PN-ABU-620

AGENCY FOR INTERNATIONAL DEVELOPMENT
PPC/CDIE/DI REPORT PROCESSING FORM

ENTER INFORMATION ONLY IF NOT INCLUDED ON COVER OR TITLE PAGE OF DOCUMENT

1. Project/Subproject Number: 931-1018

2. Contract/Grant Number: DPE-5984-Z-00-9018-00

3. Publication Date: 1994

4. Document Title/Translated Title:

Making Evaluations Work: A Cross-Cutting Review of Impact Evaluation Methodologies Used by Healthcom-I Projects

5. Author(s)

1. Barton R. Burkhalter
2.
3.

6. Contributing Organization(s)

Academy for Educational Development (HEALTHCOM project)

7. Pagination: 32

8. Report Number:

9. Sponsoring A.I.D. Office: Bureau for R&D/Office of Health

10. Abstract (optional - 250 word limit)

11. Subject Keywords (optional)

1.
2.
3.
4.
5.
6.

12. Supplementary Notes

13. Submitting Official: Willard Shaw

14. Telephone Number: 202-884-8789

15. Today's Date: 4-17-95

DO NOT write below this line

16. DOCID

17. Document Disposition: DOCRD [] INV [] DUPLICATE []

AID 590-7 (10/88)

Communication for Child Survival

HealthCom

PN-ABU-620

MAKING EVALUATIONS WORK:

A Cross-Cutting Review of Impact Evaluation Methodologies Used by Healthcom-I Projects

Prepared by

Barton R. Burkhalter

A publication of

the HEALTHCOM Project

USAID

July 1994

Acknowledgements

The quality of the ideas and the accuracy of the information presented in this paper were greatly enhanced by the contributions of Robert Hornik, Judith McDivitt, Stan Yoder and Susan Zimicki of the University of Pennsylvania. They gave generously and gracefully of their time and patience in helping me to understand and document the work they had done. It is important to note that they are in no way responsible for any inaccuracies that exist in the paper. Let me note again that this paper has not portrayed the full power of the arguments used by several of the evaluations to analyze causality. I refer the reader to the original evaluation documents. I would also like to thank Mark Rasmuson for getting me started on this project.

The research described in this document was funded by the HEALTHCOM or Communication and Marketing for Child Survival Project. HEALTHCOM is a project of the Office of Health and Nutrition, Bureau for Global Programs, Field Support, and Research of the U. S. Agency for International Development (USAID). The project is being conducted by the Academy for Educational Development (contract no. DPE-5984-Z-00-9018-00) and its subcontractors: the CIHDC of the Annenberg School for Communication at the University of Pennsylvania, Applied Communication Technology, Birch & Davis International, The Futures Group, The Johns Hopkins University, and Porter/Novelli. This document does not represent the views or opinions of USAID.

MAKING EVALUATIONS WORK:

A Cross-Cutting Review of Impact Evaluation Methodologies Used by Healthcom I

by Barton R. Burkhalter, *Academy for Educational Development*

ABSTRACT

The eight formal impact evaluations of Healthcom I (1985-1990) provide an abundant and valuable data set about the actual process of supplying long-term technical support in communications to large health programs in six countries in Africa, Asia, the Middle East, and South America. This paper describes how these evaluations were carried out, and thus provides important information on how successful evaluations of large-scale communication activities really worked in the developing world.

INTRODUCTION

The Healthcom Project applies communication technology to health and child survival activities in developing countries. During 1985-1990, a period referred to as Healthcom I, the project provided long-term technical support in communications to 11 large health programs in nine countries, and completed formal impact evaluations on the contribution of the communication component in eight of these programs.[1]

These eight evaluations provide a rich source of information about the actual process of evaluating large health programs in developing countries, especially in relation to the contribution of communication activities. The programs and the communication activities were varied and presented many different challenges to the evaluators. In response, the evaluations were equally varied and imaginative. Thus, the evaluations present an interesting story.

The original impetus for this paper arose from the need to design streamlined evaluation guidelines for the next phase of the Healthcom Project. Designing those guidelines required a careful review of the impact evaluations carried out during Healthcom I. During this review, it became apparent that these evaluations were an unusual and valuable data set for several reasons:

(1) The programs being evaluated were large-scale, often national in scope. They were not small pilot activities or experimental studies.

(2) The evaluations themselves were also large-scale and attempted to ascertain the impact of the program, especially its communication component, on the entire population targeted by the program.

(3) In spite of the many problems encountered, the evaluations were generally successful in their effort to determine

[1] Evaluations were completed for eight health programs in six countries, namely Ecuador (E), Central Java in Indonesia (C), West Java in Indonesia (W), Jordan (J), Lesotho (L), Metro Manila in The Philippines (M), the National Philippines Program (P), and Zaire (Z). Formal evaluations were not completed in three countries that initiated long-term relationships with Healthcom I (Nigeria, Papua New Guinea and Paraguay), although reports describing the programs in these countries were completed (25,26,3). Throughout this paper, the eight health programs and their associated impact evaluations are referred to in tables by their capital letters in parentheses.

BEST AVAILABLE COPY

how much effect the program had. (Note: the success of the evaluations is, to a great extent, independent from the success of the program.)

(4) The evaluations are well documented. They contain information about how they really were done and not just information on what was planned

Thus, the Healthcom I evaluations are a data set that is informative about the characteristics of successful, comprehensive evaluations of large-scale health programs in developing countries. It is unusual to have such a data set. Too often, evaluation designs are based on conceptual models that have not been validated by systematic and documented experience.

This paper is a step towards understanding the valuable lessons contained in the Healthcom I evaluations. It summarizes and compares the methodologies used by the evaluators. In other words, it describes how the evaluations were carried out. It does not address the results of the evaluations; that is, information about the success and failure of the interventions. This is reported in detail elsewhere (1,4,6,7,12,18,20). Neither does this document address the methodology used by Healthcom in applying communication technology, including formative evaluation, because it is also documented elsewhere (5,11,19).

The Healthcom project is funded by the U. S. Agency for International Development (USAID) and implemented by a consortium led by the Academy for Educational Development (AED). It includes the Annenberg School for Communication at the University of Pennsylvania, Applied Communication Technology, Porter/Novelli and PATH/PIACT. Healthcom I was preceded by the Mass Media and Health Practices Project in 1978-1985, and since 1990 has continued as Healthcom II. Healthcom I worked with host countries and USAID missions in one of three general ways:

(1) Long-term support lasting several years, usually is association with one or two national or regional health programs;

(2) Short-term assistance on particular problems; and

(3) Continuing technical support on long-term research efforts in Honduras and The Gambia begun during the Mass Media and Health Practices Project.

The evaluations addressed in this paper deal only with the long-term technical support activities noted in (1) above.

The primary sources of information for this paper were the individual evaluation documents for each of the eight programs (8,9,14,15,16,23, 24,27). All of the evaluations were undertaken by the Annenberg School in cooperation with host countries. All of the evaluation documents were prepared by the Annenberg staff, under the auspices of the Healthcom I project. The eight evaluation documents provide information about the interventions being evaluated, the methodologies by which the evaluations were carried out, and the results of the interventions. The style of the evaluation reports and the level of detail they contain about various aspects of the evaluation methodologies, however, varies considerably. Information about the methodologies was also obtained from several other documents prepared by the Annenberg staff (1,13,17, 21,22). Finally, several discussions were held with Annenberg staff, and an early draft of this paper was reviewed by its staff members.

These sources yielded information that is comparable across the eight programs. Numerous variables were reported for many or even all of the evaluations, while other variables were available for only a few. This variation in information is not particularly detrimental to this paper, because its primary objective is to depict the variety of evaluation methodologies used by Healthcom I, thereby gaining a clearer

understanding of how successful evaluations of large-scale communication programs really worked in developing countries.

In order to compare the eight evaluation methodologies, this paper organizes the information about the methodologies into seven broad categories: (1) overall evaluation design; (2) household sampling procedures; (3) validity and reliability issues; (4) approaches used to address causality; (5) data collection instruments; (6) implementation; and (7) dealing with disruptions.

