Communication Research Branch	The Centers for Disease Control and Prevention The National Institute for Occupational Safety and Health 1095 Willowdale Road, 4050 Morgantown, WV 26505-2888	304-285-6090	Zee5@CDC.GOV

Dina Borzekowski Sr. Research Officer	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6240	Dborzeko@jhuccp.org
Marc Boulay Program Evaluation Officer II	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6231	Mboulay@jhuccp.org
Maria I. Busquets Acting Division Chief Management Specialist	G/PHN/POP/CMT U.S. Agency for International Development 1300 Pennsylvania Avenue, NW 3.06-006, 3rd Floor, RRB Washington, DC 20523-3601	202-712-4684	Mbusquets@usaid.gov
Huey T. Chen Chief, Health Services and Policy Research Branch Division of Prevention Research and Analytic Methods	The Centers for Disease Control and Prevention 4770 Buford Highway, K -73 Atlanta, GA 30341	770-488-8212	Hbc2@CDC.GOV

Jim Crimmins Chief, Strategic Officer	DDB Chicago Inc. 200 East Randolf Chicago, IL 60614	312-552-6413	Jim.crimmins@chi.ddb.com
Nafissatou Diop-Sidibe Research & Evaluation Officer	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-2699	Nsidibe@jhuccp.org
Maria Elena Figueroa Chief, Research & Evaluation	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6362	Mfiguero@jhuccp.org
Fannie Fonseca-Becker Sr. Program Evaluation Officer	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6166	Ffbecker@jhuccp.org

Elizabeth Fox Sr. Advisor	G/PHN/HCBC/OHN U.S. Agency for International Development 1300 Pennsylvania Avenue, NW 3.07-036, 3rd Floor, RRB Washington, DC 20523-3700	202-712-5777	Efox@usaid.gov
Christine Galavotti Chief, Behavioral Research Unit	Division of Reproductive Health, NCCDPHP Centers for Disease Control and Prevention 2900 Woodcock Blvd. (K-34) Atlanta, GA 30341	770-488-6401	cxg2@cdc.gov
Joanne Grossi Sr. Technical Advisor	G/PHN/POP/CMT U.S. Agency for International Development 1300 Pennsylvania Avenue, NW 3.06-192, 3rd Floor, RRB Washington, DC 20523-3601	202-712-0867	Jgrossi@usaid.gov

David Guilkey Sr. Technical Advisor	The MEASURE *Evaluation* Project Carolina Population Center University of North Carolina at Chapel Hill University Square East 123 W. Franklin Street Chapel Hill, NC 27516-3997	919-966-5335	David_guilkey@unc.edu
Sarah Harbison Sr. Research Advisor, Research Division Office of Population	USAID G/PHN/POP/R Ronald Reagan Bldg. 3.06-154 Washington, DC 20523	202-712-4536	Sharbison@usaid.gov
Marge Horn Deputy Chief of the Research Division	USAID G/PHN/POP/R Ronald Reagan Bldg. 3.06-154 Washington, DC	202-712-4536	Mhorn@usaid.gov

Robert Hornik Wilbur Schramm Professor of Communication and Health Policy	The Annenberg School for Communication of The University Of Pennsylvania 3620 Walnut Street, Philadelphia, PA 19104-6220	215-898-7057	Rhornik@asc.upenn.edu
Kimm Jayne (Kim Witte) Sr. Research Officer	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	517-347-3211	Wittek@msu.edu
Larry Kincaid Sr. Advisor	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6269	Lkincaid@jhuccp.org

Gary L. Kreps Branch Chief	Health Communication and Informatics Research Branch Behavioral Research Program Division of Cancer Control and Population Sciences National Cancer Institute 6130 Executive Boulevard EPN-Suite 4087A, MSC 7326 Bethesda, Maryland 20892-7326	301-496-7984	Gary.Kreps@NIH.gov
Robert Magnani Professor and Chair of the Department of International Health and Development	Tulane University School of Public Health and Tropical Medicine Department of International Health and Development 1440 Canal Street, Suite 2200 New Orleans, LA 70112	504-587-7331	Magnani@tulane.edu

