

III. MEASURING CHANGE IN THE REAL WORLD: LEARNING FROM ONGOING AND PAST PROJECTS

How Related Fields Use Evaluation to Document Changes in Health Behaviors

What We've Learned So Far: Ten Observations for the Real World

Elaine Bratic Arkin

Health Communication Consultant

1. It is possible to change behavior and to measure those changes.

However, it all depends on what kind of change, with whom, and how one intends to intervene. Changing behavior is very complex and measuring those changes is very complicated. Very often, measuring change takes decades. While nationwide drops in smoking prevalence and stroke mortality are behavior-change success stories, we have been working on smoking for 30 years, and high blood pressure for 25 years. Behavior change cannot be measured in fiscal years, and it is important to resist policy/decisionmakers who will determine your program's success on a year-to-year basis. When smoking or high blood pressure rates are examined annually, there are fluctuations. Only by looking across decades do we see a dramatic drop.

Changing behavior—and measuring that change—also depends upon the type of intervention and the type of evaluation. Successful interventions are based on multiple theories and models, and are very complicated, multidimensional interventions with many different strategies. Their evaluations also are multidimensional with many evaluation methods and measures. We cannot expect to see similar kinds of changes with a one-dimensional program or evaluation effort.

2. If you can afford an intervention, then you can afford to evaluate it.

Often, in the “real world,” we are handed just enough time and just enough money to do the intervention. Make sure evaluation is built into the timetable up front, and make sure decision-makers and budgeters understand what is required to do the evaluation as well as the intervention.

Not every evaluation has to be elaborate. Often, there are quick and easy ways to monitor the program, so the excuse of no money or time to evaluate is rarely valid. However, there is a minimum below which the intervention and the evaluation won't work. In situations where you are not sure you have the resources to intervene or evaluate, think through whether you can make a change and measure that change before committing resources to it. Sometimes these are questions we don't ask because we have been given an assignment. As professionals in our field, we need to ask them more often.

3. Evaluate to achieve—not just measure—success.

Evaluation really is how success is achieved, and that is what is most important in the real world. Outcome evaluations are terrific for policymakers and decisionmakers, and for pilot programs before replication. But if the program has a limited time frame, process evaluation is critical for mid-course corrections, because there may not be a chance to replicate. Those changes need to be made the first time around.

Pretesting and other formative research is vital, to ensure the program is the best it can be initially. It is especially important if the time frame is short and mid-course corrections may not be able to be made. But lack of pretesting can also cause problems in large programs with very elaborate outcome evaluations: flaws in the materials may impact the outcome of the program, leading evaluators to conclude that the program didn't work when in fact the materials needed to be refined.

4. The evaluation designer and the program designer must work together on both aspects of the program.

The same theories and models need to be used for intervention and evaluation design; both parties need to be working from the same premise. Evaluators need to understand what the program can do, so they can design appropriate measures. Without close collaboration with the program designer, they may be measuring things that are unrealistic to measure and missing areas where great progress could be shown. Similarly, program designers need to build in evaluation needs from the start, to ensure they are getting the right information in the right way from the field and from the consumer.

5. Programs are designed to effect more than one type of change. Evaluations should measure more than one type of change.

As several speakers noted yesterday, the role of our interventions isn't so simple as changing behavior. Rather, our interventions must change all of the variables, or as many as we can, that we think influence the behavior. Then we hope that behavior change follows. If a program is trying to affect a whole series of variables, then evaluation needs to measure intermediate factors as well as behavior change (i.e., interpersonal, environmental, accessibility and availability changes).

6. You don't have to replicate costly evaluations if you can use other studies as proxy measures.

Not every program has to have the most elegant evaluation design. Often, other people have already done it, and their work can be used to illustrate that your program can have an effect. One example is the meta-analysis done by Isobel Contento and her team. Another is a review showing the results of well designed campaigns. Sponsored by the National Heart, Lung, and Blood Institute, it looked at hundreds of articles discussing public health, media, and community campaigns.

Once you have reviewed other studies, concentrate on evaluation measures that assure the quality of your intervention—otherwise it won't have an effect.

7. The elegance of the evaluation should match the complexity of the intervention.

Sometimes the evaluation team is stronger, in power or in dollars, than the intervention team. If the evaluation is not linked back to what the intervention realistically can be expected to do, the intervention will be shown up as a failure without ever having a chance. Conversely, if the intervention is flawed, the evaluation design cannot overcome that flaw.

8. The desired outcomes—and what the evaluation measures—should be based on realistic expectations for the length and complexity of the intervention.

Both the intervention and the evaluation need to be realistic based upon the status of the target audience. For example, at one point when the National Heart, Lung, and Blood Institute was working on high blood pressure and cholesterol education, they had two very different target audiences. For serum cholesterol, the American public did not know what it was. NHLBI worked with precontemplators, trying to introduce them to the issue and get them interested in it. For high blood pressure, the audience was people who have high blood pressure and have been on medication for years. These were people in the maintenance stage who found it very hard to comply with treatment over the rest of their lives. The program designers were being very realistic about the kinds of changes they wanted to see for the cholesterol program versus the high blood pressure program.

Also, we need to redefine what a realistic outcome is. In public health, we work with the hardest audiences to reach, then expect dramatic changes. If we make small changes, we feel we have failed. Contrast our approach to that of the private sector. Alan Andreasen sometimes uses Chevrolet as an analogy—that there is not one kind of Chevy, there are many of them for different audiences. Consider that maybe 3 percent of Americans want to buy a Chevy. Then split that 3 percent into the different Chevy models, and look at the percentage who will actually buy—perhaps 10 percent. The numbers are getting very small, but the manufacturer still shows a profit.