To a large extent, this particular set of categories flowed from the evaluations themselves rather than from a preconceived model of impact evaluations.

THE INTERVENTIONS

Oral rehydration therapy (ORT)[2] and vaccinations[3] were the most frequent child survival interventions in the eight programs, but Vitamin A, breastfeeding and, to a very limited extent, growth monitoring and child spacing were each included in one program. (The child spacing activities were not evaluated.) Some of the programs carried out comprehensive social marketing activities that included mass media campaigns to bolster demand as well as efforts to insure adequate supply of competent health personnel and products (Central Java, Ecuador, Metro Manila, Philippines and West Java).

Other programs focused on particular communication strategies rather than a comprehensive approach (Jordan, Lesotho, Zaire).

Activities were grouped into five types of communication strategies:

(1) Mass media campaigns (usually radio but sometimes television);

(2) Training health workers;

(3) Strengthening health worker outreach programs, including village volunteer programs;

(4) Improving delivery at health centers; and

(5) Improving the reliability of product supplies.

All of the programs employed the mass media and training health worker strategies. Fewer strengthened outreach, improved delivery at health centers, or increased the reliability of supply. Healthcom I activities were generally part of a larger health program of the host government, and sometimes were attached to a larger USAID supported project (Ecuador, Central Java, Lesotho). Four programs were national in scope (Ecuador, Jordan, Lesotho,Philippines) and four were limited to regions of the country (Central Java, West Java, Metro Manila, Zaire).[4]

Table 1 shows the child survival interventions and communication strategies used by each program.

[2]ORT refers to a comprehensive treatment plan for diarrhea that includes the appropriate use of oral rehydration accompanied by proper feeding during and after the diarrhea episode and appropriate referral. ORS refers to pre-packaged sugar/salt powder. SSS refers to home-made solutions of sugar and salt.

[3]This paper uses the term vaccination rather than immunization to refer to the act of giving vaccines to children and to indicators reflecting the percentage of the child population that has received the vaccines (coverage). Vaccination is not synonymous with immunization. Not all vaccinations result in the sero-conversion necessary to assure immunization, especially when frequent breakdowns of the cold chain render the vaccine ineffective.

[4]This paper uses the term program in reference to the eight large health programs carried out by the host countries, the term intervention to refer to the types of child survival interventions that comprise the programs (eg., ORT, vaccinations, vitamin A capsules, breastfeeding promotion), the term strategy to refer to the communication strategies employed in the interventions (e.g., mass media, training, outreach), and the terms project and evaluation to refer to the impact evaluation designs and activities carried out by Healthcom I.

BEST AVAILABLE COPY

TABLE I
Healthcom I Projects by Type of Intervention and Communication Strategy

Communication Strategy	ORT	Vaccination	Vitamin A Capsules	Breastfeeding	Growth Monitoring
Social marketing	E,W	E,M,P	C		
Mass media	C,E,W,Z	E,L,M,P,Z	C	J	
Train health workers	C,E,L,W,Z	L,M,P,Z	C	E,J	E
Outreach system	W,Z				
Health centers	W	M,P			
Product supply	E	E	C		

TABLE I. Table I shows that six of eight projects addressed ORT and/or vaccination, the two most frequent child survival interventions. The most frequent communication strategies were mass media (radio and television) and training of health workers, although four projects used a comprehensive social marketing strategy. Note that social marketing usually includes several of the strategies listed below it in the table. Fewer Healthcom I projects attempted to enhance the outreach system, improve the health centers or improve product supply. Capital letters denote distinct projects, namely, "E" = Ecuador, "C" = Central Java, "W" = West Java, "J" = Jordan, "L" = Lesotho, "M" = Metro Manila, "P" = Philippines national, "Z" = Zaire.

EVALUATION DESIGN

Three primary objectives of impact evaluations are to learn:

(1) The extent to which desired changes (improvements) have occurred during or after an intervention:

(2) The extent to which these changes were caused by the intervention; and

(3) Whether the intervention caused unanticipated positive or negative changes.

In this paper these three objectives are referred to as change, causality, and unanticipated effects. In practice, it is frequently the case that one or more of these objectives are neither planned nor achieved, due to limitations of access or resources. The Healthcom I evaluations focused on the change and causality objectives and generally did not address unanticipated effects.

Various strategies were employed by Healthcom I to measure change and establish causality. This section reviews broad characteristics of the evaluation strategies. Details of the strategies are left to later sections. The broad characteristics include: performance indicators (e.g., knowledge, access, individual behavior, coverage, morbidity, mortality); the target population (individuals who are the primary recipients of the communication efforts and for which change

is to be measured); **triangulation** (use of two or more independent data sets and/or arguments to ascertain change or causality); **timing** of measurements in relation to interventions; and use of **controls** (which includes pure controls not influenced by the intervention, dosage effect controls that are partially exposed to the intervention, internal models of causality, tagged outcomes to isolate the presence of the intervention, and adjustment of differences in key characteristics of the various samples). These strategies are described in general in most standard works on evaluation (2).

Performance Indicators

The Healthcom I evaluations measured many variables, including product distribution, access to goods and services, exposure to messages and services, consumer and health worker knowledge, individual consumer behaviors, provider behavior, system behavior (such as coverage) and diarrhea morbidity. Only a few of these were used to indicate performance. The most frequently used indicators of performance in the Healthcom I evaluations were changes in the behavior of consumers or the behavior of systems. Table 2 summarizes the key variables and performance indicators used by the eight evaluation projects according to type of intervention.

With regard to ORT interventions, the most frequently used performance indicators were: ORS/SSS usage in last case or recent case; ORS/SSS usage by severity of diarrhea; proportion of caretakers with knowledge to properly prepare ORS or SSS; and health worker knowledge of diarrhea and ORT. Note that vaccination coverage and ORS use rates reflect a system behavior rather than just a mother/child behavior, because health providers give the vaccinations and health center ORS use rates contribute to overall use rates. Diarrhea morbidity was not used as an indicator of performance, although the proportion of households experiencing a case of diarrhea in the past two or four weeks was reported for two programs (Lesotho and West Java).

With regard to vaccinations, full coverage (all antigens) was generally used as the performance indicator, although varying definitions of the appropriate age window were used by different projects. For example, some evaluations counted all children in the survey under two years of age who were vaccinated at the time of the survey. Other counted all children in the survey aged 12-23 months who were vaccinated by the time of the survey, or alternatively, by their first birthday. Secondary indicators of vaccination program performance included reduction in drop-outs and lowering the average age of vaccination. The vitamin A capsule intervention in Central Java also used coverage.

BEST AVAILABLE COPY

TABLE 2
Key Variables and Performance Indicators for Healthcom I Communication Strategies

Interventio	--- Variables ---							Key Performance
	Distribution	Exposure	Knowledge	Access	Individual Behavior	System/Provider Behavior	Morbidity	
Vaccination								
E		Radio TV	Purpose Proper age Number			Coverage by age		Full coverage 18m Full coverage 12m Equity of coverage
L		Radio	Purpose Symptoms		Clinic attendance Show cards	Coverage (C+R)		Full coverage 12-23m Full coverage 12-59m
M		Radio TV	Purpose Clinic day Side effects		Cards to clinic	Coverage (C+R) Missed opportunities		Measles coverage 12-23m Full coverage 12-23m Timeliness of coverage
P		Radio TV	Purpose Clinic day			Coverage (C+R)		Measles coverage 12-23m Full coverage 12-23m Timeliness of coverage
Z		Radio	Purpose Disease names Proper age			Coverage (C+R)		Full coverage 12-23m Timeliness of coverage
ORT								
E	ORS packets	Radio TV	Purpose ORS Mix ORS (demo)			ORS usage (last case and ever)		ORS usage (last case) ORS usage by severity
C		Radio Banners Volunteers Health Wkrs	Give fluids		Fl[illegible]e (last case,	Recommend fluids, foods, ORT		Fluid usage (last case)
W		Radio Film Cards Volunteers Health Wkrs	Purpose ORS Mix ORS (tell, demo)		Fluid usage Breastfeeding Appro feeding (prior 2 wks)	ORS usage (prior 2 wks) Knowledge+practice volunteers health wrkrs providers	Diarrhea	ORS usage (2 wks) Fluid usage (2 wks) Breastfeeding (2 wks) Approp feeding (2 wks) Volunteer knowledge

