

CHANNELING HEALTH:
A REVIEW OF THE EVALUATION OF TELEVISED HEALTH CAMPAIGNS

Richard Lau
Robert Kane
Sandra Berry
John Ware
Douglas Roy

December 1978

P-6262

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

1

Chronic diseases and accidents now claim more lives in the United States than infectious diseases. Because deleterious lifestyle practices such as smoking, use of alcohol and drugs, poor diet and lack of exercise are linked to the prevalence of chronic diseases and accidents, changes in these practices offer the possibility of preventing or reducing the severity of these diseases. The mass media promise a means by which large numbers of the population can be reached, informed and, it is hoped, motivated to undertake these lifestyle changes. The potential of the media to alter opinion is reflected in the large sums of money invested in advertising and the more recent efforts to restrict such advertising for products deemed dangerous to health; nevertheless, although there is a substantial body of scientific research on the effects of the mass media in general (e.g., Comstock and Fisher, 1975), relatively little is known about the media's ability to produce even minimal changes in health behaviors.

Because of television's great potential for education, it will be the primary focus of the review of health-related media campaigns. Television is ubiquitous throughout the United States; it is on for an average of six hours a day in a typical American home. On a given day, 65 percent of the population will watch television; most people spend 15 hours a week exclusively watching television (Marshall, 1977). Children spend more hours in this pastime than do adults. It is estimated that upon graduation from high school, a child in America will have spend 12,000 hours in formal education and 22,000 hours in front of a television set (Looney, 1971).

Even in the absence of scientific documentation, there is anecdotal data to suggest that health messages delivered as part of commercial television programming can produce impressive audience response. The addition of health programming to news coverage, for example, has resulted in increased audience size. When further information on health problems was offered as part of a nationally televised morning news program, more than half a million inquiries were received. The

unresolved question, though, is the extent to which this capacity for mass communication can be constructively harnessed to produce meaningful and deliberate changes in lifestyle practices. This paper will explore the evidence to date for the effectiveness of such campaigns and will examine the methodologic considerations that apply to future evaluations of health-related media efforts.

Health Behavior

That the media are likely to have only a limited role in changing the lifestyle practices of individuals can best be explained in the context of a more general model of why people undertake preventive health behaviors. This health belief model is shown in Figure 1. Basically this model, and others similar to it, suggest that individuals maintain certain perceptions about their susceptibility to a particular disease and about the potential seriousness of that disease. Their decision to take some action to prevent the disease depends upon the degree to which they perceive that disease as directly threatening them. Their behavior is a result of simply weighing the benefits of the preventive action against the barriers to that action. A cue or stimulus provided in the appropriate context may tip the scales in favor of taking the desired action. Not only can the media provide such a stimulus, but it can potentially act at a variety of other levels, including increasing the people's knowledge about the seriousness of a disease and susceptibility to it, offering new information about the additional benefits of preventive action or ways of removing barriers to such action, or providing models of how to perform the preventive health behavior.

Analysis of the relationship between beliefs and behavior examined in current literature is impeded by the lack of consistency in terminology. Several different definitions of attitudes, for example, are popular within the social sciences. One uses "attitude" as a generic term involving affective, cognitive, and conative dimensions (e.g., Brown, 1965; Kretch and Crutchfield, 1948; Newcomb et al., 1965; Secord and Backman, 1964; Sherif and Cantril, 1945, 1946). Another reserves the

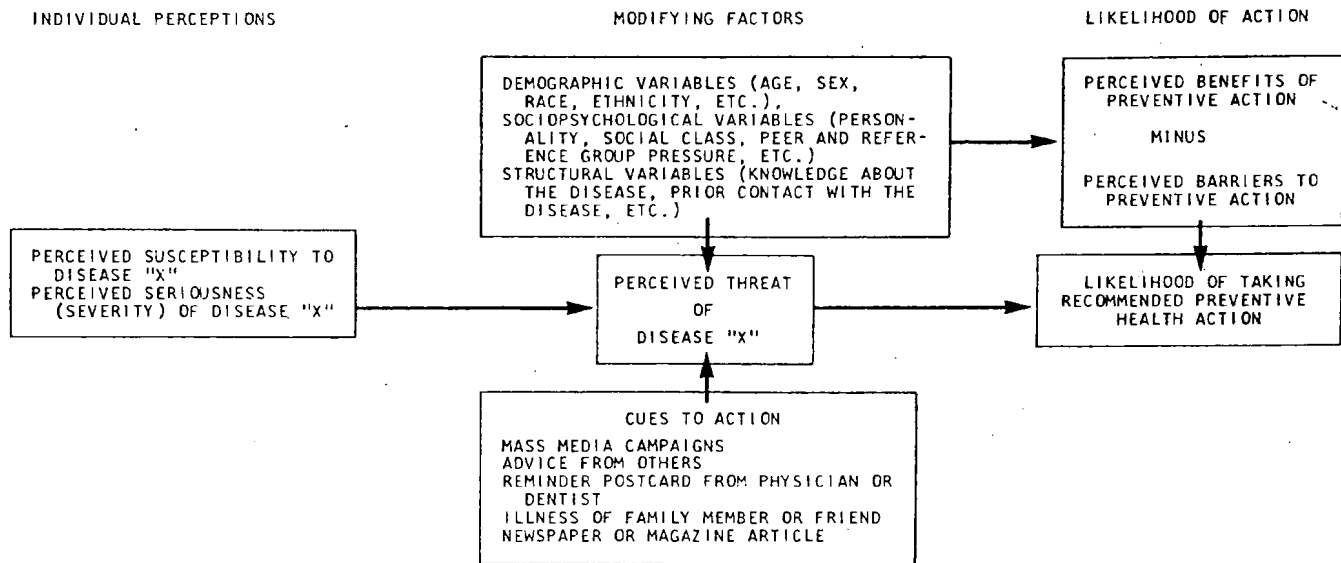


Figure 1: The "Health Belief Model" As Predictor Of Preventive Health Behavior

Source: Becker, M.H., Drachman, R.H., and J.P. Kirscht. A New Approach to Explaining Sick Role Behavior in Low Income Populations. *American Journal of Public Health*, 64:205-16, March 1974.

term strictly for the evaluative dimension, for feelings of good or bad, like or dislike about an object (e.g., Fishbein, 1967; Insko and Schopler, 1967; Thurstone, 1931). The term "beliefs" is then used for nonevaluative knowledge about an object. Although in the health field "beliefs" is more commonly used than "attitudes," it connotes some level of personal evaluation. To clarify the use of terms in this paper, "knowledge" means facts that are widely accepted (for instance, that obesity is associated with heart disease). "Beliefs" may serve the same function for the individual as knowledge, but they are much more personal cognitions in areas where the facts are less clear and therefore fairly difficult to dispute effectively. The individual often has some personal stake in the beliefs, some motivation for holding them. (Thus, "I will not get lung cancer," and "IUDs are safer than birth control pills," are beliefs.) The term "attitude" is reserved for the clearly affective dimension (e.g., "I like my doctor").

Although it was popular not too long ago in social psychology and sociology to question the relationship between attitudes and behavior (Deutscher, 1966; Fishbein, 1967; Wicker, 1969), more cogent, recent reviews have concluded that the attitude-behavior link is substantial (Ajzen and Fishbein, 1977; Shuman and Johnson, 1976). A strong association between beliefs and behaviors usually depends on their being measured at similar levels of specificity, particularly if the beliefs are based on experience (Regan and Fazio, 1977).

But the relationship of beliefs and behaviors is not perfect. As indicated in the health belief model, behavior is determined by many other things as well. In order to change behavior, the person must first be motivated to change, and then must know how to change. Beliefs and knowledge can certainly affect motivation and can lead to knowing how to change, but while they may be necessary for behavior change, they are not always sufficient. The motivation to change must be stronger than the motivation to maintain the old behavior. As anyone who has ever tried to quit smoking can attest, this is not always so. Nevertheless, belief change and knowledge are valuable

outcomes in any behavior change program; if any extrinsic event alters the balance of motivation away from the old behavior, the new beliefs and knowledge can guide the change.

Media Effectiveness

The vast potential of the mass media, particularly television, for communication and persuasion has long been recognized. This potential, however, is limited by several factors regardless of the type of message being sent. Any campaign, no matter how widespread, constitutes only a tiny proportion of what is available on the airwaves. Unless the dose and duration of the campaign is high, it may not reach even a large minority of the population. Moreover, given that most people watch television for entertainment rather than education, any explicitly educational or informational show will attract a minority of viewers. Furthermore, many people physically or mentally tune out during commercials and/or public service messages. And even when a person receives a given message, persuasion is by no means assured. Any inducement to spend one's time or money in a particular way competes with other conflicting messages. In addition, people selectively evaluate incoming messages, particularly when the messages are of direct concern to them. Psychological motivations to avoid dissonance or maintain cognitive consistency may lead to a very distinctive willingness or unwillingness to accept certain messages. "People are particularly adept at avoiding acceptance of information that is discrepant from strongly-held prior attitudes" (Sears and Whitney, 1974).

Given all the mitigating factors, any media campaign must be careful to set limited, realistic goals, employing the best social psychological variables for attitude change. Gross summarizes these persuasion techniques:

An effective, persuasive communication is one which is thought to emanate from an expert and trustworthy source which is able to capture and hold the receiver's attention while it conveys an easily comprehensible message and which offers the receiver clear and realistic channels of action in which to express his agreement. No one of these aspects can be overlooked except at the peril of the entire enterprise (1971, pp. 619-20).

Health information is available in the mass media in such forms as regular columns in magazines and daily newspapers and magazines specifically devoted to health, but only the very highly motivated read such information. There is very little health information on radio and television where it might reach less motivated people. Smith et al. (1972) found that only 7.2 percent of the content of a week of television was health-related; of that, 30 percent was judged useful and 70 percent inaccurate or misleading. Other than the popular shows portraying fictional characters that sometimes provide misleading health information, TV offers little regular, systematic health information. With an occasional exception on commercial TV, the majority of the documentaries or special programs appear on educational channels. These channels, however, do not fare well against the major networks in attracting viewers. On the other hand, shows like "Consultation" or "Medix" which occasionally appear on the commercial channels typically air at unpopular times.