Mahesh Mahalingam Communication Advisor Communication and Public Information Unit Department of Social Mobilization and Information	UNAIDS 20, Appia Avenue Geneva 27 CH-1211 Switzerland	41-22-7914918	Maheshm@unaids.org
Dominique Meekers	Tulane University 1440 Canal Street, Suite 2200 New Orleans, LA 70112		dmeekers@tulane.edu
Susie Mercado Gates Fellow, University of Philippines	JHU/CCP Center for Communication Programs 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6103	Smercado@jhuccp.org
Robert Moffitt Professor of Economics	JHU/Department of Economics Homewood Campus 429 Mergenthaler 3400 N. Charles Street Baltimore, MD 21218	410-516-7611	Moffitt@jhu.edu

Nancy Morris	Temple University School of Communications and Theater Department of Broadcasting, Telecommunications and Mass Media 06 Annenberg Hall 1601 North Broad Street Philadelphia, PA 19122-6080	215-204-8394	Amorri@unix.temple.edu
Sara Pacque-Margolis Sr. Technical Advisor	G/PHN/POP/PE U.S. Agency for International Development 3.06-035, 3rd Floor, RRB 1300 Pennsylvania Avenue, NW Washington, DC 20523-3600	202-712-1292	Spacque-margolis@usaid.gov
John Pierce Sam M. Walton Professor for Cancer Research Associate Director for Cancer Prevention and Control	University of California, San Diego Cancer Center 9500 Gilman Drive La Jolla, CA 92093-0658	858-822-2380	jppierce@ucsd.edu

Phyllis Piotrow Professor & Director	JHU/CCP Center for Communication Programs 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6304	Ppiotrow@jhuccp.org
Scott Ratzan	G/PHN/POP/CMT U.S. Agency for International Development 1300 Pennsylvania Avenue, NW 3.06-072, 3rd Floor, RRB Washington, D.C. 20523-3601	202-712-5022	Sratzan@usaid.gov
José Rimon II Sr. Deputy Director	JHU/CCP Center for Communication Programs 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6272	Jrimon@jhuccp.org

Everett Rogers Professor, Communication and Journalism	The University of New Mexico Dept. of Communication and Journalism C&J Bldg. Room 201 Albuquerque, NM 87131-1171	505-277-5305	Erogers@unm.edu
Ritu Singh Communication Program Specialist	GH/POP/CMT U.S. Agency for International Development 3.06-163 3rd Floor, RRB 1300 Pennsylvania Avenue, NW Washington, DC 20523-3601	202-712-5955	Risingh@usaid.gov
Juan Schoemaker Sr. Program Evaluation Officer	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6364	Jschoema@jhuccp.org

Leslie Snyder Associate Professor	University of Connecticut Communication Sciences Department College of Liberal Arts Phillips Building, Storrs Campus 850 Bolton Rd., Unit 1085 Storrs, CT 06269-1085	860-486-4383	Snyder@uconnvm.uconn.edu
Suruchi Sood Program Officer II	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6338	Ssood@jhuccp.org
Sharon Stash Program Officer	The Gates Foundation 511 Northeast 84 St. Seattle, Washington 98115	206-709-3108	sha-rons@gatesfoundation.org
Carol Underwood Sr. Program Evaluation Officer	JHU/CCP Research & Evaluation Division 111 Market Place – Suite 310 Baltimore, MD 21202	410-659-6142	Cunderwo@jhuccp.org