TABLE 2
Key Variables and Performance Indicators for Healthcom I Communication Strategies

Interventio	--- Variables ---							Key Performance
	Distribution	Exposure	Knowledge	Access	Individual Behavior	System/Provider Behavior	Morbidity	
L		Radio Brochure	Purpose ORS Symptoms Mix ORS, SSS (claim, tell)		SSS usage (last case, ever, recall)	ORS usage (last case, ever, recall)	Diarrhea	ORS/SSS usage by location (last case, recall)
Z		Radio	Aware SSS Purpose SSS,ORS Mix SSS (tell) Use SSS (tell)		SSS usage (last case, ever, recall)	Source of treatment ORT usage (last case, recall)		
Vitamin A Capsules								
C		Radio Banners Volunteers	Where and when to get capsules Proper age Awareness	Capsules Health posts	Attendance at health posts	Coverage		Coverage 1-5 yrs
Breastfeeding								
J		Radio TV	Early start Time to initiate first foods	Birth location	When initiated (last birth,recall) When foods started	Initiation in hospitals		Initiation time by location Time of first foods

TABLE 2 shows different variables along the continuum from distribution to morbidity that were measured by the Healthcom I evaluations, by type of intervention and by project. Exposure, knowledge and individual and system behavior were measured in most projects, while distribution, access and morbidity were measured only infrequently. "Key performance indicators" are measured variables used by the different evaluations to judge the success of the interventions and communication strategies and vary from project to project. The project codes are: "E" = Ecuador, "C" = Central Java, "W" = West Java, "J" = Jordan, "L" = Lesotho, "M" = Metro Manila, "P" = Philippines national, "Z" = Zaire. Vaccination coverage (C+R) refers to coverage as measured by cards (C) plus recall (R).

Target Population

The target population included both the primary caretaker of young children (mostly the mother) and the young children themselves. The survey interviewed the caretakers, but based the analysis on variables related to both the caretakers and the children. Thus, some analysis refers to a sample of caretakers and other analysis refers to a sample of children.

Different definitions of young were used, including under two years (Jordan, Metro Manila, the Philippines), under three years (Zaire), or under five years (Central Java, Ecuador, Lesotho, West Java), depending on the project and intervention. In addition, one project included only women aged 35 or less (Jordan), and two others only primary caretakers from poor households Metro Manila, the Philippines). This information is summarized in Table 5.

Triangulation

Triangulation can increase the reliability of the evaluation, in that it tests whether different evaluation designs arrive at the same conclusion. Triangulation was formally articulated as an important component of Healthcom I evaluations at the beginning of the project (11), and it was applied in many of the evaluations.

Triangulation during the data collection phase (obtaining different types of data from independent sources) was separated from triangulation during the analysis phase (making different arguments about causality). None of the projects ran two totally separate evaluations. During the data collection phase, information on the number of vaccines delivered or number of ORS packets given out was sometimes obtained as a check on the household survey data that were always collected. Several arguments were also frequently used to establish causality.

Timing

The timing of the measurements in relation to the intervention is not only vital for measuring changes, but can also contribute to or hinder arguments about causality. On one hand, closely spaced measurements taken just before and just after an intervention present a stronger case for causality than more widely spaced measurements. This is because fewer alternative explanations for the change could have occurred in the shorter time period. On the other hand, a post measurement that closely follows the intervention is unable to pick up longer term effects of the intervention, either positive or negative.

A variety of schedules are possible, from before and after measurements that tightly bracket the intervention to a long sequence of frequent measurements that begin well before the intervention and extend well beyond it in time. Sometimes measurements are obtained retrospectively about events in the past from recall, written records or other evidence. Periodicity is often important, because bias can be introduced if measurements are not timed to account for cyclical fluctuations in key indicators.

Table 3 summarizes the household survey schedules of the eight Healthcom I evaluations. Among the evaluations, six completed two household surveys (Central Java, Jordan, Lesotho, Philippines, West Java, Zaire) and two completed three household surveys (Ecuador and Metro Manila). (The first survey in the Ecuador program was completed in two distinct stages several months apart and it is sometimes referred to as the first and second surveys in other documents.) The fact that a minimum of two surveys were completed in all eight evaluations is a notable achievement in light of the many problems encountered. The time between the pre- and post-surveys varied from three and one-half to 28 months. The post survey occurred as soon as three months and as long as 28

BEST AVAILABLE COPY

Previous Page Blank

TABLE 3
Timing of Interventions and Household Surveys

Project and Type of Intervention [1]	Date of Pre Survey Completion	Date of Post Survey	Intervention Start Date	Intervention End Date	Duration Between Surveys (mths)	PreSurvey to Start of Intervention (mths)	Intervention Start to PostSurvey (mths)
M (v)	1/88	4-5/88	2/88	5/88	3.5	1	2-3
E[2] (v,d)	12/85	7/86	10/85	5/88	7	(2)[3]	9
		4/87			16		18
C (d,a)	10/88	10/89	6/88[4]	ongoing	12	(4)[3]	16
P (v)	8/89	8/90	3/90	9/90	12	7	5
Z (v,d)	3/89	10/90	5/89[5]	9/90[7]	19	2	17
			7/89[5]			4	15
			7/90[5]			16	3
J (b)	9/88	8/90	3/89	5/89	23	6	17
W (d)	4/88	4/90	1/87[6]		24	(13)[3]	39
			7/89[6]	7/90[6]		8	9
L (v,d)	11/87	3/90	11/87[7]	9/90[7]	28	0	28

TABLE 3 lists the dates when interventions started and ended, and when household surveys were carried out. Programs are listed from shortest to longest duration between the pre and post household surveys. [1] Capital letters denote distinct projects ("E" = Ecuador, "C" = Central Java, "W" = West Java, "J" = Jordan, "L" = Lesotho, "M" = Metro Manila, "P" = Philippines national, "Z" = Zaire), and small letters in parentheses denote types of interventions ("v"=vaccination, "d"=diarrhea/ORT, "a"=vitamin A, "b"=breastfeeding). [2] In E project, all three surveys were used to evaluate the ORT intervention, while retrospective data from the 4/87 survey were used to evaluate the vaccination component. [3] Negative numbers are in parentheses, indicating the pre-survey was accomplished after the program was started. [4] In C project, health worker training started in 6/88, and the mass media campaign started in 8/88. [5] In Z project, the training intervention started in 5/89, the vaccination mass media component in 7/90, and the diarrhea/ORT intervention in 7/89. [6] In W project, health worker training began in four regencies in 1/87 and in eight regencies in 4/89; but the mass media intervention did not start until 7/89. [7] Approximate date.

months after the start of the intervention. In six of the eight programs, the communication activity was in operation for more than a full year before the post survey was done.

There was also wide variation in the time from the pre-survey to the start of the intervention, ranging from 16 months (the longest) to minus four months (the pre-survey followed the intervention start date by four months). In two programs (Central Java and Ecuador), the pre-survey was not done until after the intervention had started. Thus, early effects of the intervention were not captured. In Zaire, the vaccination intervention was delayed and finally started 16 months after the pre-survey had been completed. This left a long duration when other events may have influenced the outcome. These variations do not appear to be related to the type of intervention nor to the amount of improvement measured. (Table 6 contains measured increases in vaccination coverage and ORS usage.)