What do we do in public health? We take the hardest market to reach, not the easiest. We are always working with people who are at high risk, who have all kinds of barriers to behavior change. Then we look for a decline by 50 percent in the stroke rate. Or we want to get all Americans to quit smoking. If we make a 5 percent change, we think we have failed and often we lose budget. We need to step back and think about what is realistic to accomplish.

9. Evaluations, like interventions, need to be designed with a purpose and a target audience in mind.

Market segmentation applies to evaluations as well as interventions. If policy/decision makers are the audience, they want to know what kind of change occurred because of your efforts. If you want to go back to a program target audience, they may be most interested in whether the intervention works for people like them. Implementation managers want to know what is working, so they know what they need to fix.

10. Some components of evaluation design are transferable from program to program; methods and instruments should be more widely shared.

It is much more common to see cross-sharing of implementation strategies, materials, and training rather than of evaluation designs and instruments. There are many lessons to be learned from conducting evaluations that should be shared. When starting to evaluate something, remember to call a peer and see if they have anything in their files that will help you get started.

The Child and Adolescent Trial for Cardiovascular Health (CATCH)

Theresa Nicklas, Ph.D.

Associate Professor and Director, Dietary Studies

Tulane Center for Cardiovascular Health

Tulane School of Public Health and Tropical Medicine

Background

The primary goal of the Child and Adolescent Trial for Cardiovascular Health (CATCH) was to assess the benefit of school-based interventions designed to promote healthful behaviors in children and ultimately to reduce their risk of cardiovascular disease. The hypothesis of CATCH was that the interventions would lower the consumption of total fat (specifically saturated fat) and sodium; increase physical activity and nonuse of tobacco; and that these changes would favorably influence blood lipids, blood pressure, physical activity, and diet.

CATCH was unique in several aspects:

- It was the first such trial to integrate a number of successful approaches in a multi-year program involving the entire school environment, school children and their parents, in an ethnically diverse population, in four geographic areas of the United States.
- It identified and measured change at two levels: 1) the institution, through changes in policies and practices for all children in the school; and 2) the individual child, through classroom and parent involvement programs.
- The trial design used the school as the unit of intervention and analysis and a large number of schools were randomized to evaluate the intervention with adequate power.

CATCH has been ongoing since 1987. It is currently in Phase III, which involves tracking the behaviors and cardiovascular risk factors of the cohort until 1997. Field centers are the University of California at San Diego, the University of Minnesota, the University of Texas at Houston, Tulane University, and a coordinating center out of the New England Research Institute. CATCH funding is provided by the National Heart, Lung, and Blood Institute.

Overview of the Students, Schools, and Interventions

CATCH involved approximately 5,100 students who were in the third grade when the intervention began in the 1991-1992 school year. Of these students, 79 percent participated in the blood assessment at followup and formed the primary cohort for student-level study findings. The cohort's average age was 8.76 years at baseline. It was approximately 52 percent male, 13 percent African American, 14 percent Latino, 69 percent Euro-American, and 4 percent other or non-classified ethnic groups.

The 5,100 students represented 61 percent of third-graders enrolled in the 96 public schools participating in the study. Schools were recruited based upon distance from the study centers, ethnic diversity, food service characteristics, their commitment to offer at least 90 minutes of physical education per week, and their agreement to participate in a 3-year study requiring cooperation with random assignment to treatment and control status. Randomization occurred after all baseline measurements were completed.

There were 56 intervention schools (14 per field center) and 40 control schools (10 per field center). Intervention schools were further randomized into two equal subgroups. One received a school-based program consisting of school food service changes, physical education, and the CATCH curriculum. The other received the same school-based program plus a home and family-based program. Control group schools received the usual, if any, health curricula, physical education and food services, but none of the CATCH interventions. The CATCH interventions began in the 1991-92 school year and continued as students progressed through grades 4 and 5.

The CATCH interventions consisted of the following components:

- In-class curricula that focused on diet, physical activity, and tobacco usage
- Home curricula and family fun nights, supporting behavioral change in children
- School environment programs relating to food service changes (lowering fat, particularly saturated fat, and sodium in school meals) and a physical education program designed to increase the levels of moderate to vigorous physical activity in children
- For fifth graders, a smoking prevention program consisting of a classroom curriculum, home curriculum, and active encouragement of CATCH intervention schools to be smoke-free environments

Tracking and Evaluation

Baseline measurements of school and student-level behavioral outcomes were made at the beginning of third grade and each spring in third through fifth grade, with one exception: menu analysis was made at baseline, and during the fourth and fifth grade spring semesters. Physiological measurements on students were done at the beginning of grade 3 and at the end of grade 5.

Eat Smart: Food Service Intervention

The intervention school districts used one of two types of food delivery systems: either the food service director planned the menus and purchased food for all schools, but each school prepared the food, or the district had a central kitchen and delivered to satellite school kitchens.

Eat Smart, the food service intervention, was designed to give children tasty meals for school lunch and, where available, breakfast (59 of the 96 schools). Menus averaged no more than 30 percent of total energy from total fat, no more than 10 percent of energy from saturated fat, and a reduction of 25 percent in sodium levels.