Wade and Schramm (1969) report the results of a national survey fielded in 1958 contrasting printed media (newspapers and magazines) and television as sources of health information. As a whole, respondents reported getting more information about health from the printed media than from television; only those with a grade-school education or less reversed this trend. Wright (1975) reported much the same pattern of results in a very small sample of respondents about 15 years later, although Massarik et al. (1977) found that television was the most frequent source of knowledge about Tay Sachs disease. These results reflect the fact that one can frequently turn to written matter about a specific problem while this is almost never the case with television.

Moreover, most short health messages are public service announcements aired at less desirable times when a station has not otherwise sold commercial time. And these health messages must compete against the great barrage of contrary commercials; any message about nutrition or obesity must be very powerful to offset the scores of commercials about snack and junk foods seen every night. Even at their highest

frequency as competitive messages, anti-smoking spots were outnumbered four to one by commercials for cigarettes (Smith, 1972).

Nevertheless, several media messages about health which were not systematically evaluated, apparently were tremendously successful. The "CBS National Driver's Test" reached a large audience and presumably resulted in a three-fold increase in the number of people enrolling in driver education or driver improvement courses (Bush, 1965). One Ann Lander's column urging readers to write the American Cancer Society for a free brochure on breast self examination elicited more than 300,000 requests rather than the expected 40,000 maximum (Rimer, 1976). Media coverage of the Betty Ford and Happy Rockefeller mastectomies was also credited with motivating a large number of women to go to their doctors about lumps in their breast. These examples of media effectiveness are of limited usefulness, however, because there was no attempt to determine the extent to which the apparent change in health behaviors could be attributed to the media event or to the communication which preceded it.

Systematic study of the effects of mass media in other fields yields mixed results but generally the effects may not be as potent as one might expect. For instance, media campaigns have generally had only minimal influence on political persuasion (Sears and Whitney, 1974); viewing the televised presidential debates tended to reinforce prior predispositions rather than to change opinions (Katz and Feldman, 1970; Sears, 1977). Likewise, information campaigns conducted by the government have met with equivocal success. Savings bond drives during the Second World War did succeed in convincing about 40 percent of the population to buy bonds, but very few people learned the reasons why the government wanted people to buy bonds (Cartwright, 1949). A campaign in Cincinnati to create support for the U.N. increased support among those already favorable toward it (Star and Hughes, 1950). Finally, the huge advertising budgets supporting the sales of many commercial products notwithstanding, most firms are delighted when an advertising campaign results in a one or two percent increase in their share of the market; no one in marketing expects an advertising campaign to result in huge shifts in buying patterns (Bauer, 1964).

HEALTH CAMPAIGNS IN THE MASS MEDIA

The literature review on health campaigns in the mass media is intended to give a representative picture of both the types and effectiveness of the campaign studied and the designs used in the evaluation. It does not presume to be comprehensive, although no field study has been purposely omitted. It is restricted to field studies of media health campaigns; laboratory studies (usually investigating attitude change) that present a message through a mass media channel have been excluded.

A major problem in evaluating the effectiveness of any mass media campaign is finding comparable groups of viewers and nonviewers. Within any given population, viewers and nonviewers will differ on a variety of characteristics--including motivation to seek out health information--which may affect the dependent variable of interest. That is, any difference between viewers and nonviewers on some criterion may be a function of the message, or may be a result of some extraneous variable associated with viewing. Hence, the very limited value of simple comparisons between viewers and nonviewers.

The first group of studies summarized in Table 1 are of this limited nature. Viewers are contrasted with nonviewers in terms of knowledge about or preventive steps taken against certain diseases. Moreover, only viewers are queried about their reactions to some health program. Although better than having no evaluation, these studies are inadequate for making definitive statements about the effectiveness of health-related messages in the mass media. Unfortunately, by far the largest number of studies fall into this category.

Haskins' (1969) review of safety communication campaigns conducted prior to the 1970's that found almost no good evaluations concluded "The principal flaw in campaign evaluation studies has been the reliance on before-and-after research designs with no control group" (p. 59). Likewise, in Young's (1973) review of studies of family planning media campaigns, most of which were conducted in third-world countries, almost all involved interviews after the campaign had begun, frequently with self-selected samples. Concerned as much with the

Table 1: Inadequate Studies of Media Health Campaigns

Study	Topic	Media	Population	Design	Results
Haskins (1969)	Safety	Various	Various English & American samples	Mostly single group pretest-posttest & after-only correlational	Most frequently no conclusions could be drawn
Young (1973)	Family Planning	Mostly radio Public Service Announcements (PSAs)	Mostly 3rd-world adults	After-only correlational	Many heard media message, but word-of-mouth was often more effective
Aldis et al. (1969)	Prenatal Education	4 TV programs	174 Canadian Drs., nurses, social workers expectant mothers most of whom were invited to watch	Single group after-only correlational	Some aspects of programs preferred
Edwards et al. (1973)	Mental Health	16 program series on Public Broadcasting Service (PBS) channels	76 Educators enrolled in tele-course	Single group after-only correlational	Some aspects of program preferred
Yarnell (1976)	Measles Vaccinations	Various	81 mothers of young children in Bristol known to be susceptible	After-only correlational	Knowledge of vaccination campaign thru media did not distinguish those who had children vaccinated from those who did not
O'Keefe (1971)	Anti-smoking	TV and radio PSAs	621 Students & a random sample of 300 adults in Florida	After-only correlational	Perceived effectiveness of commercials greater among non-smokers and smokers predisposed toward quitting; very little perceived influence on behavior
Dickman & Keil (1971)	Alcoholism	4 programs on PBS	Random survey of 1200 people in Pennsylvania	After-only correlational	Very low exposure; some help in recognizing alcohol problem, but little behavior change
Mendelsohn (1973)	Anomie	65 episode soap opera on Span. language station	Sample of Mexican-American population in L.A.	After-only correlational	15% of target population saw some of the series, 6% joined community organization
Baran (1977)	Mental Retardation	4 programs on local commercial station	Random sample of 40 adults in Cleveland surveyed before programs; 40 volunteers who had seen program	Separate sample, pretest-posttest	Volunteers who had seen program more favorable toward retarded
Salzer et al. (1977)	Disease Screening	Short announcements on localized cable TV	330 elderly tenants of high-rise in New York City	After-only correlational	Those who saw announcements were more knowledgeable & more likely to participate in screening
Schmeling & Wotring (1976)	Drug Abuse	TV PSAs	Random sample of 48 middle-aged women in Tallahassee	Pretest-posttest correlational	No differences in self-esteem or perceived seriousness of problem
			Random sample of 614 Florida women	After-only correlational	
Greenberg & Gantz (1976)	Venereal Disease	PBS Program "VD Blues"	Random sample of 923 people in Lansing, Mich.	After-only correlational	Those who had seen show more knowledgeable about VD

dissemination of the information as with acceptance of the message, they usually reported that a small proportion (.3 to 25 percent) of women who heard the message went to family planning centers for birth control devices.

Aldis et al. (1969) examined the reactions of a group of invited viewers to a series of prenatal teaching communications shown on a local commercial station. The evaluation, however, consisted solely of what was liked and disliked about the programs. Similarly, Edwards et al. (1973) sent questionnaires to a group of educators enrolled in a "telecourse" offering mental health information to school teachers. Again, the questions dealt merely with the effectiveness of different portions of the show.

Yarnell (1976) evaluated the effectiveness of a measles vaccination campaign in England. All mothers in Bristol with young children known to be susceptible to measles were sent appointment invitations to bring their children in for shots. A random sample of mothers of young children who had recently visited two medical clinics for reasons other than getting vaccinations (all of whom were previously sent appointment cards) were interviewed sometime later. Those who had brought their children in for shots were contrasted to mothers whose children were susceptible to measles but who had not brought them in. Hearing about the vaccination campaign in the media did not differentiate attenders from nonattenders.

O'Keefe (1971) examined the effects of anti-smoking commercials on people in Florida. During the time of the study between 80 to 100 televised anti-smoking commercials appeared each week. Almost 90 percent of the respondents reported seeing at least one such commercial and 50 percent could remember the message of at least one commercial. Those who were predisposed toward quitting were much more likely to report that the commercial had some effect on them (50 percent) than were those who did not want to quit (17 percent). Only one percent of the nonsmokers and exsmokers, however, credited the commercials with preventing them from smoking or helping them quit.

Dickman and Keil (1977) assessed the effects of four weekly ninety-minute programs about alcoholism on public television in Pennsylvania. They conducted a large survey after the series, but only 2.3 percent of their sample had seen any part of any of the shows; another 3.8 percent had some awareness of the campaign. Of respondents who reported experiencing alcohol problems, 18 percent were familiar with the campaign compared with 4.4 percent of those who had not experienced alcohol problems. Almost 41 percent of those experiencing an alcohol problem and who were familiar with the campaign credited it with helping them realize the presence of a problem, but barely 23 percent said the campaign had stimulated them to take any corrective action.

Mendelsohn (1973) evaluated the production of "Cancion de la Raza," a Mexican soap opera shown on the Spanish-language channel in Los Angeles designed to reduce anomie among the local Mexican-American population. A survey taken after the series was completed indicated that about 15 percent of the target population had seen at least some of the programs. Most viewers said the series was helpful to them in one way or another and six percent reported they had actually joined a community organization (the primary goal of the project) because of the series.

Baran (1977) reported the effects of four thirty-minute dramas about mental retardation shown on commercial television in Cleveland. The last program solicited viewers to take part in a study, and 40 people complied. These 40 volunteers were contrasted with a randomly-selected group of 40 people who had been interviewed before the airing of the four programs. The volunteers were more educated than the control group, but did not differ in the amount or nature of contact with retarded people. They were, however, significantly more favorable in their attitudes toward mental retardation than were the control group.