Clearly, the survey schedules were not always optimal as a result of these problems. Ecuador provides an interesting example of how one program managed its way around scheduling difficulties. The survey schedule started out in trouble because the intervention was started two months before the pre-survey could be implemented. To further complicate matters, the initial survey (12/85), which was designed and implemented under time pressure, later proved to lack reliability and adequate rural representation. To correct these deficiencies, a supplemental rural survey was conducted in 5/86 to create a nationally representative sample when combined with the 12/85 sample. Two additional surveys were planned and implemented (7/86 and 4/87). Lessons from the 12/85 survey were used to redesign the questionnaires for the 7/86 and 4/87 surveys, making them more reliable and comparable. As a result, the 7/86 and 4/87 survey data were used to establish trends in ORS usage, which was significant in spite of the relatively short duration between the two surveys. Another strategy altogether was used for the vaccination intervention in Ecuador. Retrospective data based on vaccination cards and mothers' recall were obtained in the 7/87 survey and the data from that single survey were analyzed extensively in order to estimate change in vaccination coverage.

Seasonality can be a problem when measuring the effect of ORT interventions, and as a result, it is generally preferable to hold pre- and post-surveys during the same season of the year, preferably the high season for diarrhea. As can be seen in Table 3, two of the five ORT interventions held pre- and post-surveys in the same season (Central Java and West Java), while three of the five held the pre- and post-surveys in different seasons (Ecuador, Lesotho, Zaire).

Controls

The primary use of controls relates to the causality question—what were the causes of any observed change, especially did the intervention cause the change (or, in the case of negative or no change, did the intervention prevent things from getting worse than they did). In some instances, controls are used to help determine whether observed changes really occurred, rather than being the result of a random fluctuation or an artifact of the measuring procedure. When observed changes are not artifacts, controls are needed to establish the role of the intervention in producing the change.

Healthcom I used a wide variety of control strategies, some of which were pre-planned and others which were developed during the course of the program or during the data analysis. The most important strategies are identified below and summarized in Table 6. Their application in the various projects is discussed at more length in a later section.

Prior control group (pre-post). All eight projects used a prior control group, although some were started after the intervention as noted above.

BEST AVAILABLE COPY

Pre-post in geographic proximity. In this strategy, the pre-sample is supposed to be the control for the post sample. Because the pre-sample is drawn from the same or a nearby community, it is likely to be relatively similar to the post-sample. One can be reasonably sure it has not been influenced by the intervention, except when the pre-survey follows the start of the intervention. A special case of this strategy is a time series design with more than one pre- or post-survey. (This strategy does not include panel designs in which the same individuals are surveyed pre- and post-intervention.)

Concurrent control group. Some programs planned and collected data on separate control groups which were as similar to the activity sample as possible, except of course for the intervention. The challenge was to find a separate control population that was neither exposed to nor influenced by the intervention.

Adjust differences in groups using multivariate analysis. Some programs went further, testing whether the program and control samples were similar on a whole host of socio-economic variables, and if not, weighing the samples in order to make them more similar with respect to those variables. This approach can also be applied to differences between different program samples.

Short duration; no other explanation. These refer to the timing strategy described above, rather than to controls.

Correlation in time. This refers to increases or decreases in the performance indicators that closely correspond to the timing of the intervention. This approach often uses time series data and includes observations about plausible alternative explanations. (A special case of this strategy is a reversible design in which the intervention is purposely withdrawn and reintroduced. The reversible design, however, was not used by Healthcom I.)

Tagged outcomes. In this strategy, certain goods or services promoted by the intervention are uniquely tagged and thus traceable back to the intervention. For example, a mass media campaign urges mothers to bring their children to the clinic to be vaccinated on a particular day of the week, and the evaluation analyzes the changes by day of the week.

Internal models. This approach attempts to track the causal pathways leading from the intervention to the outcomes. It requires measurements of the internal variables along the hypothesized path. For example, a high correlation between exposure to a message, knowledge of its contents, and desired behavior change is evidence that the message caused the desired change. Many different plausible pathways are possible, however, so that a negative finding with respect to one path does not rule out the others. The Healthcom I evaluations looked at many different possible internal models in their attempt to establish causality. The different internal models are discussed in more detail in Section 6 and in the notes to Table 6.

SAMPLING

All eight evaluations completed two or more household surveys. The surveys all used cluster sampling techniques to select the sample of households, restricted the sample to households with young children and interviewed the caretakers (mostly mothers) of those children. However, within this general structure, the approaches used by the evaluations differed widely. These differences are summarized in Tables 4 and 5.

The definition of young children varied from under two years to under five years. In addition to restricting the sample to households containing young children, some evaluations placed other restrictions. The program in Jordan restricted the sample to households with mothers 35 years old or younger. Two programs (Metro Manila and the Philippines) purposely selected a preponderance of poor households.

Sampling occurs within a frame that defines the set of all households that are eligible to be selected for the sample. The Healthcom I evaluations generally matched their sampling frame to the area covered by the intervention. Three were national (Ecuador, Jordan, Lesotho), three were limited to several major regions of the country (Central Java, West Java, Philippines), and two were in one major urban area (Metro Manila, Zaire). One project (Central Java) included two regencies in the first stratum of both the pre- and post-surveys, one reached by the intervention (the program group) and one that was not reached by the intervention (the simultaneous control group).

Cluster sampling requires two or more steps. In a two-stage cluster sample of households, the first stage selects clusters (such as villages or neighborhoods) and the second selects households within the clusters. The procedure for selecting the clusters can also occur in stages or strata, in which large geographic regions are sampled, and then sub-regions, etc. until the final clusters are selected. The number of strata used by the eight evaluations varied from two to five. The five strata in the Central Java evaluation are described in Note 3 of Table 4.

Many different procedures can be used to select the sub-areas within each stratum and the households. Ideally, the procedures select households randomly from among all eligible households in the sampling frame with equal probability. This is difficult if an accurate census of the cluster is not available, and so approximate methods are often used. An important issue is what to do when no one is home at a selected household. Although it is usually best to survey the same number of households in every cluster, this is not always done for practical reasons.

The selection procedures used by the Healthcom I evaluations include purposive, probability proportional to population (PPP), simple random sampling (SR), and approximate random sampling (AR). The first stratum is often purposive. As noted above, two evaluations (Central Java, West Java) purposely separated the first strata by program and control areas. Two other evaluations (Jordan, Lesotho) purposely separated the first strata into urban and rural areas. Another evaluation (Metro Manila) selected areas near health centers. The use of the PPP method requires an estimate of the population in each sub-area, usually obtained from census data. At least three evaluations (Central Java, West Java, Philippines) used PPP in the intermediate stratum. The advantage of PPP is that the sample is statistically more representative of the population than with other methods. Simple random selection (SR) is used in lieu of PPP when a complete list of sub-areas or households is available but the populations are not known or are known to be approximately equal, such as with households. This method selects each area or household from the list with equal probability until a predetermined number are selected.

Approximate random (AR) methods are used when a complete list is not available. This is often the case in the final stratum, when particular households are finally selected. For example, a frequently used AR method is to select a starter household randomly from among a group of households near the center of the cluster, and then survey all the eligible households in some random direction until the desired sample size for the cluster is satisfied.

Several evaluations (Central Java, West Java, Zaire) used exactly the same sub-areas in their pre- and post-survey samples in order to increase the comparability of the pre- and post-samples. In the pre-survey, the areas were selected using a random method (PPP or SR).