K. Viswanath Sr. Health Communication Scientist	Health Communication & Informatics Research Branch Behavioral Research Program Division of Cancer Control and Population Sciences National Cancer Institute 6130 Executive Blvd., EPN 4070 Bethesda, MD 20892-7363	301-594-6644	Viswanav@mail.nih.gov
Charles Westoff Professor Emeritus of Demography and Sociology and Senior Research Demographer	Princeton University Office of Population Research 251 Wallace Hall Princeton, New Jersey 08544	609-258-5867	Westoff@lotka.Princeton.edu
Joan Woods Education Advisor	AFR/SD U.S. Agency for International Development 4.06-079U 4th Floor, RRB 1300 Pennsylvania Avenue, NW Washington, DC 20523	202-219-0481	jwoods@afr-sd.org

Susan Zimicki Co-Director	Change Project Academy for Educational Development 1825 Connecticut Ave., NW Suite 800 Washington, D.C. 20009-5721	202-884-8825	Szimicki@aed.org

Appendix B: Agenda

EVALUATING THE IMPACT OF COMMUNICATION PROGRAMS

First Day, October 4

8:00 – 9:00 Breakfast

9:00 – 9:15
Welcoming by Phyllis Piotrow (JHU/CCP) and Ties Boerma (MEASURE *Evaluation*)

9:15 – 9:30
Evolution of Communication: from IEC materials to National Strategic Communication Programs, 1970-2000 by Jose Rimon II (JHU/CCP)

9:30-9:45
The Challenges in Evaluating the Impact of Communication Programs, by Maria Elena Figueroa (JHU/CCP)

9:45-11:00
Session 1. Cross-Sectional Surveys
Moderator: Juan Schoemaker

Charles Westoff, "Teasing Out Communication Effects from DHS Data."

Robert Magnani, "Experience With and Prospects for Bivariate Probit Estimation in Measuring the Impact of Adolescent Communications Interventions."

Discussion

11:00-11:15
Break

11:15 – 12:30
Session 2. Longitudinal Panel Data
Moderator: Stella Babalola

Robert Hornik, "Evaluation of the National Youth Anti-Drug media Campaign: Research Design and Analysis Strategies."

John P. Pierce, "Evaluation of Tobacco Control Programs, Cross-sectional and Cohort Designs."

Discussion

12:30 – 1:30 Lunch

1:30 – 2:45

Session 3. Experimental and Quasi-Experimental Research Designs to Evaluate the Impact of Mass Media Communication

Moderator: Kimm Witte

Everett Rogers, "Lessons Learned from a Field Quasi-Experiment in Tanzania."

Ruth Bessinger. "The Impact of Mass Media Exposure on Knowledge and Use of Condoms for STI and HIV/AIDS Prevention."

Discussion

2:45 – 3:00
Break

3:00 – 4:15
Continuation Session 3
Moderator: Susan Zimicki

Steve Booth-Butterfield, "The 1% or Less Milk Campaign and Reasoned Action."

Leslie Snyder. "Meta-analysis of the Impact of some Evaluation Designs: Issues on the Ability to Detect Campaign Effects."

Discussion

4:15 - 4:30
Synopsis of Day 1: Jane Bertrand

5:00 Reception

Second Day, October 5

8:00 – 9:00 Breakfast

9:00 – 10:15
Session 4. Using Theory to Strengthen the Argument for Impact
Moderator: Carol Underwood

Huey-tsyh Chen, "Theory-Driven Outcome Evaluations."

Larry Kincaid. "Communication and the Cumulative Effects of Ideation on Behavior."

Discussion

10:15-10:30
Break

10:30-12:00
Session 5. Panel: Overcoming Threats to Validity – Endogeneity, Self-selection, Competing Media, and Other Confounders

Moderator: Jane Bertrand

Panelists:
David Guilkey, "Endogenity Issues"
Robert Moffitt, "Issues of Self-Selection"
Jim Crimmins, "Evaluating Communication Programs."

12:00 – 1:00
Lunch

1:00 – 2:00
Development of Recommendations
Jane Bertrand

2:00
Final Words
Phyllis Piotrow, Ties Boerma

End of Meeting