Salzar et al. (1977) studied the effects of six different health commercials shown on cable television in a high-rise apartment complex for the elderly poor in New York City. The commercials discussed a specific disease and gave information about where in the building to

obtain free medical screening for the disease. All who came down for a screening and an equal number of randomly-selected nonparticipants in the building were interviewed during the week following the screening. Of those interviewed, people who had seen the commercials participated in the screening significantly more often, and also gave significantly more correct answers about the particular disease than did those who had not seen the commercials; however, given the inherent bias of the sample and the ease of participation in the screening, the commercials were disappointingly unsuccessful in attracting residents to the screenings. From 2.6 to 14.4 percent of the residents (median, 7.7 percent) participated in any given one of the six screenings available over the course of the project.

Schmeling and Wotring (1976) evaluated the effectiveness of a series of programs designed to reduce prescription drug abuse among middle-aged women. Some women were interviewed both before and after the commercials appeared on television while others were interviewed only after the campaign. The commercials did not change the attitudes of women from the pre- to postinterviews, nor did viewers differ from nonviewers in their attitudes about this issue.

Greenburg and Gantz (1976) interviewed a group of respondents after the airing of the television show "VD Blues." Viewers knew more and perceived themselves as being more knowledgeable about venereal disease than did nonviewers. Greenberg replicated these results in a laboratory using a group of students who had not previously seen the show on TV assigning them randomly to view or not view the program.

These studies present a mixed picture of the effectiveness of media health campaigns. They offer only limited guidance, however, because they are all correlational; there is no experimental manipulation of the independent variables. Viewers are contrasted with nonviewers, and there is no guarantee that the two groups are in any way comparable. Some variable which is confounded with viewing (such as education which is associated with viewing public television) might be associated with the behaviors the campaign was trying to influence (such as knowledge about drug abuse). Using an individual as his own

control in a before-and-after design does not solve the problem. Changes in the viewer group could be due to the effects of other stimuli that distinguish viewers and nonviewers.

The following studies which are summarized in Table 2 avoid this self-selection bias and thus offer much more useful data for evaluating the effectiveness of media campaigns on health issues. Basically three different evaluation designs have been used, and while all three are better than the correlational studies discussed above, each design has its own set of weaknesses.

The first two studies utilized a time series design which allows more inferential conclusions. Bailey (1973) examined the effects of a radio campaign about family planning in Columbia. Radio messages about family planning were aired for 17 1/2 months interspersed throughout a three-year period. By examining the trends in the number of new women acceptances each month for three years at the sixteen family planning clinics, Bailey concluded that the initiation of the radio campaign was associated with an increase in this number. It might be suggested that these increases were simply borrowing from the future, that they were only attracting people earlier who would have come to the clinic at a later date. Bailey argued, however, that this was probably not the case; people who went to family planning clinics during this period because they heard about the clinics on the radio differed on a variety of demographic variables from the typical family planning acceptor.

Warner (1977) evaluated the effects of the major events in the anti-smoking campaign in the United States by examining annual per capita cigarette consumption (controlling for the relative price of a cigarette) from 1947 to 1975. The events included the first smoking health scares in 1953 and 1954, the Surgeon General's report in 1964, and the many anti-smoking commercials appearing on TV and radio between 1968 and 1970. Each of these events was associated with a significant decrease of three to eight percent in annual per capita cigarette consumption. Had none of this anti-smoking publicity occurred (assuming that other factors like economic conditions remained stable), Warner

Table 2: Adequate Evaluations of Media Health Campaigns

Study	Topic	Media	Population	Design	Results
Bailey (1973)	Family Planning	Radio PSAs	16 family planning clinics in Columbia	Time series	The campaign increased the number of FP acceptors at the 16 clinics
Warner (1977)	Anti-smoking	Various	U.S. yearly per capita cigarette consumption, 1947-75	Time series	Each major event in anti-smoking campaign assoc. with significant per capita cigarette consumption decrease
Robertson et al. (1974)	Seat belt use	Cable TV PSAs	Drivers in city of 230,000	Pretest-posttest	No effect
Morrison et al. (1976)	Alcohol & Drug Abuse	Radio & TV PSAs	Residents of 2 small cities	Nonequivalent control group (flawed)	No differences in PSAs heard; no difference in knowledge & attitudes in groups
Kline et al. (1974)	Family Planning	Radio PSAs	Residents in 2 cities	Nonequivalent control group	Very low exposure; No difference in knowledge between groups
Fleisher (1973)	Seat belt use	Radio & TV PSAs	Drivers in 3 central Calif. cities	Nonequivalent control group	Seat belt use temp. increased in 1 experimental & control group during campaign
Udry et al. (1972)	Contraceptive Use	Various	Residents of 10 southern cities	Nonequivalent control group	Temp. increase in acceptances at FP clinics in experimental cities; no difference in contraceptive sales
Douglas et al. (1970)	Mental retardation	Various	145 residents of 2 small Wisc. cities	Nonequivalent control group	Significant information gain & attitude change in experimental city
Farquhar et al. (1977) Maccoby et al. (1977)	Heart Disease	Various	1004 residents of 3 northern Calif. cities	Nonequivalent control group	Media campaign alone & media campaign + personal contact associated w/reduction in various risk factors associated w/heart disease
Minor and Bradburn (1976)	"Feeling Good"	18 PBS shows	468 residents of lower income area of Dallas	Pretest-posttest; experimental group induced to watch	Increases in knowledge about some diseases, some health behaviors & overall interest in health
Cohen & Abelson (1976)	"Feeling Good"	18 shows on PBS	7028 residents of 4 large cities in the U.S.	Moving control group (see text)	Increases in some health behaviors
Ferstl et al. (1977)	Obesity	TV	Residents of German Fed. Rep. & 4 special study sub-groups	Pretest-posttest comparison of TV & bibliotherapy	TV less effective in promoting weight loss than bibliotherapy

estimated that annual cigarette consumption would have been almost 20 percent higher in 1975 than it actually was. Warner suggested further that this study greatly underestimated the effectiveness of the campaign because it ignored the tar and nicotine levels of the cigarettes currently sold.

Both studies found significant effects of media health campaigns. Both benefited from relatively large dosage and relatively long duration campaigns. The major problem with these two studies is the inability to exclude the possibility that the changes might have taken place in the absence of any information campaign. The reduction in cigarette consumption, for example, might be a result of factors other than those measured by Warner. The problem obtains with Bailey's study because the family planning radio commercials were almost always in the latter half of the year and it is possible that the number of new acceptors of family planning was always higher during those months. Since Bailey does not provide data for years other than the three in which the radio commercials were aired, this possibility cannot be ruled out.

The self-selection biases inherent in correlational studies, and the problems of history or naturally occurring changes which cannot be ruled out with time series designs are eliminated if people can be randomly assigned to receive or not receive the communication. This is easy to accomplish in a laboratory, but much more difficult in the real world.¹ Nevertheless, one study was able to approximate fairly closely the laboratory researcher's ability to randomly assign subjects to media and nonmedia groups. Robertson et al. (1974) took advantage of a dual cable television system designed for marketing studies to show commercials advocating safety belt use to homes on one of the cables. Homes on the other cable and homes without cable television served as controls. The proper control group, however, is drivers in homes with the cable TV line that did not receive the safety belt commercials. Drivers from homes without cable television may differ from those in homes with cable TV on a variety of factors. Systematic observation of seat belt use was conducted for one month prior to the

campaign and throughout the duration. License plate numbers identified in which type of home the driver lived. There were no observable differences in seat belt use among drivers in homes with either cable or in homes with no cable television. Unfortunately, there was no discussion of the size of effect from the campaign necessary in order to be detected with this design.

This degree of control over what can be seen on television is very rare. Two reasonable methods do exist to approximate random assignment in a laboratory however. One is to select fairly comparable communities and randomly assign reception of a media campaign to some of the communities and not to others. This type of study is possible when local media are utilized for the campaign. Such a nonequivalent control group design was attempted by Morrison et al. (1976) who studied two small cities in the U.S., one of which was subjected to special alcohol and drug abuse commercials. Unfortunately for the study, there were other groups besides the investigators publicizing the problems of drug abuse, so that there were no differences between the experimental and control cities in the proportion of people interviewed who had heard an alcohol or drug abuse commercial. Hence the study degenerated into a correlational design and found no differences in knowledge or attitudes about drug abuse between those who had heard a commercial and those who had not.

Kline et al. (1974) studied the effects of a radio campaign about family planning aimed at adolescents. Survey respondents in the experimental city where the commercials were played did hear significantly more family planning messages than respondents in a similar control city which was not given the commercials. Subjects in the experimental city, however, heard less than one such message on the average, and not surprisingly, the study found no differences in knowledge about family planning between the experimental and control cities.

Fleisher (1973) reports the results of a six-week campaign in central California to increase safety belt use. One city was subjected to an intense media campaign, another to a moderate campaign, and a third city served as a control and received no seat belt messages.

Both telephone interviews to determine attitudes and direct observation of seat belt use were conducted before, during, and after the campaign in all three cities. Seat belt use did increase significantly during the campaign in the high-intensity city, but then returned to pre-campaign levels after its conclusion. No significant changes in seat belt use were found in the moderate intensity city. Inexplicably, seat belt use in the control city that received no messages followed the same pattern as that in the high-intensity city with an increase in safety belt use during the campaign and a subsequent decrease in use after the campaign. Hence, it would be inappropriate to attribute any changes in seat belt use to the media campaign. Moreover, there were no major changes in attitudes toward safety belt use among the three cities which could be attributed to the campaign.

Udry et al. (1972) evaluated a six-month multimedia campaign in four southern cities aimed at increasing contraceptive use. Six other southern cities which had not received the campaign served as the controls. New acceptances at family planning centers in two of three experimental cities increased significantly more than control cities during the campaign. Data from a fourth experimental city was not available because the family planning center did not exist in the city prior to the campaign. This one clinic did experience "rapid" early growth, to which the media campaign might have contributed. The growth trends, however, dipped below pre-campaign levels during the last month of the campaign in the experimental cities and the author suggests the campaign may have simply borrowed from the future to achieve the early spurts. No differences were found between experimental and control cities in the sale of birth control pills or condoms.