BEST AVAILABLE DOCUMENT

TABLE 4
Procedures Used in Cluster Sampling for Healthcom I Household Surveys

Project[1]	Survey[2]	Concurrent Control Group	Sampling Frame	Total Hseholds in Sample[2]	Number of Strata	Total Sub-areas by Strata [3]	Method Used to Select Sub-areas and Hseholds [4]	Approx. HHs per Cluster	Second survey paired to first	Census data used in sample
E	1st[5]	No	Limited[5]	1,474	2	60		25		
	2nd		National	2,702						
	3rd		National	1,460						
C	1st	Yes	2 regencies	799	5	2/10/50/50[6]	purp/PPP/PPP/SR/SR[7]	16	Yes	Yes
	2nd	Yes		791			same/same/same/SR/SR			
W	1st	No	4 regencies	1000	4	50/50/100	PPP/SR/SR/SR	10	Yes	Yes
	2nd	No		1000			same/same/SR/SR			
J	1st	No	National	930	3	2/50	purp/SR/AR[8]	11-23	No	Yes
	2nd			966			purp/SR/AR[8]	16-22		
L	1st	No	National	1,040	3	2/40[9]	purp/SR/AR[9]	25/35[9]	Yes	Yes[8]
	2nd			1,016			purp/adjacent/AR[9]			
M	1st	No	City of Manila	320	4	8/8/32	purp/-?-/purp/AR[10]	10		
	2nd			600				6		
	3rd			600				6		

TABLE 4
Procedures Used in Cluster Sampling for Healthcom i Household Surveys

Project[1]	Survey[2]	Concurrent Control Group	Sampling Frame	Total Hseholds in Sample[2]	Number of Strata	Total Sub-areas by Strata [3]	Method Used to Select Sub-areas and Hseholds [4]	Approx. HHs per Cluster	Second survey paired to first	Census data used in sample
P	1st 2nd	No	4 regions	1,200 1,200	4	4/11/120	purp/PPP/SR/ AR[11]	10		No
Z	1st 2nd	No	City of Lubumbashi	1,125 1,153	2	75	SR/AR same/AR[12]	15	Yes	

NOTES: [1] The project codes are: "E" = Ecuador, "C" = Central Java, "W" = West Java, "J" = Jordan, "L" = Lesotho, "M" = Metro Manila, "P" = Philippines national, "Z" = Zaire. [2] Program and control samples are grouped together in this table, so that the total number of households in a survey includes households from both samples. [3] First figure is number of sub-areas within sampling frame (first strata), and so on. For example, in C project, the first strata consisted of a program regency and a control regency, the second strata had five sub-districts per regency for a total of 10, an average of five villages per sub-district for a total of 50, and one neighborhood per village for a total of 50 neighborhoods, and an average of 16 households per neighborhood for a total of 800 households in both the first and second surveys. The last strata, namely the number of households, is not included in this column. [4] Selection methods include: "purp" = purposive, "PPP" = probability proportional to population, "SR" = simple random selection among all sub-areas or households in that strata with equal probability, "AR" = approximately random selection among all sub-areas or households, such as a random starting house, "same" = the exact same sub-areas as used in the earlier survey, "adjacent" = sub-areas adjacent to ones selected in earlier survey. The initial items in each list correspond to the strata. The final item in each list refers to the method of selecting the households. [5] The first survey in E project was done in two stages, an urban sample of 964 households in 1985 and a rural sample of 500 households in 1986. [6] In C project, the program group sampled six villages per sub-district for a total of 30 villages, and the control group sampled 4 villages per sub-district for a total of 20 villages. (See also the example in note [2].) [7] In C project, the first strata is divided into program and control areas. [8] In J project, census blocks were divided by urban and rural, and then 35 urban and 15 rural census blocks were selected randomly. [9] In L project, the first strata was urban and rural, and the second selected 34 rural enumeration districts and six urban enumeration districts randomly in the first survey, and the second survey used enumeration districts adjacent to the ones selected in the first survey. [10] In M project, strata one and three were purposively selected near to health centers. [11] In P project, the sampling frame consists of four regions (Manila, Luzon, Visages, Mindano). Ten "cities" were selected by PPP from the last three areas, when combined with Manila giving 11 cities in strata two. Then a random sample of lower income voting districts in the 11 cities was made, and households selected AR from these voting districts. [12] In the second survey of Z project, 66 of the 75 clusters were the same clusters used in the first survey and nine were newly selected by SR. [13] Blanks mean that no data were found for this item in the sources reviewed.

BEST AVAILABLE DOCUMENT

Two sets of smaller sub-areas or households were selected, one set for the pre-survey sample and the other for the post-survey sample. Thus the households in the two surveys are certain to be in the same geographic area. The program in Lesotho selected sub-areas for its post-survey that were geographically adjacent to the randomly selected sub-areas in the pre-survey. These pairing strategies were done to increase the similarity of the clusters and households in the pre- and post-surveys.

Sample sizes varied from 600 (Metro Manila) to 1,474 (Ecuador), after excluding the small first survey in Metro Manila and the large second survey in Ecuador, as summarized in Table 5. The number of sub-areas in each strata, number of final clusters and approximate households per cluster are summarized in Table 4. The number of final clusters ranged from 32 to 120, and the average households per cluster ranged from six to 25. (Note this is slightly smaller than the 30 households per cluster often used by WHO in their cluster surveys for vaccination coverage.)

The number of caretakers interviewed was approximately equal to the number of households in the sample. Sub-samples were obtained in order to provide information about particular child survival interventions: sample sizes of children aged 12-23 months for evaluation of vaccination programs for four programs (Ecuador, Lesotho, Philippines, Zaire); sample sizes of children aged one to four years for the evaluation of the vitamin A capsule program in Central Java; sample sizes of households with recent (two or four weeks prior to survey) cases of diarrhea or just "last case" of diarrhea for six programs (Ecuador, Central Java, West Java, Lesotho, Philippines, Zaire); and sample sizes of births and of children engaged in complementary feeding for the evaluation of the breastfeeding intervention in the Jordan program.

TABLE 5
Sample Sizes of Healthcom I Household Surveys

Project[1]	Survey[2]	Total	Age	Children	Children	Households with case of diarrhea		
						Last 2 wks[5]	Last 4 wks[5]	Last cases
E	1st	1,474	U5[6]	nd				
	2nd	2,702	U5	nd		893		
	3rd	1,460	U5	1,966		341		
C	1st	799	U5		734[7]			611[8]
	2nd	791	U5		733[7]		146	559[8]
W	1st	1000	U5			167	203	920
	2nd	1000	U5			176	216	893
J[9]	1st	930[10]	U2					
	2nd	966[10]	U2					
L	1st	1,010	U5	288			261	
	2nd	1,016	U5	289			288	
M	1st	320	U2[4]					
	2nd	600	U2[4]					
	3rd	600	U2[4]					
P	1st	1,200	U2	446				722
	2nd	1,200	U2	461				
Z	1st	1,125	U3	418		444		
	2nd	1,153	U3	427		536		951

NOTES: [1] The project codes are: "E" = Ecuador, "C" = Central Java, "W" = West Java, "J" = Jordan, "L" = Lesotho, "M" = Metro Manila, "P" = Philippines national, "Z" = Zaire. [2] Program and control samples are grouped together in this table, so that the total number of households in a survey includes households from both samples. [3] Used in analysis of vaccination coverage. [4] Used in analysis of vitamin A capsule coverage. [5] Includes diarrhea on day of survey. [6] Youngest child in household only. [7] The one to four year old sample in project C includes 454 and 280 children respectively in the program and control groups in the first survey, and 438 and 295 children respectively in the program and control groups in the second survey. [8] The cases of diarrhea sample in project C includes 370 and 244 households respectively in the program and control groups in the first survey, and 401 and 258 households respectively in the program and control groups in the second survey. [9] In its analysis of the breastfeeding intervention, project J used number of mothers giving birth after the intervention started (500 in first survey and 525 in second survey), and number of children complementary feeding (383 in first survey and 409 in second survey). [10] In its analysis, project J restricted its sample to children 0-20 months old (800 in first survey and 777 in second survey). [11] Blanks mean that no data were found for this item in the sources reviewed.

BEST AVAILABLE DOCUMENT

VALIDITY AND RELIABILITY

Validity

Validity refers to the extent to which a variable measures what it purports to measure. Validity errors are almost always biased, and systematically underestimate or overestimate the intended measure. Healthcom I evaluations were frequently concerned with issues of validity and searched for ways to estimate the amount of bias using independent measurements. Unfortunately, large biases were often found that could not always be resolved.

In the eight evaluations, two did not address validity (Central Java, Jordan), five addressed validity issues in vaccinations (Ecuador, Lesotho, Metro Manila, Philippines, Zaire) and three addressed validity issues in ORT (Ecuador, Lesotho, West Java). Two methods were used to measure vaccination coverage: the card method only counts vaccinations that are recorded and dated properly on the child's vaccination card which the caretaker must produce during the interview; and the card plus recall method which counts all card vaccinations plus all vaccinations recalled by the caretaker when no card is available or no date is on the card. Both methods use the full cohort as the denominator.