The program involved four major intervention areas: menu planning, food purchasing, food preparation (including recipe modification in food production), and program promotion. Thirty *Eat Smart* guidelines were developed to assist with program implementation. The guidelines were based on the assumption that if a school cafeteria could meet all applicable guidelines and use the nutrient criteria, the dietary objectives of the *Eat Smart* program could be attained. Guidelines included activities such as skinning chicken, de-fatting ground beef, and whipping butter.

Materials developed to assist food service personnel included:

- A school meal program guide for school food service directors, managers, and technicians
- A recipe file box that included quantity recipes meeting both the USDA reimbursable meal pattern and the *Eat Smart* fat and sodium criteria
- A vendor product handbook that contained a list of products meeting fat and sodium criteria (and acceptable to students)
- *Newsline*, a one-page, two-sided bimonthly bulletin distributed to all cafeterias
- Intervention posters to display the *Eat Smart* guidelines and food preparation techniques so staff could use them as a quick visual reference

Implementation of *Eat Smart* was supported by comprehensive training sessions, annual booster sessions for food service staff, and monthly visits by CATCH personnel to observe and document implementation of the guidelines, solve problems, and provide feedback and support to food service staff.

Process and Outcome Nutrition Measures

The *Eat Smart* process evaluations were designed to assess implementation of the intervention, contextual factors, and external and competing programs that may affect implementation and influence study outcomes. Measures included:

- *Exposure (dose)*, to document the amount of the intervention participants received. Dose was measured by training attendance forms, intervention visits, promotional activities, school meal participation, and nutrient content of school meals relative to program goals.
- *Fidelity*, to document the degree to which protocols were followed by food service staff.
- *Characteristics of the schools* and staff that could mediate the impact of training and the extent of program implementation (food service staff knowledge, demographic characteristics, experience, self-efficacy, turnover rates, other programs independent of CATCH that may have impacted CATCH outcomes) and characteristics of the students that could affect program implementation through school meal participation.

Main outcome measures for nutrition were designed at the school and individual level. At the school level, changes in food service offerings were assessed using the average of five menu days. Vendor product information was also collected at baseline, interim and followup. At the individual level, a health behavior questionnaire assessed reported knowledge, intentions, choices, efficacy, and support. Dietary intakes were assessed using a food-record-assisted 24-hour dietary recall. Physiologic measures related to nutrition also were assessed.

All *Eat Smart* process and outcome measures have been published in a supplement of *Health Education Quarterly*, No. 2, 1994.

Results

Student participation in school lunch did not change as a result of the intervention. When meals were examined, there was a slight decrease in total calories of the school lunch menus in the intervention schools. However, the energy levels still met the one-third RDA. Significant decreases in both total and saturated fat were observed in intervention schools. Although reductions were also seen in control schools, they were less dramatic. Sodium content increased in both intervention and control schools. When sites were analyzed independently, it became apparent that vendor products were responsible for the increase.

Looking at student-level outcomes, a comparison of baseline to followup data shows that fat intake was significantly reduced by children in the intervention schools, and the reduction was much greater than in control schools. Much of the reduction in fat came from lower intakes of saturated fatty acids.

There was a corresponding increase in the percentage of calories from protein and carbohydrate in the intervention schools compared to the control schools. Data are showing that the vitamin and mineral intakes of the children were not affected due to these changes in the diet.

Physiologic measures showed no differences in body size at baseline or followup between the intervention and control schools. Growth was within normal limits of expected patterns for the age group. Similarly, pulse rate and blood pressure were not significantly different for students in intervention versus control schools. However, students' total blood cholesterol fell .4 milligrams more in the intervention than in the control schools.

Summary

CATCH was successful in demonstrating that a school-based program involving school food service, physical education, classroom curricula, and the family can be successfully implemented in diverse populations in four areas of the United States. CATCH successfully changed both the policies and the practices of schools and children's behaviors by the end of the 3 years, even given these modest exposure levels.

National Dairy Council

Madlyn Daley, M.B.A.

Senior Vice President, Marketing and Economic Research

Dairy Management, Inc.

Background

As part of an effort to update the dairy industry's *Food, Your Choice* nutrition education program, research was commissioned on the needs for dairy product information in the classroom. The goal was to provide a nutrition education program that teaches children the importance of choosing nutritious foods, while stressing the benefits of dairy products. The study had three objectives:

- ❶ Determine whether there is a role for dairy farmer organizations in schools, and if so, define the attributes and dimensions of that role

- ② Assess potential opportunities for enhancing dairy product exposure and usage in the school setting by generating alternative and ideal methods for presenting nutrition information to students and education professionals
- ③ Determine perceptions about the optimal grade level for presenting nutrition information so that resources could be channeled efficiently

Research to date has covered four phases: 1) focus groups, to investigate the scope and boundaries of both nutrition curricula and product issues; 2) a large-scale, national study to provide statistically reliable and actionable strategic marketing and planning information; 3) a pilot test; and 4) tracking utilization and acceptance of the program over time.

Research Highlights

Nutrition was rated important by all audiences surveyed (administrators and school board members, teachers, food service personnel, parents, and students). Food service personnel were more likely to rate it extremely important.

Teaching nutrition was thought to be more important at the elementary level than upper grades. Current teaching patterns reflected this perception, with nutrition more likely to be taught at the elementary level (43 to 48 percent of teachers said they taught nutrition in first through sixth grade). Eighth grade was the point at which teaching nutrition begins to drop. Consistent with these findings, there was strong support for nutrition education at all grade levels, but competition (from subjects such as math, drug education, physical education, and sex education) was less intense at the elementary level.