Douglas et al. (1970) studied a six-month multimedia campaign concerning mental retardation in a small community in Wisconsin. The campaign included a number of stories in the local newspaper, posters, displays of goods made by the retarded, countless radio broadcasts, items in church bulletins, and speakers at a variety of community organizations. Information about and attitudes toward mental retardation were measured both before and after the campaign in this

experimental city and a matched control city. There was a significant gain in information and change in attitudes in the experimental city; no changes were found in the control city. Moreover, information gain and attitude change were significantly correlated (.40) in the experimental city but not in the control city (-.05).

The Stanford three-community study was an effort to reduce the risk of cardiovascular disease in several small northern California cities (Farquhar et al., 1977; Maccoby et al., 1977; Nash and Farquhar, 1978; Stern et al., 1976). A media educational campaign designed to teach specific behavioral skills was carried out in two of the cities; the third city received no educational campaign and served as a control. A random sample of men and women between the ages of 39 and 59 was interviewed in each city before the campaign started to measure knowledge about heart disease. Daily intake of cholesterol, fats, sugar, alcohol and tobacco was recorded, and various physiological measures of cardiovascular health were also taken. In addition, two-thirds of the study participants in one of the two media cities received intensive face-to-face instruction and other educational stimuli. The same basic interview was repeated twice at one-year intervals after the campaign began to determine its effectiveness. All behaviors except weight loss were significantly affected in a positive direction by the media, with and without face-to-face instruction.

As the Stanford three-community study illustrates, the nonequivalent control group design can, if it is well-executed, be very powerful. The design does have several drawbacks, however. One is that usually only a handful of sites are studied and these will never be exactly comparable; that is, randomly assigning one of two sites to receive a treatment only approximates random assignment at the individual level. Another problem is the possibility that something independent of the research design (e.g., a new law) which affects the behavior under study will occur in one site and not the other so that the groups are no longer comparable. A third problem is the added cost of setting up research groups in several locations. Furthermore, only the local media may be used in this design; national media presumably reach all potential experimental and control cities equally.

The second method of approximating a laboratory study with a random assignment is to randomly encourage some subjects to watch the communication of interest. This results in two randomly-selected groups with a higher proportion of viewers in the experimental, or encouraged, group than in the control, or nonencouraged group. This encouragement is not designed to simply induce more people to watch the show; such a design would cause problems of external validity or generalizability. Rather, prompting is intended to increase the proportion of viewers in the sample being studied from a population in which viewing is infrequent. This design eliminates the problems of the nonequivalent control group design discussed above. Random assignment is done at the individual level, extraneous events should affect both experimental and control groups equally, and only a single site (and therefore a single experimental research group) is required.

This design has its own special problems, specifically, the effectiveness of the inducements to watch and the possible confounding effects of the inducements *per se* above and beyond the health communication itself. To assess the effects of such prompting, a properly designed study would allow for four subgroups as seen in Figure 2. Examining main effects and interactions will test the possibilities of the design limitations noted.

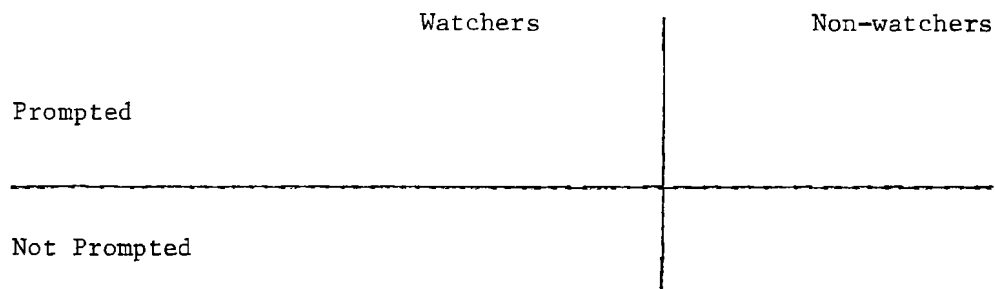


Figure 2: Testing for Inducement Effects

Only two studies were found that utilize this design and only one deals specifically with health. (The other was an evaluation of "Sesame Street" which will be discussed below.) One of the evaluations of the television program "Feeling Good" conducted by the National Opinion Research Center (NORC) involved a survey of women in a low-income neighborhood in which a random half of the respondents were paid \$50 to take part and watch the show. Because "Feeling Good" was the most ambitious attempt yet to convey health information through the mass media, and because a strong evaluation component was included in the initial planning, it deserves a somewhat more lengthy discussion.

The show was designed by the Children's Television Workshop (CTW) to be an entertaining, prime time series on Public Broadcasting Service dealing with a variety of health topics. The show began with weekly, hour-long programs that interposed entertainment with information about and motivation to perform a variety of health-maintenance and disease-prevention behaviors. It was hoped the show would attract a larger audience than usual for a PBS program, and particularly a larger audience of lower income viewers who were seen as most in need of the information. Originally intended to last 26 weeks, the show was stopped by CTW after only 11 weeks (Season A) because of disappointing ratings. It returned after an 8-week hiatus (Season B) as a new half-hour show hosted by Dick Cavett dealing exclusively with a single health topic.

Four evaluations of "Feeling Good" were subcontracted by CTW. One was undertaken by Nielsen to estimate the national audience for the series (Nielsen Television Index Report on "Feeling Good," 1976) and another was a series of surveys by the Gallup Organization to determine trends in the awareness of the series and in health behaviors (The National Audience for "Feeling Good," 1975). As the latter study properly did not try to equate viewing "Feeling Good" with performing health behaviors (which would have been a purely correlational design), neither of these two evaluations are of interest here.² The NORC evaluation began with the assumption that very few low-income viewers would normally see "Feeling Good." By providing a substantial

financial inducement to watch "Feeling Good," it was hoped that 80 percent of the induced group would regularly watch the show. Because respondents were randomly assigned to the experimental (induced) and control groups, a true experiment results. The problem with this basic design is that the independent variable of interest (viewing "Feeling Good") is confounded with the inducement. To determine the effect of the inducement *per se*, a portion of the control group was also paid \$20. No mention was made of "Feeling Good" for this group; the money was simply paid for participation in the study. Because the series was intended to last a long time (26 weeks), a three-wave panel design was used in which all respondents were interviewed before the series began, at the conclusion of the study, and at an intermediate point which varied for different subgroups. Although some problems resulted from an unexpectedly high number of respondents in the control group watching at least one of the "Feeling Good" shows and a disappointingly low number of subjects in the experimental group watching, the following conclusions were reached:

1. The series did prove effective in increasing some fairly easy behaviors like writing away for health information and urging others to get regular medical examinations.
2. It did increase knowledge in a few areas.
3. It did increase overall interest in health.

The inducement *per se* was not found to have any effect.

The fourth evaluation of "Feeling Good" falls somewhere between a correlational and a quasi-experimental study. The Response Analysis Corporation (RAC) conducted a three-wave panel study in four large cities across the country (Cohen and Abelson, 1976). An initial telephone screening allowed oversampling from groups more likely to watch the show. Most respondents received an initial baseline questionnaire, an interim questionnaire during the series and a postseries questionnaire. All of the interim questionnaires, spaced one month apart, contained items specific to each of the previous four programs as well as items concerning topics that were covered in practically all shows.

Two types of analyses were presented. Frequent viewers, infrequent viewers, and nonviewers were contrasted with regard to questions addressed by the whole series. Although these three groups did not differ significantly on a variety of demographic variables and degree of worry about health collected during the initial baseline survey, there was evidence for some self-selection in the results. These analyses, therefore, should be interpreted very cautiously.

The analyses of the questions specific to a given show were more sound. Viewers of a particular episode of "Feeling Good" were contrasted with viewers of other shows in the series, but not to the specific one under analysis. There is no reason to expect self-selection to be a problem here, because among viewers of the series seeing or not seeing one particular show was probably a random event. These analyses were performed on interim questionnaires taken during Season A, when each show covered a variety of health topics. Advertisements about a particular show could, therefore, not be specific to any health topic, thus eliminating the most likely source for viewer selection of a particular program. The results from the RAC evaluation are similar to those of the NORC evaluation.

The series as a whole did not have nearly as much effect as was originally hoped, however, for although it was in the top third of popularity among PBS shows, it did not really compete effectively with the commercial station. Despite the large publicity campaign associated with "Feeling Good," almost half of those who had never seen any of the shows said that they had never even heard about them (Cohen and Abelson, 1976). Thus, CTW concluded in its summary evaluation that a prime-time PBS series like "Feeling Good" was not a cost-efficient manner of providing health information (Mielke and Swinehart, 1976). They were careful, however, not to discourage other uses of the mass media for transmitting health information.

A study in the German Federal Republic explored the effectiveness of television in promoting weight loss. Seven "packages" were developed, each consisting of a 45-minute show in prime-time and a series of 3-minute reminders on subsequent evenings during that month. An experimental evaluation was utilized establishing four groups: television

only, mailed information only, both and neither. The results suggest that TV was not very effective. Although the weight loss in the TV-only group was significantly greater than the control group (5.5 lbs. compared to 2.2 lbs), the weight loss in the bibliotherapy groups, with and without television, averaged 16-18 lbs. and was significantly greater than that in the television group.

Some other findings of interest emerge from that study. A subset of each of the four groups was exposed to a more extensive investigation. In the television-only group, the intensive study efforts were reactive; subjects lost twice as much weight under the intensive study conditions as without it. This reactivity was not found in the other three groups. The investigators also carried out a validity on the self-reported weights. Although no specific data are reported, they conclude that the "measured weights were remarkably close to the self-reported weight" (Ferstl et al., 1977).

In summary, there are only a handful of studies evaluating health campaigns in the mass media that have not used a correlational design and even when the latter are considered, the evidence for any measurable or positive effect of health communication in the media is far from overwhelming. Most of these campaigns have suffered from very low exposure. The health message is either delivered as a public service announcement on commercial television during other than prime-time viewing hours or is a special program on the educational channels. Where the exposure was great enough, however, as in Warner's study of two decades of anti-smoking information or the Stanford three-community study that saturated two small towns with specific health information for almost two years, reasonably major effects of health campaigns have been found.