The five evaluations addressing validity in vaccinations all concluded that the card method seriously underestimates the true coverage, and some concluded that the card plus recall method was probably fairly close to a true estimate. The strongest evidence for this conclusion is found in plots of coverage versus age of child in the Ecuador and Philippines programs, where card plus recall coverage rises steadily but ever more slowly as age increases while card-only coverage stops rising and drops off as age increases. Further, the program in The Philippines found that coverage estimated by the presence of a BCG scar rises steadily in a fashion similar to, but a bit below, the card plus recall coverage, while card-only coverage drops off after six months of age.

Other evidence was offered which indicates that some of the vaccination cards were in error. The Ecuador program reported numerous cases with missing dates (some of which are likely to have been on time) and the Zaire program found 5.5% of the cards reported all vaccinations complete except BCG (a condition it thinks unlikely). The Lesotho program found a large discrepancy between card showing rates reported in the Healthcom surveys (65%) and that reported officially by MOH, CCCD, WHO and UNICEF (90-96%).

The percentage of caretakers who know how to prepare ORS or SSS was obtained in three ways:

(1) Asking the caretaker if she knows how (claimed);

(2) Asking the caretaker to describe the steps in the preparation (described); and

(3) Asking the caretaker to demonstrate the preparation (demonstrated).

A fourth technique—observing the preparation during real life conditions— was used by at least one program (West Java), but not described in the documents reviewed for this paper.

Three programs compared the proportion of women who claimed knowledge of how to prepare ORS against those who could describe it. The results were mixed. One (West Java) found only 48% of those who claimed knowledge could correctly describe how to prepare it. Two others (Ecuador, Lesotho) found 91% and 86% of those claiming could correctly describe how to prepare ORS. Note that correct preparation of ORS does not insure correct giving of the ORS to the child. One program (Lesotho) found only 20% of those claiming knowledge of how to prepare SSS could describe the preparation correctly. None of the drafts reported comparisons between described and demonstrated knowledge of ORS or SSS preparation.

A special study of ORS usage measurement in Healthcom I activities (13) addresses the following validity issues related to the definition and measurement of ORS usage:

- Last case use/ever used;
- Recall period (24 hours, two weeks, four weeks, ever);
- Correct use/use at all;
- One (oral rehydration) liquid/many liquids (e.g., ORS, SSS, tea);
- Child's age (under one/over one); and
- Unaided recall questions/aided recall questions.

The findings of this study are interesting and useful.

(1) Last case usage was consistently 25-40 percentage points below ever used in four programs.

(2) Diarrhea prevalence went down and last case usage went up as the duration of the recall period increased. In one project, as the recall period increased from the day of the survey to the prior two weeks, to the two weeks before that, the incidence of diarrhea cases was 7%, 17% and 8% respectively. Similarly, ORS usage during the last case of diarrhea was 16%, 21% 24% and 40% in one program and 5%, 6%, 8% and 8% in another for the recall period equal to the day of the survey, the prior two weeks, two weeks before that, and more than a month, respectively.

(3) Correct use was lower than use at all, as expected.

(4) Usage of one liquid was lower than usage of many liquids, as expected.

(5) Usage by children under one year of age was marginally lower than by children over one year.

(6) Usage based on aided recall questions was higher than usage based on unaided recall questions by 25% (5-10 percentage points) in two programs.

Reliability

Reliability refers to the precision of the various measurements. Although the degree of reliability does not influence the expected value of a measurement, it can influence conclusions based on statistical significance (less reliable measurements are less likely to produce significant differences). Healthcom I evaluations did not discuss the reliability of their measurements.

MORE ON CAUSALITY

Sections on triangulation, timing and controls identified the broad strategies used by Healthcom I evaluations to establish causality. Table 6 summarizes the use of these strategies by each program. It provides a convenient organization for discussing the application and success of these strategies in the Healthcom I evaluations. Some of the evaluations apply more elegant arguments than can be adequately described in this cross-cutting summary, especially Central Java, Ecuador, and the Philippines.

The first column of Table 6 lists the different arguments used by Healthcom I evaluations to establish causality. In addition, lines 2, 8, 11 and 14 give the approximate change in percentage points (pp) in the key performance indicators during the intervention. All of the increases over 10 percentage points were statistically significant.

BEST AVAILABLE DOCUMENT

Lines 1a through 1e present information on the type of controls used by the programs. All eight evaluations used a pre/post design with the pre-survey serving as a prior control group (line 1a). Line 1b, Pre/post in geographic proximity is a "poor person's" panel design. Five programs used this approach (Central Java, Lesotho, Philippines, West Java, Zaire), generally without problems. Two programs (Central Java, Zaire) used concurrent control groups (line 1c). The strategy identified in line 1d, "adjust differences in groups using multivariate analysis," was also used by five programs (Ecuador, Central Java, Philippines, West Java, Zaire), in order to increase the comparability of pre- and post-samples and of program and concurrent control groups.

Finally, evaluations using triangulation for data collection are noted in line 1e. Several programs interviewed clinic health workers and volunteers (Lesotho, Philippines, West Java, Zaire). Others undertook side surveys of caretakers, one with clinic exit interviews (Philippines) and another with indepth, open-ended interviews (Jordan). The program in West Java interviewed retailers. Ecuador and Lesotho collected central data on production and supply. (Information on triangulation during analysis has not been summarized in this paper.)

The program in Metro Manila argued that the time between the pre- and post- surveys was short, only three months, and the intervention was the only plausible cause for the substantial increase in vaccination coverage (line 3). At the other extreme, in Jordan the pre-survey occurred seven months before the campaign started and the post-survey 15 months after the campaign ended. This longer time period potentially allowed for more alternative causes to have occurred between the two measurements.

The program in Ecuador found that increases in coverage in time series data obtained retrospectively from vaccination cards and recall correlated in time with the start of the intervention (line 4). The only program to use control groups with respect to vaccination coverage (Zaire) found that although coverage in the program group increased, so did coverage in one of five control groups (line 5). Two programs used control groups with respect to ORT usage (Central Java, Zaire)—one negating and one partially supporting the causal connection between the intervention and the change (or lack thereof) in usage (line 9). The program in Central Java also used control groups with respect to Vitamin A capsule coverage (line 12).

Four programs noted that coverage increased substantially during the months of the campaign (Ecuador, Zaire) or the special day of the week promoted for vaccinations (Metro Manila, Philippines), an argument termed tagged outcomes (line 6).

Many different internal models were used in the effort to establish causality, some much stronger than others. The weakest internal models claim that if an intermediate variable thought to be on a chain of causality between the intervention and the performance indicator (such as exposure to the intervention, knowledge, access or trial usage) increases on average in the community, then this is some evidence that the intervention caused whatever increase in the performance indicator was observed (lines 7a,7d,10a,10d,15a). A slightly stronger link is established between the intervention and increased knowledge when respondents specifically identify the intervention as the source of their knowledge, as in Ecuador (line 10b).

A stronger argument can be made when substantial correlations are found between exposure, knowledge and practice for individual respondents, supporting the model that individual exposure increases individual knowledge which improves individual practice. Five programs (Ecuador, Lesotho, Metro Manila, Philippines, Zaire) used some form of this

argument for vaccination coverage (line 7a), three (Ecuador, West Java, Lesotho) used it for ORS usage (line 10c), Central Java used it for vitamin A coverage (line 13b), and Jordan used it for breastfeeding practices (lines 15a,b). It was the most widely used type of argument and possibly not explicitly part of most of the original evaluation plans.

When applied to vaccinations, this argument can be attacked on the grounds that increased vaccination visits might be the cause of greater knowledge rather the other way around. The program in Lesotho countered this attack with its finding that exposure and knowledge were correlated for clinic non-attenders who had never been exposed to training from the clinic. Ecuador also considered, and found support, in alternative causal models involving community knowledge and community expectations as key intermediate variables between community exposure and individual practice (line 7c).

Further discussion of the internal models can be found in other reports (9,10,11,15).