However, satisfaction with existing nutrition programs was not optimal; only 13 percent of teachers rated them excellent (although another 69 percent rated them very good or good). Qualitative work revealed that teachers were trying to format their own nutrition education program by looking at articles in magazines, newspapers, etc. Consequently, interest in something new was very high, with 52 percent saying they would definitely be interested in a new program.

Developing the New Program

Research identified the following decision criteria for teachers selecting a new program:

- Cost—budgets are tight
- Fit with regular curriculum, such as math, science and reading, otherwise, there may not be enough time to justify it
- Number of hands-on activities
- Intrinsic interest for both students and teachers

The National Dairy Council decided to target educational efforts at the elementary level, while reaching out

to teens with the dairy message through the entertainment media. For nutrition education efforts, the target audiences were teachers and superintendents, supplemented with activities targeted toward food service directors.

Two versions of *Nutrition, It is Elementary* were developed: *Snack Stars* targets second and third grade and *Snack Tricks* targets fourth and fifth grade. Both programs were built around the smart snacking concept, which was rated highly by educators. They reasoned that because students eat a lot of snacks, it is important for them to make the right choices. Also, snacking is an area of eating that is fun and an area kids can control.

Each version of the program consisted of: 1) six 30- to 60-minute lessons; 2) a teacher guide offering a teaching plan and suggestions for in-depth student involvement; 3) posters; 4) black-line masters; and 5) portfolio covers.

Evaluation: The Pilot Test and Monitoring

The pilot test had four specific objectives: 1) measurement of the extent to which the program increased students' knowledge of nutrition overall; 2) students' positive perceptions and knowledge of dairy products as they relate to snacking; 3) exploration of teachers' reactions to the program based upon their participation in the pilot test; and 4) identification of whether any modifications were necessary before full-scale marketing of the program.

The pilot test included eight second-grade and eight fourth-grade classrooms, recruited from the quantitative study. Teachers were asked to teach the program over a 3-week period, and they administered the pre- and post-tests to students. Results of the pilot test showed that students increased their knowledge of nutrition and the five food groups. After the program was taught, primary students' knowledge jumped from 17 percent to 81 percent correctly identifying all five food groups. Intermediate students' knowledge increased from 35 percent to 84 percent. Perceptions of specific dairy products being snacks that are really good for you also increased significantly.

The program was modified slightly based upon teacher input and produced for the 1993-94 school year. Dairy Council members distributed the program in a variety of ways, including direct mail, workshops, and key teacher contacts. Some charged for the program while others gave it away. Once received, 45 percent of the teachers used the program (regardless of distribution method), and another 27 percent said they planned to use it before the end of the school year. The primary reason for not using the program was time constraints. The program was used as it was intended—as a supplemental nutrition program rather than a core program. Teachers planned to modify the program just as they do other curricula, adding their own creativity and tailoring it to their particular class.

Minnesota Heart Health Program

Pat Snyder

Nutrition Coordinator, School of Public Health

University of Minnesota

Background

The Minnesota Heart Health Program (MHHP) was a community-based 13-year research and demonstration program that started in 1980. It involved six communities, three of which were intervention sites and three of which were controls or comparisons. The goal was to improve the health of the people of the community by 1) lowering the population levels of blood cholesterol, blood pressure, and cigarette smoking; 2) increasing physical activity; and 3) reducing the morbidity and mortality from heart disease.

The Grocery Store and Restaurant Programs

Needs assessment indicated that, rather than pamphlets and classes, what people in the community wanted was help where they purchased food and where they ate food. Programs in the grocery stores, the restaurant program *Dining à la Heart*, and school lunch programs were designed to address these needs.

As the MHHP evolved, pretesting with consumers taught us how to talk to them—and how not to. For example, the program's original message was, "Increase the use of various foods of plant origin, including grain, cereal products, legumes, seeds, vegetables and fruits." People did not understand. They wanted to know four things: 1) what foods to eat, 2) how much food to eat, 3) how frequently to eat the food, and 4) how to prepare the food. What the MHHP planners learned about how to communicate with people has been used by many other nutrition educators since. For example, the metaphor of a meat portion being the size of a deck of cards or the palm of one's hand came out of this work.

Most of the evaluation data collected was qualitative data and process data. Participation was examined first, then whether people were aware of the programs, behavior, and sales. For example, with the grocery store program, all of the major stores in two cities participated in the program, and two-thirds of those in the other city participated. Awareness of the shelf labels was about half the people responding to community surveys in all three cities. For *Dining à la Heart*, the restaurant program, about three-fourths or more of the people in each city noticed the hearts on the menu. Of those eating in *Dining à la Heart* restaurants, a little more than a third of those in each city reported that the hearts changed their menu choice.

In the grocery store program, video presentations, taste testing, and signs were used in the stores. Effectiveness of these tactics was evaluated by looking at sales data. For example, sales of top round increased significantly for the two weeks following a video presentation and taste testing featuring it. However, sales data for both grocery stores and restaurants was much more difficult to obtain when the MHHP program was conducted than it is now. At the time, it involved many site visits and creative ways of measuring change in product movement.

The School Lunch Program

The focus of the school lunch program was to lower the fat and sodium in individual menu items. After the program was in place, data collection consisted of analyzing the menu to see how often various items were offered.

Substantial change was seen. Examples include: offerings of buttered vegetables went down, while those of plain vegetables went up, and offerings of sausage pizza decreased, while those for cheese pizza went up.