METHODOLOGIC CONCERNS IN EVALUATING MEDIA CAMPAIGNS

Any evaluation of a media campaign raises a number of methodological issues. Every design alternative involves its own specific problems. Although many of these issues and problems were briefly mentioned in the above review, they deserve more systematic discussion.

Breadth of Stimulus

Whereas short public service announcements are usually restricted by time to a single topic, longer programs are not so inherently limited. For the purposes of conducting a good evaluation, however, the ideal media campaign, regardless of length, should be concerned with one or a number of closely related health topics. Where multiple topics are involved in the evaluation, less time will be spent on them, resulting in decreased precision. A diffuse set of educational objectives impedes any precise evaluation. Moreover, given a large number of dependent variables, some will likely show significant change just by chance.

Sampling

Critical to any evaluation is the choice of whom to study; convenience and cost must be weighed against representativeness. The easiest method is to solicit volunteers for participation in the study, and if the solicitation is done during the message itself, one can be assured that those who do volunteer will have actually been exposed to the message under study. Another popular technique is to pass out questionnaires at shopping centers or other crowded places. The problem with this method is that those who volunteer will be a very select group. Not only will they have had to have seen or heard the message and been to the shopping center but they will have had to have been motivated enough to agree to volunteer. Rosenthal and Rosnow (1975) have written extensively about how volunteer subjects differ from nonvolunteers. This alternative should be considered only as a last resort.

Another fairly easy and relatively inexpensive alternative is to contact pre-existing groups that could be expected to be interested in the project. For instance, a weight watcher's group might be easily persuaded to view a television program about obesity. Again, however, this is a very self-selected group and one could not validly generalize the results of the study beyond this special group.

The best method of obtaining subjects representative of the population under study is to randomly sample the population of interest. This could mean sampling from all potential viewers in the city, state, or county in which the campaign has been conducted. If the campaign message were aimed at a particular subset of the population, however, the sample would include only members of that special population. To give an obvious example, if a campaign were aimed at increasing breast self-examination it would be reasonable to exclude males from the sample. This is not to say the communication cannot have beneficial effects on the non-target population; males who learned about the dangers of breast cancer might urge their female friends to use breast self-examination. But limited funds might argue for studying only the specific targeted group. Initial screening surveys to identify members of a particular subset of the population are relatively inexpensive compared with the rest of the research process. Of course, the smaller the subgroup of interest, the larger the initial screen will have to be to identify a desired number of members of that subgroup for study.

Moreover, it might be wise to over-sample that portion of the population most likely to be reached by the media campaign. This type of screening is done routinely when the local media are used in a health campaign and sampling involves only persons in the city or community within range of the particular television or radio station. This procedure can be taken a step further, however, by obtaining information about normal television viewing or radio listening habits of respondents as part of the initial screening survey. For example, if one were interested in evaluating the effectiveness of short messages about exercise aired during broadcasts of baseball games, one could find out who normally listens to baseball games on radio and television.

A compromise between a representative sample of the population as a whole and an exclusion of all non-targeted individuals is to over-sample the particular subgroup of interest. This usually means interviewing every member of the target group reached and a specified proportion of all other people contacted. This yields a sample which includes not only a large number of targeted individuals but also, with

statistical weighting to compensate for the over-sampling, is representative of the population as a whole.

Sample Size

The size of the sample to be studied depends on a variety of factors. The larger the number of individuals finally included in the study the greater the cost, but also the greater the statistical power one will have in drawing conclusions. "Power" here refers to one's ability to detect differences between experimental and control groups which are unlikely to have occurred by chance (i.e., those which are statistically significant). If large differences are expected to result from one's treatment, relatively little power is needed to detect an effect, and a relatively small sample will suffice. For the reasons discussed above, however, one should not expect any media campaign to have a large impact. Hence, any study of a media campaign usually requires a fairly large sample.³ Of course, the initial screening size grows proportionately to the extent that one is screening out certain individuals from the study.

Dose and Duration

For any media campaign to have a measurable effect, the dose and duration of the campaign must be relatively large. "Dose" refers to the frequency with which the message can be perceived--the number of times in any short period of time that it is on the air on radio or television, the number of different newspapers and magazines in which it appears. "Duration" refers to the period of time that the message can be perceived--that is, the length of the media campaign. The larger the dose, the larger the number of people who can be reached in a short period of time, and the greater the short-term effect of the campaign. The longer the duration of the campaign (assuming a constant dose throughout), the greater the period of time in which the message can be perceived and thereby reinforced and the greater the long-term effects. Unfortunately, there are no reliable rules about what constitutes a threshold--the minimum dose necessary for a large proportion

of the population to receive the message--nor about the need for sustained or repeated messages to prevent its extinction.

A campaign, for example, which is trying to transmit new knowledge (like Brand X causes cancer) can be relatively more effective than one seeking to change attitudes. Exposure to one or two messages from a credible source should be enough for most people to perceive the message and perhaps to accept it. Moreover, if readily available substitutes for Brand X exist, the campaign can also be successful in changing behavior. If, on the other hand, the campaign is trying to change beliefs or attitudes to which many people are committed or if the campaign is trying to change habitual behavior for which there is no good substitute, the campaign will require a very large dose and duration before any effect can be expected. Further, to the extent that beliefs and behavior counter to the campaign message are also being advocated through the media, the dose and duration will have to be greater yet to overcome this conflicting advertising.

Pretest Sensitization

A popular and powerful method for determining the effect of media campaigns is to measure the dependent variable of interest for each person twice, once before and once after the campaign. When pre-campaign or pretest scores are used to control for initial individual differences in a dependent variable, the power of finding group differences on the post-test is greatly enhanced. The danger with any pretest-post-test design is that the pretest might sensitize subjects to the treatment they are about to receive; that is, pretested subjects may react differently to the media campaign than non-pretested subjects. Learning might occur from simply taking the pretest.

Fortunately, there is a large body of literature on pretest sensitization to provide estimates of the size of this problem in a given situation. In the "Solomon four group design" (Solomon, 1949), used to study sensitization effects, two groups of subjects get a pretest and two groups do not. Next, one pretested and one non-pretested group hear a persuasive communication. All four groups are then compared.

on a post-test (see Figure 3). If the pretest and the communication interact--that is, if the pretested group that heard the communication show more, or less, attitude change than the non-pretested group that heard the communication--then pretest sensitization has occurred.

	Pretest	Communication	Post-test
Group 1	X	X	X
Group 2		X	X
Group 3	X		X
Group 4			X

Figure 3: The Solomon Four Group Design

Lana's (1969) review of the literature on pretest sensitization finds considerable evidence for pretest sensitization when the communication is bidirectional (two-sided) but almost no evidence for sensitization when the communication is unidirectional. Lana argues that pretesting might commit subjects to opinions which are discrepant from those advocated in the communication, but there should be no such resistance to change if the subject initially supported the opinions advocated. With a bidirectional message, all subjects oppose some of what is said, and the pretest commitment could cause resistance to persuasion. But with a unidirectional communication, assuming half of the subjects initially agree with the message, no freezing of attitudes should occur for this half. This reasoning is particularly reassuring for the purposes of media evaluation on health topics; not only will most health communications presented by unidirectional, but also few people will initially believe the position opposite of that advocated (e.g., believe that cancer is good for you).

One study subsequent to Lana's review, however, clouds the picture, noting a three-way interaction between the volunteer or nonvolunteer status of the subjects, pretesting, and a unidirectional communication. Rosnow and Suls (1970) solicited volunteers from a psychology class for a psycho-linguistics study. This solicitation merely established whether

or not subjects were willing to volunteer for experiments, because the psycho-linguistics study was never actually run. A short time later, all students in the class participated in an attitude change study. Pretested subjects who had volunteered to participate in the experiment were more likely to show change consistent with the communication than were nonpretested volunteers. But pretested subjects who had not volunteered were less likely to show change than were nonpretested subjects. Thus, pretesting facilitated change among volunteers and inhibited change among nonvolunteers. One could infer that volunteers were trying to help the experiment work while nonvolunteers were resisting influence.

Whether (voluntary) respondents in any health campaign study are going to be as likely to willingly accept the communication as Rosnow and Suls' volunteers is unknown. Nevertheless, the issue of pretest sensitization seems sufficiently unresolved to warrant the use of the Solomon four group design in any study using pretests and post-tests.

The Solomon four group design was used as part of the evaluation of "Feeling Good" conducted by both the National Opinion Research Center (Minor and Bradburn, 1976) and by the Response Analysis Corporation (Cohen and Abelson, 1976). The NORC study found no sensitization effects, but the RAC study found that pretested subjects were generally higher in knowledge at the post-test while subjects who did not get the pretest were frequently higher in reported levels of health behaviors performed in the recent past. Cohen and Abelson suggest that pretested subjects learned from the pretest itself while the greater behavioral effects in the non-pretested group probably resulted from "telescoping." Telescoping is more controlled in the pretested group because it was "picked up" by the pretest.

Measuring the Dependent Variable

There is a variety of ways of measuring different types of dependent variables. With beliefs, attitudes, and knowledge, one is restricted to some type of self-report. With behavioral dependent variables or physiological states such as blood pressure or weight, either self-reports or direct observation are possible.

Self-reports have the advantage of being very inexpensive; they can be gathered during any contact with the subject. But self-reports can be unreliable, one of the constant problems being social desirability. (See Crano and Brewer, 1973, for a discussion.) If social norms dictate that a certain response is acceptable and other responses are not, then subjects may be reluctant to admit to violating those norms. Voting in elections is dictated by good citizenship, for example, and studies indicate that voting is over-reported in surveys by about 10 percent (Clausen, 1968-69).

On the other hand, direct observation of behaviors or physiological states can be very reliable, particularly if the observer is blind to the hypotheses and/or experimental manipulations. But direct observation is very expensive relative to self-report, and it usually destroys anonymity. Fortunately, many behavioral and physiological measures (such as weight) can be accurately reported by the individual. One can rely chiefly on self-reports of these behaviors and states, validating them with direct observation on a small subset of the population. Such a combination should provide an indication of the extent of bias or unreliability, if any, in self-report responses.