BEST AVAILABLE DOCUMENT

TABLE 6
Arguments Used to Establish Causality

	--- Project[1] ---							
	E	C	W	J	L	M	P	Z
1. Types of Control ("CG" = control group):								
a. Prior CG (pre-post)	yes	yes	yes	yes	yes	yes	yes	yes
b. Separate CG		yes	no					yes
c. Pre-Post in geog proximity		yes	yes		yes		yes	yes
d. Triangulation (data)	yes		yes	yes	yes		yes	yes
e. Adjust differences in groups using multivariate analysis	yes	yes	yes				yes	yes
Vaccination Coverage								
2. Coverage increase(appr) [2]	20pp	////////////	////////////	////////////	none	15pp	12pp	3pp
3. Short duration; No other explanations		////////////	////////////	////////////		support		
4. Correlation in Time	support[3]	////////////	////////////	////////////				
5. Control group evidence		////////////	////////////	////////////				partial support
6. Tagged outcomes (day or month of vaccination)	support(m)	////////////	////////////	////////////		support(d)	support(d)	support(m)
7. Internal models: [4]		////////////	////////////	////////////				
a. E hi; K up		////////////	////////////	////////////		K up	E hi; K up	
b. E,K,P correlate	support,EKP	////////////	////////////	////////////	support,EKP	weak,EK	weak,EK	negate,EK
c. Communicate K and expect correlation	support	////////////	////////////	////////////	[5]			

TABLE 6
Arguments Used to Establish Causality

	--- Project[1] ---							
	E	C	W	J	L	M	P	Z
d. Clinic supply + prom up		//////////	//////////	//////////			negate	
e. Missed opportunities down		//////////	//////////	//////////		support		
ORS/SSS/Fluid Usage								
8. Usage increase (approx) [3]	15pp	17pp	none[6]	//////////	19pp	//////////	//////////	12pp
9. Control group evidence		negate (CGs up)	//////////	//////////		//////////	//////////	partial support[7]
10. Internal models:				//////////		//////////	//////////	
a. E hi/K up/P(trial use) up	support	negate[8]	support	//////////	support	//////////	//////////	
b. K up/project named	support			//////////		//////////	//////////	
c. E,K,P correlate pos	support,EKP		weak[9]	//////////	support,EK	//////////	//////////	
d. Packet distr up	support			//////////		//////////	//////////	
Vitamin A Capsule Coverage								
11. Coverage increase (approx)	//////////	small	//////////	//////////	//////////	//////////	//////////	//////////
12. Control group evidence	//////////	support (CG constnt)	//////////	//////////	//////////	//////////	//////////	//////////
13. Internal models:	//////////		//////////	//////////	//////////	//////////	//////////	//////////
a. E hi/K up/access(A) up	//////////	support,EKA	//////////	//////////	//////////	//////////	//////////	//////////
b. E,P correlate pos	//////////	negate	//////////	//////////	//////////	//////////	//////////	//////////
c. Health post attendance,Coverage correlate	//////////	support	//////////	//////////	//////////	//////////	//////////	//////////

BEST AVAILABLE DOCUMENT

TABLE 6
Arguments Used to Establish Causality

	--- Project[1] ---							
	E	C	W	J	L	M	P	Z
Breastfeeding Practice								
14. Change in practice:	////////////	////////////	////////////		////////////	////////////	////////////	////////////
a. Early initiation increase [3]	////////////	////////////	////////////	20pp	////////////	////////////	////////////	////////////
b. Approp first foods increase	////////////	////////////	////////////	none	////////////	////////////	////////////	////////////
15. Internal models:	////////////	////////////	////////////		////////////	////////////	////////////	////////////
a. E hi/K up	////////////	////////////	////////////	support	////////////	////////////	////////////	////////////
b. E,K,P correlated	////////////	////////////	////////////	support	////////////	////////////	////////////	////////////

TABLE 6 summarizes the arguments used to establish causality, by type of intervention and project. Approximate increases in coverage, usage and practice deemed to be caused by the communication strategy are presented in lines 2, 8, 11 and 14. [1] Project codes are: "E" = Ecuador, "C" = Central Java, "W" = West Java, "J" = Jordan, "L" = Lesotho, "M" = Metro Manila, "P" = Philippines national, "Z" = Zaire. [2] "pp" = percentage points. [3] In project E, line 4, coverage for children with completed immunization cards showed an increase during the period of the program. [4] In reference to internal models in lines 7, 10, 13 and 15, "E" = exposure, "K" = knowledge, and "P" = practice (coverage or usage). [5] In project L, line 7b, E and K are positively correlated for clinic attenders, supporting the hypothesis that the causal direction is from E to K to P rather P to E to K. [6] In project W, line 8, usage increased for children perceived to be very sick but not for other children. [7] In project Z, line 9, in four of five prior CGs coverage remained constant while coverage increased significantly in one prior CG. [8] In project C, line 10a, P remained low while K increased in both the program and control groups, indicating that the program was not the cause of the increase in K. [9] In project W, line 10c, E and K correlate positively, but E, K and P (trial usage) correlate only in children perceived to be very sick. [10] Blank cells reflect arguments not used in that project; crosshatched cells reflect interventions not applied in that project.

DATA COLLECTION INSTRUMENT

The nature of the data collection instrument influences the reliability of the data and the cost of data collection. Three attributes of the instrument were noted in some of the evaluation reports:

(1) the length (in number of questions);

(2) whether a written translation to the local language was done; and

(3) whether it was pretested in field conditions.

With respect to the length, one evaluation (Lesotho) reported 144 questions, a second (West Java) reported that the questionnaire was extended, and a third (Zaire) indicated that each interview took 30-40 minutes. Data were not found on this issue for the other five evaluations. Six (Ecuador, Central Java, Lesotho, Philippines, West Java, Zaire) reported the questionnaire had been translated into the local language, translated back into English, and pretested in the local language before being printed in the local language (21).

IMPLEMENTATION PROCEDURES

Information on procedures for implementing the data collection include: the number and background of the interviewers; the amount and type of training received by the interviewers; the organization of teams and supervision in the field; the amount of time spent per interview and in total; and data compilation procedures. Information about implementation of the evaluations was sparse.

Three evaluations provided information about who carried out the surveys. The two Indonesian programs reported using a local firm to develop and implement the sampling plan, translate and pretest the questionnaire, collect data in the field and produce clean copies for analysis. The evaluation of the program in Jordan was done entirely by the counterpart organization.

Three evaluations provided information about training the interviewers. Both the Indonesian projects provided four days of training for the 12 interviewers and two supervisors that included classroom presentations, role playing and field practice. The training of the 18 interviewers in Zaire included role playing. This evaluation also indicated that the 18 interviewers were organized into six teams of three each. Each team had a leader who checked the material, established schedules and routes, and verified responses at the end of the day. Each interview took 30-40 minutes to complete, each team did about 15 interviews per day, and in all it took 12 days of work to collect the data. Data entry to computer was performed in the provincial capital of Lubumbashi.

DISRUPTIONS

Many of the evaluations experienced problems outside their control. Some of the types of problems experienced included interventions that were begun before the baseline study or first survey, reduced funding that caused a decrease in planned sample size, or final surveys not done due to lack of funds or lack of access. These unforeseen disruptions caused the adoption of alternative patching strategies to try to measure change or establish causality.

For example, the program in Ecuador had planned to do at least one additional survey at the end of the intervention which was canceled, a disruption that led to the extensive use of retrospective data. The intervention in Central Java was started before the first survey was undertaken, and the West Java program began its expansion prior to the first survey, leading to reliance on concurrent control groups to estimate change in Central Java.

BEST AVAILABLE DOCUMENT

SUMMARY DISCUSSION

Although the eight impact evaluations were done in four different parts of the world in relation to several different types of child survival interventions and communication activities under a wide variety of conditions, they had several important characteristics in common. The interventions and evaluations were all large scale (often national in scope). How the evaluations were conducted was relatively well documented. The evaluations were successful in the sense that they yielded convincing arguments about the success of the interventions.