Summary

The Minnesota Heart Health Program evaluation taught program planners:

- What questions to ask.
- What data are useful and how to use them.
- How to work with experts at the implementation sites to design evaluation plans and to set up the intervention and evaluation so that it fits into existing systems.
- How to measure success—sometimes, significant differences are not the only measure of success. For example, increasing sales a half percent at a grocery store represents a great deal of money to the store owners.
- How to be flexible.

Final results were published in the *American Journal of Public Health* in 1994.

Project LEAN

Sarah E. Samuels, Ph.D.

Health Program and Policy Consultant

Steps to Program Implementation

- ① ***There must be agreement in the field about a particular recommendation.*** First, there is a level of scientific consensus and then there is agreement about a particular recommendation. Such agreement is often hard to achieve and often does not last very long. The scientific evidence can change, requiring modifying a recommendation.
- ② ***Find a window of opportunity and seize it.*** Planners must be ready and able to act and respond to available funding and institutional support. Sometimes the opportunity can come and go quickly based on scientific consensus, availability of funding, and other variables.
- ③ ***Move quickly.*** Begin the planning process quickly.
- ④ ***Build partnerships.*** This step is important so that others are involved and buy into the agenda for the program, creating a broader range of support for it. More partners provide more credibility and legitimacy.

- ⑤ **Conduct formative research.** Focus groups give a voice to the consumer by providing an understanding of what the consumer needs and wants.
- ⑥ **Develop policy and advocacy strategies** to give the program a presence in the policy environment and to begin creating environmental change that will support the program's message.
- ⑦ **Launch the program with as much visibility as possible.** The more people that hear and know about it, the more credible its implementation will be.
- ⑧ **Support community implementation** of the program as a way to disseminate the message and tailor it to different audiences.

Evaluation

In the past, in the "real world" evaluation was often an afterthought or something to tie into the program after it was designed. It may have been designed when funders were asking about the cost-effectiveness of the program, how their money was spent, and what the impact of the program was. In this time of government downsizing, the days of funding big national programs with intervention and control or comparison communities over a 10-year period are gone. Today's programs will require a variety of funding sources. Sometimes there will not be time to think out all aspects of program design and evaluation planning. The challenge is to put together programs that will be sustained, are fundable, and can be evaluated.

Issues to consider when designing an evaluation:

- ① **Cost.** Don't make the evaluation larger than the program itself. Avoid trying to answer broad outcome questions. Don't let the evaluation dictate the program by adding in components because they can be evaluated.
- ② **The magnitude of the intervention.** Particularly with community programs, a long process of planning and development may be required before the program is in place. Look for reasonable outcomes given the time period of the program, rather than outcomes that cannot be achieved in a short time-frame. Also, don't make the evaluation an intervention. Don't collect so much data that the evaluation winds up influencing people's behavior, their thoughts, or how they relate to the program issues.
- ③ **The audience for the evaluation.** This is a critical consideration. Thinking through the kinds of questions the audience may have about the program helps focus the kind of evaluation that is done. Sometimes audiences for evaluation change, or their questions change. The evaluation cannot be retrofitted to answer new questions.
- ④ **Know what is important beyond health outcomes.** For example, it might be cost savings, gaining a marketing edge, or positioning nutrition in a positive light to garner greater support in the future.

Example: Project LEAN

Project LEAN—Low-fat Eating for America Now—was initiated by the Kaiser Family Foundation in 1988 when a consensus was attained about the importance of dietary fat as a major risk factor for chronic disease. The foundation immediately created the Partners for Better Health to help develop the program. Partners for Better Health represented 34 Federal agencies and professional and industry associations that shared a commitment to reducing fat in the American diet. The foundation funded Project LEAN for 3 years. In 1991, it was transferred to the American Dietetic Association's National Center for Nutrition and Dietetics.

The goals of the program were to accelerate a trend toward lower fat consumption by increasing the availability and accessibility of low-fat foods and to promote greater collaboration.

Focus groups conducted across the country provided information that helped identify strategies to address needs. For example, greater awareness of the need to reduce dietary fat intake was necessary, so mass media strategies were developed to increase awareness—a public service advertising campaign sponsored by the Ad Council, and a national publicity effort to bring attention to the Project LEAN message. The focus groups also revealed that taste and convenience were the most important factors in determining behavior change. People wanted the food to taste as good as the food they were currently eating and were very interested in convenience and quick preparation. Chefs and food journalists were selected as important and effective spokespersons for influencing popular taste.

Partnerships were formed to strengthen, reinforce, and multiply the message. Ten community programs were funded to create institutional change in the marketplace. Many other states and communities picked up the campaign on their own, tailoring it to their local needs.

Evaluation

Evaluation of Project LEAN consisted of a number of activities:

National media. Tracking of public service advertising showed that PSAs appeared in 50 percent of television viewing households. Print ads reached 16.5 million readers, and 2,800 radio stations played the PSAs. Print publicity consisted of 291 articles reaching 34 million readers. Television and radio appearances reached 27 million viewers and listeners.

Hotline. The hotline number included in the public service advertisements generated 300,000 calls in 12 months, peaking at 25,000 to 28,000 calls per month.

Leveraging of funds was tracked because the Kaiser Family Foundation was interested in whether its investment helped to bring additional funds to support the program. The foundation allocated \$3.5 million over 3 years; an additional \$354,500 was raised from collaborating organizations, many of whom were members of Partners for Better Health and who co-sponsored programs, training programs, special events, and various other activities. The corporate sector contributed \$94,000, government provided \$236,000, and professional associations provided \$23,000.