Data Gathering Techniques

One of the most popular among the variety of data gathering techniques available is a laboratory study which allows the researcher a great deal of control over experimental procedures. When one is studying the effectiveness of a media campaign, however, the laboratory setting is sufficiently different from any natural environment where the media message could be heard as to limit generalizations of the findings to the real world. Laboratory research is, therefore, not recommended for studying the effectiveness of media campaigns, at least not as the major thrust of the investigation. This is not to say laboratory research cannot be useful in a media campaign, particularly in determining effective communication.

The most common data gathering technique for studying media effects is the survey of which there are three basic types. Mail surveys have the advantage of being very cheap to administer; almost everyone with

an address can be contacted. But a good sampling frame (a complete list of all addresses from which a random sample of addresses will be drawn) is difficult to construct. Moreover, only simple, direct questions can be asked, and no followup or elaboration of answers is possible. The chief drawback of mail surveys is that the response or return rate is low relative to other types of surveys: only the very interested or highly motivated will take the initiative to answer the questionnaire and return it to the sponsor of the study.⁴ Therefore, the typical mail survey is not very representative of the population as a whole. (See Dillman, 1978 and Heberlein and Baumgartner, 1978, for recent reviews of issues concerning mail surveys.)

In person interviews are generally preferred by most authorities on survey research (e.g., Babbie, 1973; Selltitz et al., 1976; Weisberg and Bowen, 1977). Ideally the interviewer can develop a good rapport with the respondent and collect accurate, truthful, and even intimate data. Interviews of an hour or longer are common, and a great deal of probing and followup of answers is possible. The biggest disadvantage of in-person interviews is their cost; it is very expensive to train and transport interviewers. Cluster sampling (randomly choosing a number of geographic clusters such as census tracts or congressional districts and then randomly selecting a number of individuals from each chosen cluster) can cause problems due to the non-representativeness of the sample. Data gathering is relatively slow, and the response rates, once considered a chief advantage of in-person interviews, may be falling in large urban areas (Marquis, 1977; Quinn et al., 1978); potential respondents are reluctant to admit a stranger into their home for an interview and interviewers are reluctant to go into certain less safe areas of cities.

Telephone interviews are a good compromise between mail and in-person surveys. They are a relatively cheap and very fast method of surveying: As much probing and followup of answers is possible over the telephone as in person, and telephone surveys of thirty minutes or longer are not atypical. The cost is only one-third as much per respondent as for in-person interviews. Response rates may be

somewhat lower compared to in-person surveys, but training and monitoring phone interviewers is much easier. Some comparisons of telephone and in-person surveys conclude that data quality (e.g., proportion of missing data, number or length of open-ended questions) is lower with telephone surveys (Bushery et al., 1978; Groves and Scott, 1976; Hochstim, 1967); other studies report the opposite conclusion or find no differences between telephone and in-person surveys (Hochstim, 1962; Locander et al., 1974; Lucas and Adams, 1977; Quinn et al., 1978; Selltitz et al., 1977). The greatest limitation to phone surveys is usually considered drawing a representative sample: There is a bias in telephone ownership and in having one's number listed in the telephone book (an obvious sampling frame for a telephone survey). The advent of random digit dialing has eliminated the latter problem, and recent research suggests that "by 1976 saturation [of telephone ownership] was so high that the exclusion of non-telephone households was no longer a liability for telephone survey sampling in most parts of the country (Lucas and Adams, 1977, p. v). Telephones also offer anonymity which may prove a great advantage with certain types of data. With random-digit dialing, a panel can be followed and even prompted without ever having to know the respondents' names or addresses.

A final method of data gathering is categorized under the general rubric of unobtrusive measures (Webb et al., 1966), so named because the information is gathered without contacting individual respondents. These techniques are possible when the dependent variable of interest is some easily observable behavior (Robertson et al., 1974) such as sales of certain products (Udry et al., 1972) or attendance at special clinics (Bailey, 1973). Unobtrusive measures, however, are not very good at picking up small or subtle effects of the media campaign. Changes in beliefs and knowledge cannot be measured very reliably, and unless one is willing to accept serious violations of individual privacy, individual-level data is difficult to obtain.

Methodological Problems Specific to Individual Designs

Although problems inherent in each of the four basic designs used to study the effectiveness of health campaigns in the mass media have been mentioned previously, it is important to discuss them in greater detail. First, the correlational studies suffer from an irrefutable and generally fatal self-selection bias. One cannot simply compare viewers to nonviewers and draw any definite conclusions about the effectiveness of any media message, because viewers and nonviewers are most likely not comparable groups.⁵

Time series *per se* refers simply to repeated measurement (usually ten or more observations) at the same unit of analysis, but virtually all time series studies use aggregate, as opposed to individual, level data. As such, they almost always use unobtrusive measures and therefore share the problems and limitations of those measures. Aggregate data do not allow nearly so precise an understanding of the process under study as do individual data; the richness of individual differences and individual subjective reactions to media messages are lost. Moreover, as the name implies, time series analyses entail availability of data over an extended period of time. The data gathering itself must not change over that time. Similarly, changes in the environment or in the population under study that are independent of the media campaign which could result in changes in the dependent variable being measured are always a possible source of confounding. Variations on basic time series designs like multiple or repeated time series do help control for this confounding (Campbell and Stanley, 1963).

The nonequivalent control group design involves three primary problems, one being the added expense of setting up separate research sites in different locations. If the towns or localities are relatively close together, increasing the possibility they will have shared media, this added expense might be minimal. A second problem arises if some event affecting the dependent variable being measured occurs during the media campaign in one town but not another. If, for example, one were studying the effectiveness of a campaign for getting flu shots and some prominent figure in one of the towns died from flu

during the study, this might very well motivate a number of people in the town to get a flu shot who would not otherwise have done so. Although this problem can never be completely eliminated, if more than one experimental and one control city are used the likelihood is greatly reduced that some extraneous event will confound the events of the study. It does, of course, lead to more research sites and increased costs.

The third problem with the nonequivalent control group design is that, as the name implies, the control and experimental cities are never exactly comparable. Whereas two small communities in close geographic proximity might be reasonably similar, the larger the communities are or the greater the distance between them, the less confidence one can have in their comparability. Nevertheless, it is often a reasonable assumption that the control and experimental cities do not differ in any theoretically important way. But it is always an assumption. If, however, four or more cities are studied and if they are randomly assigned to experimental or control status, then one is more confident that the experimental and control groups are equivalent. But again, as more communities are added the costs of the design increase proportionately.

Only two examples were found of the design in which half of the sample is induced or prompted to view, read, or listen to the media message in an effort to create experimental and control groups differing only in the proportion encountering the campaign's message. One such study is the evaluation of "Feeling Good" described above. In the second, five sites were selected around the country for an evaluation of the first year of the children's television program "Sesame Street" (Cook et al., 1975). Four of these sites were lower socioeconomic states (SES) areas and one was a middle class suburb. Within each site, two-thirds of the households selected for study were randomly assigned to the encouragement or experimental group. This group was told about the program, given publicity material about the show, and visited once a week by an evaluator during the show. The control group did not receive these inducements. Hence, the experimental group differed from the control group in ways other than the percentage of

children who watched the show, particularly in the presence of an outside evaluator once a week during the telecast.

There are two chief problems with the prompt design, and the "Sesame Street" study suffered from both of them. The first is that the inducement or prompt may not be effective enough to create an experimental group having significantly more viewers than the control group. This failure can occur either when the prompt is weak and few people in the prompted group watch the communication or when the communication proves so popular that a high proportion of the non-prompted group watch it. In either event, when relatively similar proportions of both groups see a communication, any effect of the communication is underestimated by comparing experimental and control groups. In the two sites highest in SES involved in the "Sesame Street" evaluation, the experimental and control groups did not differ in the proportion of viewers. This resulted primary from the unexpected popularity of "Sesame Street" rather than the ineffectiveness of the inducements. Nevertheless, no differences could be expected (and none were found) between experimental and control groups in these two sites.

The authors of the NORC evaluation of "Feeling Good" also complained that the proportion of viewers in the experimental (induced) group and control (noninduced) groups did not differ as much as was hoped. By the middle of the second season, 23 percent of the control group had seen at least some of the "Feeling Good" shows, whereas 77 percent of the induced viewers had seen some programs and 62 percent had seen at least half. Considering that almost two-thirds of the control group had been asked at least once during the course of the study if they had watched "Feeling Good" (which would seem to be a mild stimulant to watch the show) it is probably unrealistic to expect to control viewing more than this.

The two studies suggest that simply encouraging the experimental group to watch the program of interest may not be enough to insure that the proportion of viewers in that group will be greater than the proportion of viewers in the unencouraged or noninduced control group. If the program is very popular anyway, then it might be necessary to

discourage the control group from watching, or even better to encourage the control group to watch some other program which precludes their viewing the experimental communication. Such a design would make the treatment of the experimental and control groups more nearly comparable, for both can get the same type of encouragement although directed at different sources. If one's message, rather than being a completely new show with an unknown audience, is shown or heard during an already existing program, it would be possible to use the program ratings to estimate the audience for the message. If this proportion of viewers can be tolerated in the control group, discouragement or inducements to watch another program may not be necessary.

The second problem with this design is that the prompt or encouragement may prove reactive, that is, the effects of the communication on people who watch after they are prompted to do so may be different from the effect of the communication on nonprompted viewers. Certain types of prompts clearly will be more reactive than others, but although no research was identified that specifically assessed the effect of prompts or inducements on subject response,⁶ two studies offer some insight into the problem.

In their study of the evaluations of "Sesame Street" Cook et al. (1975) concluded that, while there was solid evidence that in the three sites where the proportion of viewers differed between the experimental and control groups (the two groups differed on a variety of measures), there was little evidence that viewers in the noninduced group improved more than nonviewers in this same noninduced group. The combination of the inducement to watch the show and the actual watching of the show produced the most change. The type of prompt used in the "Sesame Street" evaluation certainly had the potential of being very reactive. In defense of the original evaluators, it should be mentioned that the outside evaluator/prompter in the home served an important purpose in the evaluation other than encouraging the child to watch the program under study--recording exactly what the children were doing while they were watching.