The rarity of this data set suggests the question, Why were these evaluations successful? This paper, based primarily on the written final evaluation reports, undertakes to answer this question, but a fuller understanding requires deeper probing.

A common problem faced by many of the evaluations was instability. Evaluations plans were disrupted by unforseen events. One manifestation of this problem is the variation between the start and end of the interventions and the pre- and post-measurements among the different evaluations. In several cases, pre-measurements occurred after the interventions. In general, schedules were disrupted. Another manifestation is the three programs that did not complete impact evaluations. Such instability, which is common in large scale evaluations, adds to the interest of the central question.

All of the evaluations defined success in terms of a few carefully selected behaviors, rather than in terms of knowledge or morbidity/mortality. Most of the evaluations used pre- and post-measurements to ascertain change in the key indicators. The exceptions occurred in some of the vaccination interventions where the pre-measurement was discarded in favor of retrospective data from vaccination cards and mothers' recall obtained in post measurements. The measurements all used cross-sectional multi-stage cluster samples. The number of clusters and method of selecting the clusters varied across evaluations and stages. No panel designs were used, although a variety of techniques were used to obtain clusters and samples in the pre- and post-measurement that were comparable.

The approaches used to establish causality were not consistent across the evaluations. Many different techniques were used. Simultaneous control groups were used in some evaluations but not in others. In part, this was because national programs make it difficult to find post intervention control groups. Many of the analytic techniques used to establish causality appeared to emerge from the data and experiences of the interventions and evaluations themselves. They did not appear to be preplanned, except in broad outline. In other words, the key to the analysis of causality appeared to be the use of considerable technical expertise and imagination.

Triangulation, the strategy of using two or more independent data sources and independent methods of analysis to reach conclusions about change and causality, was identified early as a desirable approach. Several of the evaluations used this approach to some extent, although extensive use is not reported in the evaluation reports.

Many issues were raised with respect to the validity of the key indicators and their measurements. These issues were addressed thoroughly in several special studies that cut across evaluation sites rather than in each evaluation.

Sample sizes were substantial and questionnaires lengthy. Data collection was done mostly by host country personnel, but the analysis received heavy input from senior technical advisors from Healthcom. As a result, the cost of each evaluation was substantial.

BEST AVAILABLE DOCUMENT

Previous Page Blank

The Healthcom I evaluations demonstrate that it is possible to carry out successful impact evaluations on a variety of large-scale child survival and communication interventions in moderately unstable environments in developing countries. The evaluation strategies included a mixture of standard techniques (e.g., multi-stage cluster sampling), preplanned approaches adopted to each case, and imaginative adaptation to overcome the inevitable disruptions that arise during the evaluation.

REFERENCES

1. Cabanero-Verzosa C, Bernaje MG, de Guzman EM, Nernandez JRS, Reodica CN, Taguiwalo MM. *Managing a Communication Program on Immunization*. Department of Health, Republic of the Philippines and Healthcom Project, Academy for Educational Development, Washington DC. December 1989.

2. Cook TD and Campbell DT. The design and conduct of quasi-experiments and true experiments in field settings. In Donnette MD, *Handbook of Industrial and Organizational Psychology*, Chicago: Rand McNally College Publishing Co. 1976.

3. Foote DR. "The Healthcom Project in Paraguay: A Case Study." Healthcom Project, Academy for Educational Development, Washington DC and Applied Communication Technology, Menlo Park. 1989.

4. Healthcom. "Behavioral Research in Child Survival: Studies from the Healthcom Project." Academy for Educational Development, Washington DC. April 1992.

5. Healthcom. "Institutionalizing a Methodology for Public Health Communication: A Midproject Report." Academy for Educational Development, Washington DC. (No date.)

6. Healthcom. *Lessons from Evaluation Studies*. Academy for Educational Development, Washington DC. (No date.)

7. Hornik R, McDivitt J, Yoder PS, Zimicki S, Contreras-Budge E, McDowell J, Rasmuson M. "Communication for Child Survival: Evaluation of Projects in Ten Countries (Draft)." Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 10, 1992.

8. Hornik R, Contreras-Budge E, Ferenic N, Koepke C, Morris N, Torres MP, Pareja R, Smith W. "Results from the Evaluation of the PREMI/Healthcom Project in Ecuador 1985-1988." Healthcom Project, Academy for Educational Development, Washington D.C. and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 1991. (Ecuador program evaluation report)

9. Hornik R, Zimicki S, Less MB, Hernandez JR, de Guzman EM, Verosa CC, Abad M, Fausto A, Pabalan MC, Bagasao TMP, Pilipino KP. "The Healthcom Project in the Philippines: The National Urban Immunization Program 1989-1990." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 1991. (The Philippines program evaluation report)

10. Hornik R. "Alternative Models of Behavior Change." Working Paper No. 131, Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. October 1990.

11. Hornik R and Foote D. "Evaluation Strategy for Healthcom." Healthcom Briefing Paper, Healthcom Project, Academy for Educational Development, Washington DC. January, 1986.

12. Hornik RC, McDivitt JA, Yoder PS, Zimicki S. "Explaining Health Communication Success: Evidence from the Healthcom Program (Draft)." Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. (No date.)

13. Koepke CP, Zimicki S and Hornick R. "Measuring Use of Oral Rehydration: Conceptual Issues and Evidence from HealthCom Surveys." August, 1988.

14. McDivitt JA and Abulaban A. "The Healthcom Project in Jordan: Final Case Study Evaluation Report." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 1991. (Jordan program evaluation report)

15. McDivitt JA, McDowell J, Satoto, Palmer A. "Results from the Evaluation of the Healthcom Project in Central Java 1988-1989." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 1991. (Central Java program evaluation report)

16. McDivitt JA, McDowell J and Zhou F. "Evaluation of the Healthcom Project in West Java: Results from Surveys of Mothers and Volunteer Health Workers." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. February 1991. (West Java program evaluation report)

17. McDivitt JA, Zimicki S, Hornik R, Abulaban A. "The Mass Media in Breastfeeding Promotion: The Impact of the Healthcom Project on Early Initiation of Breastfeeding in Jordan (Draft)." Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. (No date.)

18. Rasmuson M, Fluty H and Clay R. Healthcom: Lessons from 14 years of experience in health communication. (Accepted for publication in *Development Communication Report*.)

19. Rasmuson MR, Seidel RE, Smith WA, Booth EM. "Communication for Child Survival." Healthcom Project, Academy for Educational Development, Washington DC. June 1988.

20. Seidel R. *Results and Realities: A Decade of Experience in Communication for Child Survival*. Academy for Educational Development, Washington DC. January 1992.

21. Yoder PS and Hornik RC. "Perceptions of Severity of Diarrhea and Treatment Choice: A Comparative Study of Healthcom Sites (Draft)." Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. May, 1992.

BEST AVAILABLE DOCUMENT

22. Yoder PS, McDivitt JA and Hornik R. Knowledge of oral rehydration and response to diarrhea: A comparison among Healthcom sites. *Int'l Quarterly of Community Health Education*, 13(3): 201-217, 1992-93.

23. Yoder PS and Zheng Z. "Healthcom in Lesotho Final Evaluation Report." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 1991. (Lesotho program evaluation report)

24. Yoder PS, Zheng Z, Zhou F, Rugagi N, Ngwarsungu C. "Results of the Healthcom Evaluation in Zaire, Lubumbashi, 1988-1990." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. January 1991. (Zaire program evaluation report)

25. Yoder PS and Oke A. "Healthcom in Nigeria Final Case Study Evaluation Report." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. August 1991.

26. Zimicki S. "The Healthcom Project in Papua New Guinea: Final Case Study Evaluation Report." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 1991.

27. Zimicki S, Hornik RC, Lee MB, Pabalan M, Bagasao MT, Almario E, Pilipino KP, Fausto A. "The Healthcom Project in The Philippines: Final Case Study Evaluation Report (Draft)." Healthcom Project, Academy for Educational Development, Washington DC and Center for International Health and Development Communication, Annenberg School for Communication, University of Pennsylvania, Philadelphia. March 1991. (The Metro Manila program evaluation as well as a draft of the Philippines program evaluation)