Community program evaluation of the 10 foundation-funded sites was done by an external evaluator and involved tracking the process of implementation at the community level. All of the sites were provided with na-

tional materials but also developed their own and tailored their strategies to reach local populations. Institutionalization—whether the program continued beyond the foundation funding by obtaining additional resources to continue—was also examined. Most sites were found to have formed alliances with State or community agencies and to have obtained some type of additional funding, either as permanent space, dedicated staffing, or in-kind support.

Evaluation methods included key informant interviews, followup institutionalization interviews, analyses of monitoring reports that each site submitted, and site visits. All information was compiled into a case study analysis of each site. These analyses were then examined to identify themes. Themes included: 1) a grocery-store component with high-level commitment of key corporate personnel; 2) media activities enhanced by mutually beneficial partnerships (e.g., other agencies in the community that wanted to team with Project LEAN); and 3) strong and active community coalitions.

Conclusions for Nutrition Program Planning and Future Evaluations

- ① Intermediate measures provide valuable short-term indicators of success. For example, looking at whether the program was fully implemented and whether it was sustained is an important marker of implementation success.
- ② An action-oriented evaluation design provides useful feedback to participating organizations. It lets them know how they are doing and how they can improve what they are doing. It allows the program to be changed to adapt to environmental changes.
- ③ The evaluation should be on a scale no larger than the intervention. It needs to look at the issues that are going to be affected by the program intervention itself.
- ④ There is a need to create sustainable programs at low cost. As the policy climate changes, we need to rethink the kinds of programs that can be supported over the long term, and who is going to support those programs. We need to educate and inform foundations and other potential funders of the importance of nutrition.
- ⑤ We need to continue to create national templates that local programs can use and adapt to their population, like Project LEAN and 5 A Day.
- ⑥ We need to build on the cumulative effect of previous programs.

We need to be aware of the negative consequences of our efforts and be frank about the limitations of what we can accomplish. Don't oversell; it will cause more damage than good. Also, don't undersell. Make the efforts you have put into your programs known to all.

5 A Day

Jerianne Heimendinger, Sc.D.

Program Director

National Cancer Institute (NCI)

National Institutes of Health

The 5 A Day Program

The goal of the 5 A Day program is one of the Year 2000 Health Objectives: To increase the average consumption of fruits and vegetables to five servings a day by the year 2000. In addition, there are two specific program objectives: 1) increase public awareness of the importance of eating at least five portions of fruits and vegetables every day, and 2) give them the skills to do so.

The program began in California in 1986, funded by a National Cancer Institute grant. The State developed the public-private partnership underlying the program and demonstrated that it could work. They also developed the theme, logo, and the first set of materials.

When the program came to the national level 2 years later, it was not clear how to proceed. This confusion was due in part to the fruit and vegetable industry having a large number of small players that were not as organized as the Meat Board or the Dairy Council. The solution was to form the Produce for Better Health Foundation, enabling NCI to work with a nonprofit foundation to partner with the industry. The foundation now has more than 800 members, including the Produce Marketing Association, the United Fresh Fruit and Vegetable Association, the Food Marketing Institute, a variety of individual commodity boards, manufacturers, supermarkets, suppliers and merchandisers.

Program components include 1) supermarkets, as the major point of sale; 2) the mass media effort, which is designed to keep 5 A Day in print, radio and television media by coming up with newsworthy information and trying to keep the program fresh; 3) redirected advertising dollars provided by the produce industry; 4) food service, which 5 A Day is beginning to move into; 5) the community component; and 6) the research component.

The Community Component

The community component is based on the premise that awareness can be created through the media. What seems to be more effective in the community is also to be personally relating with people. We want to explore all the creative ways to draw people into the 5 A Day message. We want to get them to be aware of it, help them build skills, etc.—techniques to bring home to people the message and some ideas and skills about how to do it.

The public-private partnership combination is also important at the community level. Having a health authority connected with the industry to guarantee that this message is good for a person's health has been shown to work well. At the national level, NCI provides the credible health source. The produce industry provides the skills and capacity to reach the public in ways that NCI could not. At the local level, the health department is the appropriate partner since it is also part of the Public Health Service. The health department generally has developed working relationships with other agencies in the government with similar agendas.

One of the purposes of developing a network for the program is to provide at the community level the necessary interactive components of successful behavior change interventions. By working with the health department, hopefully a huge network is already in place. State agencies participating in 5 A Day are licensed. As part of the license, they are asked to conduct a variety of activities, including at least one function each year or one major theme-related program event.

Theoretical Constructs

Social cognitive/learning theory, community information processing, the health belief model, social marketing, and stages of change are the major theories incorporated into the program. NCI attempted to use a matrix approach to get at the major constructs that are important for creating behavior change. Creating awareness is one of these. So is motivation—that it is important for the population to be motivated if they are going to make changes. Consumers need to have the skills and feel that they can make the changes. Social support from family, co-workers and schoolmates is also important. The food system and environmental support are also critical. A person might have the best of intentions and motivation, but if the options are not available in the school or work cafeteria, then the probability of doing the desired behavior is low.

Channels

There is a window of opportunity for doing channel-specific projects with specific target populations. Ultimately, we hope to get community change, but we are not going to get there by a broad effort. We need to focus on specific channels, such as schools, food assistance programs, work sites, and supermarkets—so we can be effective and hopefully expand from there.