As previously mentioned, the NORC evaluation included a manipulation to examine the effects of monetary inducement. A portion of the control

group was also paid, but they were told this payment was for participating in the study and there was no mention of "Feeling Good." Hence, induced and noninduced members of the control group could be compared, and any differences between the two could be attributed to the inducement. In more than 100 such comparisons, less than five percent were significantly different--the number expected by chance if there were no actual differences (Minor and Bradburn, 1976). Hence, the authors' conclusion that there were no important inducement confounds in the study.

While it is certainly not inevitable in utilizing this design that any inducement or prompt will prove reactive and confound the results, an attempt should be made to develop a prompt that is as nonreactive as possible. Moreover, it would also seem advisable to build into the overall evaluation design some method for determining if the prompt used were reactive, such as that shown in Figure 2 or the one used by the NORC evaluators of "Feeling Good."

AN EXAMPLE OF AN IDEAL EVALUATION OF A HEALTH-RELATED MEDIA CAMPAIGN

The major points of the previous section are summarized in Table 3 as a model for an ideal evaluation of a media health campaign. The campaign itself should deal with an important health problem, one with serious consequences for a large number of people, but also one which can be prevented or cured. The campaign should concentrate on a single health topic rather than deal with a diffuse conglomerate of health hints. The dose and duration of the campaign should be large in order to increase the effectiveness of the campaign and thus the possibility of measuring that effect. A representative sample of the population should be selected for study; if specific subgroups of the population are of interest, however, those subgroups might be oversampled to the point of excluding everyone else. Although the general design utilized will always be a function of the particular setting and the circumstances of the study, any design should allow for more definitive causal statements than does a simple correlational study. Moreover, it should also take advantage of the added power of a pretest-post-test design;

Table 3

Components of an Ideal Evaluation of a Media Health Campaign

Steps	Critical Elements
1. Selecting a Health Problem	Seriousness, prevalence, effectiveness of remedial or preventive actions
2. Breadth of Stimulus	Single health topic
3. Dose and Duration	The larger and longer the better
4. Sampling	Representative survey, perhaps oversampling specific target groups
5. Design Considerations	<p>Experimental or quasi-experimental design</p> <p>a. Nonequivalent control group design, if small communities with separate media are of interest</p> <p>b. Inducements for watching if single large community is being studied; include techniques for minimizing and measuring reactivity of prompt</p> <p>c. Incorporate Solomon four group design into a or b; pretest-posttest more powerful in detecting small effects</p>
6. Data Gathering Technique	Probably telephone survey, possibly with mailed followup
7. Measurement of Dependent Variable	Self-report, with direct observation of subsample for validation

the Solomon four group technique should be utilized to measure any possible sensitization effects. A telephone survey is probably the most efficient way of gathering data, although after establishing contact with a respondent, followup measures may be gathered by mail questionnaire if the loss of anonymity is acceptable. Reliance on self-report measures for most dependent variables of interest seems necessary, although some direct observation of the dependent variable is desirable to determine whether there is any report bias.

FOOTNOTES

1. Random assignment is fairly easy to accomplish if the communication is mailed (or not mailed) to subjects. Greenberg et al. (1953) selected a random sample of new parents in North Carolina. Shortly after the birth of their first child and continuing for a year, half of these parents were mailed a series of pamphlets entitled "Pierre the Pelican." These pamphlets, which arrived once a month, were designed to cover the major points of good child rearing. After a year all mothers in the sample were interviewed. No differences were found between mothers who had received or not received the pamphlets in handling their children's eating habits. Marginal differences (in the recommended direction) were found for nonwhite mothers who had received the pamphlets in encouraging independence in their children.
2. The Gallup study did include a very interesting technological innovation. When it appeared that self-reported performance of a variety of health behaviors in the past two months seemed unusually high in the first two surveys, a small experiment was conducted during the third survey. It was suspected that many respondents were "telescoping," reporting they had performed a certain behavior in the previous two months when it had actually been longer than two months since they had performed it, so half of the respondents in the third survey were simply asked which of a variety of health behaviors they had performed in the past two months. This was exactly the way the question had been asked in the previous two surveys. The other half of the respondents were first asked which of those behaviors they had performed in the previous six months, and then were asked which they had performed in the previous two months. It was hoped that this would limit the amount of telescoping by providing it with a previous outlet (the six-month question) and by more clearly delimiting the two-month period of interest. There was clear evidence for telescoping. For instance, 31 percent of the respondents who were simply asked if they had a physical examination in the past two months said they had, but only 18 percent of the respondents who were first asked if they had a physical examination in the past six months also said it had been in the previous two months. On 14 health behaviors, the mean over-reporting due to telescoping was 11.5 percent.
3. Power is a function of the variance in the dependent variable. The smaller the variance, the greater the power. A good rule of thumb is that the larger the sample, the smaller the variance. The more heterogeneous the sample, the larger the variance. If one knows how large an effect or how large a difference between experimental and control groups one would like to be able to detect and if one can estimate the variance in the dependent variable, one can calculate how large a sample will be required (Cohen, 1977).

4. Response or return rates of mail surveys are, of course, subject to the same variables that affect response rates to any survey: the type and interest of the population being surveyed, the prestige of the institution sponsoring the survey, the length of the questionnaire, the number of call (or mail) backs to nonrespondents, and (specifically for mail surveys) the use of a special class of mail or telephone calls on the last contact with nonrespondents.
5. A recently developed multivariate, non-matching technique (Sherwood et al., 1975) could prove useful in drawing stronger conclusions from correlational studies with self-selection biases like these. In the past, the standard remedy for self-selected groups was to match individuals from each group on pretest scores and use these matched individuals (ignoring people in each group who could not be matched) as experimental groups. The problem with matching like this is that other unmatched factors which affect the dependent variable are also confounded with the self-selected groups. Therefore, controlling on pretest scores by matching does not control all relevant differences between self-selected groups. However, most studies in the past which have attempted matching have matched on only a single variable. A multivariate matching technique (that is, matching on multiple variables) hopes to form a "net" which captures and controls for other, unmatched, unmeasured, not-thought-of variables which could affect the dependent variable. The more theoretically important confounding variables which have been measured and controlled for by matching, the more effective the net should be. This technique does not yet have the empirical support necessary for us to endorse or reject it, but it is an interesting and potentially useful idea.
6. After drawing a blank in our own efforts to find research on prompts, a number of experts in the evaluation field were contacted (UCLA, Northwestern, Florida State). These people could not reference any research on prompts either.

REFERENCES

- Ajzen, I and M Fishbein "Attitude-behavior Relations: A Theoretical Analysis and Review of Empirical Research" *Psychology Bulletin* 84: 888-918, 1977.
- Aldis, RM, et al. "Prenatal Teaching on Commercial TV" *Canadian Journal of Public Health* 60:315-20, 1969.
- Atkin, CK "Research Evidence on Mass Mediated Health Communication Campaigns" Paper presented at the National Academy of Sciences Conference on "Adolescent Behavior and Health" Washington, D.C., June 27, 1978.
- Babbie, E (1973) *Survey Research Methods* Belmont, Cal., Wadsworth.
- Bailey, J "An Evaluative Look at a Family Planning Radio Campaign in Latin America" *Studies in Family Planning* 4:275-8, 1973.
- Baran, S "TV Programming and Attitudes Toward Mental Retardation" *Journalism Quarterly* 54:140-2, 1977.
- Bauer, R "The Obstinate Audience" *American Psychologist* 19:319-28, 1964.
- Becker, MH, RH Drachman and JP Kirscht "A New Approach to Explaining Sick Role Behavior in Low Income Populations" *American Journal of Public Health* 64:205-16, 1974.
- Brown, R (1965) *Social Psychology* New York, Free Press.
- Bush, WV (1965) *The National Drivers Test* New York, Random House.
- Bushery, JM, CD Cowan and LR Murphy "Experiments in Telephone/Personal-Visit Surveys" Paper presented at the 1978 meeting of the American Statistical Association, San Diego, Calif.
- Campbell, DT and JC Stanley (1963) *Experimental and Quasi-experimental Designs of Research* Chicago, Rand-McNally.
- Cartwright, D "Some Principles of Mass Persuasion" *Human Relations* 2:253-67, 1949.
- Clausen, RR "Response Validity for Surveys" *Public Opinion Quarterly* 32:588-606, 1968-69.
- Cohen, J (1977) *Statistical Power Analysis for the Behavioral Sciences* San Francisco, Academic Press.
- Cohen, MN and HI Abelson (1976) *Impacts, Benefit and Consequences of "Feeling Good"* New Jersey, Response Analysis Corporation.

- Comstock, G and M Fisher (1975) *Television and Human Behavior: A Guide to the Pertinent Scientific Literature* R1746-CF, Santa Monica, Cal. The Rand Corporation.
- Cook, TD et al. (1975) *Sesame Street Revisited: A Study in Evaluation Research* New York, Russell Sage.
- Crano, WD and MB Brewer (1973) *Principles of Research in Social Psychology* San Francisco, McGraw-Hill.
- Deutscher, I "Words and Deeds: Social Science and Social Policy" *Social Problems* 13:235-65, 1966.
- Dickman, FB and TJ Keil "Public Television and Public Health: The Case of Alcoholism" *Journal for the Study of Alcoholism* 38:584-92, 1977.
- Dillman, D (1978) *Mail and Telephone Surveys...The Total Design Method* New York, Wiley.
- Douglas, D, B Wertley and S Chaffie "An Information Campaign That Changed Community Attitudes" *Journalism Quarterly* 47:220-7, 1970.
- Edwards, JE, EC Penick and B Suway "Evaluating the Use of Television in Community Mental Health Education" *Hospital and Community Psychiatry* 24:771-3, 1973.
- Farquhar, JW et al. "Community Education for Cardiovascular Health" *Lancet* 1:1192-8, 1977.
- Ferstl, R et al. (1977) *Die Beeinflussung des Übergewichts* Munich, Ger. The Max Planck Institute for Psychiatry cited in A Stunkard "Behavioral Medicine--and Beyond: The Example of Obesity" in JP Brady and OF Pomerleau (eds) *Behavioral Medicine: Theory and Practice* New York, Williams & Wilkins (in press).
- Fishbein, M "Attitudes and the Prediction of Behavior" in M Fishbein (ed) (1967) *Readings in Attitude Theory and Measurement* New York, Wiley.
- Fleisher, GA "A Study of a Radio/TV Campaign on Safety Belt Use" *Journal of Safety Research* 5:3-11, 1973.
- The Gallup Organization (1975) *The National Audience for "Feeling Good"* Princeton, NY, The Gallup Organization.
- Greenberg, B and W Gantz "Public Television and Taboo Topics: The Impact of 'VD Blues'" *Public Telecommunications Review* 4:56-9, 1976.
- Greenberg, BG, et al. "A Method for Evaluating the Effectiveness of Health Education Literature" *American Journal of Public Health* 43:1147-55, 1953.

- Gross, LP "Can We Influence Behavior to Promote Good Nutrition?" *Bulletin of the New York Academy of Medicine* 47:613-23, 1971.
- Groves, RM and SC Scott "An Attempt to Measure the Relative Efficiency of Telephone Surveys for Social Science Data Collection" Paper presented at the American Association for Public Opinion Research Conference, Ashville, NC, 1976.
- Haskins, JB "Effects of Safety Communication Campaigns: A Review of the Research Evidence" *Journal of Safety Research* 1:58-66, 1969.
- Heberlein, TA and R Baumgartner "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature" *American Sociological Review* 43:447-62, 1978.
- Hochstim, JR "Comparison of Three Information-gathering Strategies in a Population Study of Sociomedical Variables" *Proceedings of the Social Statistics Section American Statistical Association*, 1962.
- Hochstim, JR "A Critical Comparison of Three Strategies of Collecting Data from Households" *Journal of the American Statistical Assoc.* 62:976-89, 1967.
- Insko, CA and J Schopler "Triadic Consistency: A Statement of Affective-Cognitive-Conative Consistency" *Psychological Review* 74:361-76, 1967.
- Katz, E and JJ Feldman "The Kennedy-Nixon Debates: A Survey of Surveys" in WS Cratty (ed.) (1970) *Public Opinion and Politics: A Reader* New York, Holt, Rinehart and Winston, 1970.
- Kline, FG, P Miller and A Morrison "Adolescents and Family Planning Information: An Exploration of Audience Needs and Media Effects" in J Blumler and E Katz (eds) (1974) *The Uses of Mass Communications: Current Perspectives on Gratifications Research* Beverly Hills, Cal., Sage Publications.
- Krech, D and RS Crutchfield (1948) *Theory and Problems in Social Psychology* New York, McGraw-Hill.
- Lana, RE "Pretest Sensitization" in R Rosenthal and R Rosnow (eds) (1969) *Artifact in Behavioral Research* New York, Academic Press.
- Locander, W, S Sudman and N Bradburn "An Investigation of Interview Method, Threat and Response Distortion" *Proceedings of the Social Statistics Section American Statistical Association*, 1974.
- Looney, G "Television as a Major Factor in the Ecology of Childhood" *Clinical Pediatrics* 10:445-8, 1971.
- Lucas, WA and WC Adams (1977) *An Assessment of Telephone Survey Methods* R-2135-NSF, Santa Monica, Cal. The Rand Corporation.

- Maccoby, N et al. "Reducing the Risk of Cardiovascular Disease" *Journal of Community Health* 3:100-14, 1977.
- Mark, N "How Television Tries to Close the Health Information Gap" *Today's Health* 54:31-3, 52, 1976.
- Marquis, KH (1977) *Survey Response Rates: Some Trends, Causes and Correlates* P-5863, Santa Monica, Cal., The Rand Corporation.
- Marshall, CL (1977) *Toward an Educated Health Consumer: Mass Communication and Quality in Medical Care* (Teaching of Preventive Medicine monograph #7, US DHEW Publ # (NIH) 77-881) Washington, DC, GPO.
- Massarik, F et al. "Community Based Genetic Education, Communication Channels and Knowledge of Tay-Sachs Disease" *Progress in Clinical and Biological Research* 18:353-66, 1977.
- Mendelsohn, HA "Some Reasons Why Information Campaigns Can Succeed" *Public Opinion Quarterly* 37:50-61, 1973.
- Mielke, KW and JW Swinehart (1976) *Evaluation of the "Feeling Good" Television Series* New York, Children's Television Workshop.
- Minor, MJ and NM Bradburn (1976) *The Effects of Viewing "Feeling Good": Results from a Field Experiment in a Low-income Community* Chicago, National Opinion Research Center.
- Morrison, AF, FG Kline and P Miller "Aspects of Adolescent Information Acquisition About Drugs and Alcohol Topics" in R Ortman (ed) (1976) *Communication Research and Drug Education* Beverly Hills, Cal., Sage Publications.
- Nash, JO "Mass Media and Behavior Change" unpublished manuscript, Stanford University, 1976.
- Nash, JD and JW Farquhar "Community Approaches to Dietary Modification and Obesity" in AJ Stunkard (ed) (1978) *Psychiatric Clinics of North America: Obesity* Philadelphia, WB Saunders.
- Newcomb, TM, RH Turner and PE Converse (1964) *Social Psychology* NY, Holt.
- AC Nielsen Company (1976) *Nielsen Television Index Report on "Feeling Good"* New York, AC Nielsen.
- O'Keefe, MT "The Anti-smoking Commercials: A Study of Television's Impact on Behavior" *Public Opinion Quarterly* 35:242-8, 1971.
- Penick, S et al. "Behavior Modification in the Treatment of Obesity" *Psychosomatic Medicine* 33:49-55, 1971.
- Quinn, RP, BA Gutek and JT Walsh "Telephone Interviewing: A Reappraisal and a Field Experiment" (mimeo) 1978.

- Regan, DT and R Fazio "On the Consistency Between Attitudes and Behavior: Look to the Method of Attitude Formation" *Journal of Experimental Social Psychology* 13:28-47, 1977.
- Rimer, II "The Impact of Mass Media on Cancer Control Programs" in JW Cullen et al. (eds) (1976) *Cancer: The Behavioral Dimensions* NY Raven Press.
- Robertson L et al. "A Controlled Study of the Effect of Television Messages on Safety Belt Use" *American Journal of Public Health* 64: 1071-80, 1974.
- Rosenthal R and RL Rosnow (eds) (1969) *Artifact in Behavioral Research* New York, Academic Press.
- Rosnow, RL and JM Suls "Reactive Effects of Pretesting in Attitude Research" *Journal of Personality and Social Psychology* 15:338-43, 1970.
- Salzar, JE, CL Marshall and ER Glazer "The Use of Cable Television as a Tool in Health Education of the Elderly: Screening" *Health Education Monographs* 5:363-78, 1977.
- Schmeling, D and E Wotring "Agenda Setting Effects of Drug Abuse Public Service Ads" *Journalism Quarterly* 53:743-6, 1976.
- Sears, DO "The Debates in the Light of Research: An Overview of the Effects" Paper presented at the 73rd Annual Meeting of the American Political Science Association, Washington, D.C., 9/3/77.
- Sears, DO and RE Whitney "Political Persuasion" in IdS Pool and W Schramm (eds) (1974) *Handbook of Communication* Chicago, Rand-McNally.
- Secord, PF and GW Backman (1964) *Social Psychology* New York, McGraw Hill.
- Selltiz, C, L Wrightsman and SW Cook (1976) *Research Methods in Social Relations* New York, Holt.
- Sherif, M and H Cantril "The Psychology of Attitudes: I" *Psychological Review* 52:295-319, 1945.
- Sherif, M and H Cantril "The Psychology of Attitudes: II" *Psychological Review* 53:1-24, 1946.
- Sherwood, CD, JN Morris and S Sherwood "A Multivariate, Nonrandomized Matching Technique for Studying the Impact of Social Interventions" in EL Struening and M Guttentag (eds) (1975) *Handbook of Evaluation Research (Vol. 1)* Beverly Hills, Cal., Sage.
- Shuman, H and MP Johnson "Attitudes and Behavior" *Annual Review of Sociology* 2:161-207, 1976.

- Smith, FA et al. "Health Information During a Week of Television" *New England Journal of Medicine* 286:516-20, 1972.
- Solomon, RL "An Extension of Control Group Design" *Psychological Bulletin* 46:137-50, 1949.
- Star, SA and HM Hughes "Report of an Educational Campaign: The Cincinnati Plan for the United Nations" *American Journal of Sociology* 55:389-400, 1950.
- Stern, MP et al. "Results of a Two-year Health Education Campaign of Dietary Behavior" *Circulation* 54:826-33, 1976.
- Stuart, RB (1972) *Slim Chance in a Fat World: Behavioral Control of Obesity* Champaign, Ill., Research Press.
- Thurstone, LL "The Measurement of Attitudes" *Journal of Abnormal and Social Psychology* 26:249-69, 1931.
- Udry, JR et al. "Can Mass Media Advertising Increase Contraceptive Use?" *Family Planning Perspectives* 4:37-44, 1972.
- Wade, SE and W Schramm "The Mass Media as Sources of Public Affairs, Science, and Health Knowledge" *Public Opinion Quarterly* 33:197-209, 1969.
- Warner, KE "The Effects of the Anti-smoking Campaign on Cigarette Consumption" *American Journal of Public Health* 67:645-50, 1977.
- Webb, EJ et al. (1966) *Unobtrusive Measures: Nonreactive Research in the Social Sciences* Chicago, Rand McNally, 1966.
- Weisberg, HF and BD Bowen (1977) *An Introduction to Survey Research and Data Analysis* San Francisco, Freeman.
- Wicker, AW "Attitude Versus Action: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects" *Journal of Social Issues* 25:41-78, 1969.
- Wright, WR "Mass Media as Sources of Medical Information" *Journal of Communication* 25:171-3, 1975.
- Yarnell, J "Evaluation of Health Education: The Use of a Model of Preventive Health Behavior" *Social Science and Medicine* 10:393-8, 1976.
- Young, M "Review of Research in Studies on the Health Education and Related Aspects of Family Planning" *Health Education Monographs* 1:13-30, 1973.