Intervention and Evaluation Efforts

The Centers for Disease Control and Prevention are funding roughly 25 States to do 5 A Day interventions. Many states organize coalitions for 5 A Day in a variety of ways based on what fits in their environment. NCI has funded 5 A Day evaluation efforts in four states and is in the process of funding more. In addition, nine flagship evaluation projects will be using randomized designs to look at the effect of 5 A Day in target populations in specific channels—work sites, schools, churches, and the WIC program. Some common measures will be used across the nine sites and have been made available to the States as well.

As for national evaluation activities, a baseline study was conducted before the industry began the national program in 1991. The purpose of the study was to measure consumption and get information on knowledge, attitudes, and stages of change. A followup survey will be conducted in the fall of 1996. Process evaluation activities include retail activity reports, quantities of materials sold by the Produce for Better Health Foundation, the growth of State agency licensees, and the analysis of State activity reports. NCI also tracks national media coverage of the program and has done so since it began.

We want to develop substudies to connect 5 A Day with outcomes. One example is developing case studies to get a better handle on what happens at the State level. In specific localities within the four evaluation States, we would like to measure what sales and other activities take place during 5 A Day week. Ultimately, we would like to develop an index that would show the intensity of the intervention in the various States and connect that with various outcomes. For example, many States conduct the Behavioral Risk Factor survey; we might make some connections there.

Charting the Course from Lessons Learned

Robert Hornik, Ph.D.

Professor, Annenberg School for Communication

Director, Center for International Health and Development Communication

University of Pennsylvania

Problems with Current Outcome Evaluation Models

Using randomized, controlled trials at the community level as the gold standard for evaluating nutrition and communication education programs is a major mistake. It is a major mistake for practical reasons. We don't have the resources to do it very often. But of more importance, it is a mistake for theoretical reasons. It answers the wrong question very well.

Randomized trials, such as the Stanford Community Program, the Minnesota Heart Health Program, the Pawtucket Heart Health Program, and the COMMIT smoking trials, have either had very small effects or no effects at all, according to their evaluators. At the same time as those gold standard programs appear not to be working, we see very large changes in health behavior associated with other public health education and communication programs, such as the National High Blood Pressure Education Program and the AIDS communication programs that have gone on around the world.

For example, the Stanford program did its evaluation using treatment and control cities over a 4- or 5-year period. One of its outcome measures was the level of cholesterol in the diet. During that period, treatment cities' level of cholesterol in the diet declined slightly more than control cities. However, after the end of the treatment period, dietary intake of cholesterol in control cities was declining three times as fast as in the treatment cities during the treatment period. So the major treatment control trial found no effects until afterwards in the control cities, when there were very nice effects.

What is going on here? There are two major hypotheses. One possible explanation is that these inferences about effects are inappropriate because, in fact, there really were no effects. That isn't even slightly credible. The alternative is that public health education and communication programs were at least partly responsible for the observed sharp changes in behavior and that the treatment-control model that fails to find such effects is actually at fault. But the model of influence on population change implicit in controlled trials and testable in controlled trials does not match our model of what influences widespread behavior change.

Evaluations need to respect the model of change under which we operate—not somebody else's model of change. The model of those who are doing these sorts of programs should be identified. What evaluation designs are consistent with programs such as the National High Blood Pressure Education Program, Project LEAN, or 5 A Day? To help us think about that, a brief description of how the process for change occurs for the National High Blood Pressure Education Program is shown below.

A woman sees some public service announcements, a local TV health reporter's feature story, and a discussion on Oprah about the symptomless disease of hypertension. She checks her blood pressure in a newly-accessible shopping mall machine. Those results suggest a problem. She tells her spouse, who has also seen the ads and encourages her to get a checkup. She goes to a physician, who confirms the presence of hypertension and encourages her to change her diet and then return for monitoring.

Meanwhile, the physician has become more sensitive to the issue because of a recent *Journal of the American Medical Association* article and some recommendations from his professional society, a conversation with a drug retailer, recent conversations with colleagues, and exposure to television discussions of the issue. The patient talks with friends at work about her experience, which increases their concern, so they go to have their blood pressure checked. She returns for another check-up and her pressure is still elevated although she has reduced her level of cooking salt. The physician decides to treat her with medication. The patient is ready to comply because all the sources around her, personal, professional and mediated, are telling her that she should.

This program is effective not because of PSAs or specific programs in physician education. It is successful because the National High Blood Pressure Education Program has changed the professional and public environment as a whole around the issue of hypertension. That is the basic model of effects, or influences on behavior change, that we have talked about.

What does this imply about the model of population-wide behavior change? That we have a national culture, not just the local culture alone. That local activities build on a spine of national programs that work together. Don't evaluate, don't try to compare treatment and controls that are geographically defined unless there is really going to be a difference in exposure to messages. Don't accept trials as negative evidence until you look hard at the evidence for differences in exposure between the so-called treatment and control areas.

Message Exposure

In general, we have done very well in developing good theory and practice in developing messages. But we need to look at how we assure exposure with an appropriate frequency over time for those well-developed messages. That is, we need to be sure we achieve not only good messages, but messages people are actually exposed to.

A randomized design that establishes that a great intervention has had an effect is meaningless if the intervention is not replicable on any scale that matters. We have to spend as much energy designing interventions that are workable on a real scale as we do designing the perfect message. We spend too little time worrying about exposure and how to get it.

Alternative Models of Change

We often strive to achieve very rapid change in a short time, and there are instances where this has occurred. A Philippine immunization program resulted in a change from 32 percent of the children fully immunized on time to 56 percent in one year. There is evidence that programs trying to affect babies' sleeping position to avoid Sudden Infant Death Syndrome (SIDS) have had very rapid effects. In Switzerland, condom use associated with AIDS communication went from 8 percent of those with casual partners using a condom with those partners all the time to 48 percent.

So we have examples of rapid change. But, these are unusual programs and each is not always replicable under the conditions in which we ordinarily work. In the Philippine immunization program, people were already getting their children immunized, but they were getting them immunized late. They did not know they needed to bring them in by 12 months. Information was enough to change behavior. SIDS is greatly feared and moving babies from sleeping on their stomachs to sleeping on their backs was a very easy change in behavior.

Condom use is a harder behavior. But the Swiss Stop AIDS campaign was accompanied by massive media coverage of the risks of HIV and its association with unprotected sex. Attributing the change to the campaign alone would be difficult, although that was probably important.

Some alternative models of change to consider:

- *A slow-change model versus a rapid-effects model.* Depending upon which of these is appropriate, your intervention will be designed differently.
- *An individual cognitive model versus a collective social norms model* or an environmental model. Sometimes we think people will change just because they know something different, but other times we think they will change only when the community around them changes and changes its ideas.
- *A direct-exposure model versus diffusion through interaction.* The first model posits that people hear the messages and are affected. The second, that people hear messages and talk with other people, and the other people are as likely to be affected as they are.
- *A single learning model versus a multiple-channel, multiple-exposure, wear-them-down model.* That is, do we really expect someone to change upon hearing 5 A Day once? Clearly not. It is a multiple-exposure, multiple-channel, wear-them-down model. Keep going out with the message; at some point it becomes part of a collective social norm.

Each of these implies something about who will be affected, through what channels, and how fast they will change behavior. These intervention models are what should be used in our decisions about how to evaluate. If we think that the second model of each pair described above (except the first pair) is really the more likely intervention model, our evaluation should be designed accordingly. Similarly, if our program will be happening nationally, or partially nationally, we won't get discrete treatment and control cities. We should consider that when selecting an evaluation model—i.e., a controlled trial would be inappropriate because there will be no true control community.

Alternate Evaluation Models

Are there ways to do credible evaluations that don't require randomized or controlled communities? Yes, at times. Will these designs produce believable effects for lab researchers? No. However, we have two tacks. We can use their design—the vaccine or placebo control design—and surely find that our programs don't work because they don't fit the model. Alternatively, we can use our designs and try to convince them of the credibility of the evidence we put forward. We have to take the second tact, though it is not easy. It would be a mistake for us to try to live within those designs that are inappropriate for the way we think our programs work. Here are four alternative evaluation designs:

- ① *Time series analysis*, which is a very powerful procedure. It is a longer-term design, but very convincing. It is tough to argue that there was not some effect with national programs using it. This is particularly true if an indicator has a known trend over time and the effects of the program can occur fairly quickly and are substantial. This type of analysis will work for programs like the National High Blood Pressure Education Program. It is very hard to argue that something else important was going on starting in 1972, exactly when the national program began, and that it was associated with a very sharp decline in rates.

- ② These analyses can also be more complex, such as *comparing indicator levels between groups with differential likelihood of exposure to a program* in before and after cross-sectional samples. For example, consider a case in Zambia involving a weekly radio soap opera trying to get people to talk about AIDS and encourage safer behavior, particularly more condom use. We compared people who had greater and lesser probabilities of being exposed to the program. We based the comparison upon who owned and listened to a radio, and looked at the proportion of women who had ever used a condom, which increased during the time period. If the radio program was responsible for the increase in women using condoms, we would expect a more rapid rate of change among the high-access sample (those who own and listen to radios). We didn't see that. So our conclusion was that access to the program had no effect.

However, it is unlikely that in a nine-month period a radio soap opera was going to profoundly influence condom-use behavior. What should have happened was some intermediate variables would have changed, such as talking in the household about AIDS and the risks of AIDS, which the program encouraged. Here we saw a greater advantage to the high-access sample than the low-access sample; those who had easy access to radio were more likely to talk than those without access. However, when we looked at whether those who actually listened to the program were more likely to talk to their spouse, we found no difference.

- ③ *When change occurs rapidly and there are no other plausible explanations*, we can make a case that our program caused the change. For example, with the Philippine immunization program, the change in proportion of 11-month-old children who received all eight vaccinations on time is tough to explain, other than by the presence of the immunization program.
- ④ *When the observed outcome is credibly explained by a process we thought would lead to change*. If we can show that the people who are exposed to the program and the people who are more knowledgeable are those who are in fact engaging in the appropriate behavior, we can make a stronger case for the effects of the program. For example, our program increases knowledge in order to stimulate behavior change, like the Philippine immunization program that told parents children should have their vaccinations by 1 year of age.

Summary

We need to be sure the program model is what is driving the evaluation, not vice versa. It will rarely be the case that control-treatment designs will serve your purposes very well, although at times they will. Make sure and think through the model of influence being used. Know how the program is supposed to work, who it is supposed to affect, how fast, and what levels of exposure have to be achieved. Design the evaluation around the model of influence rather than borrowing a model of evaluation from fields that are largely inconsistent with our models of influence.