UNINTENTIONAL CROSSOVER EFFECTS IN COMMUNITY INTERVENTION STUDIES: EVIDENCE FROM A NUTRITION SURVEY AIMED AT AFRICAN AMERICANS

Shelley P. Brock, Westat, Inc., Joseph V. Mauney and William D. Kalsbeek, University of North Carolina
Shelley P. Brock, Westat, Inc., 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: crossover, nonsampling error, telephone survey

1. INTRODUCTION

In the past twenty years, there has been tremendous growth in public health research in the area of community intervention trials. These kinds of trials have been defined by Koepsell as "programs (that) are aimed at entire populations, which are usually geographically defined, and they attempt to change health behavior and disease risk through mass media campaigns, activation of existing community organizations, or changes in the physical or sociocultural environment" [Koepsell (1992)].

One potential problem in community intervention trials occurs when study participants are not exposed to the regimen prescribed for them. This contamination problem, called unintentional crossover in this study, can occur in several ways. First, migration between intervention and control communities can occur, which compromises the assumption that responses of study subjects from different clusters are independent. If crossover intervention subjects adopt the qualities of control subjects, this can cause underestimation of the intervention effect. Another more likely form of crossover is for control community members to become aware of and be affected by intervention activities. This type of contamination in the study design also reduces the size of the estimated effect of intervention [Donner (1996)].

1.1 Purpose

Crossover in this survey appeared to be a likely possibility for several reasons. Eight of the ten counties chosen for the study were bordered by a county in the opposite treatment group, allowing for easier migration between treatment groups. Also, there was known existing networking within primary sampling units, and a known general desire to participate in intervention programs. The primary purpose of this research was to look at this survey data to see if there was crossover present, to measure the extent and direction of it, and to determine its effect on the intervention. Individuals who completed the survey were profiled and measured using a crossover scale. Generally, subjects fell into one of the following four groups regarding crossover:

- *Not Crossover Intervention Group* --- intervention subjects who showed little evidence of being influenced by the control group;
- *Crossover Intervention Group* --- intervention subjects who displayed evidence of crossing over to the control group;
- *Not Crossover Control Group* --- control subjects who showed little evidence of being influenced by the intervention group; and.
- *Crossover Control Group* --- control subjects who appeared to cross over to the intervention group.

Once the subjects were classified, analysis consisted of measurement of the effect of crossover on total daily fruit and vegetable consumption.

1.2 Description of the Five-A-Day for Better Health Study

The Five-A-Day for Better Health Study in the Black Churches United for Better Health Project, run through the Department of Nutrition at the University of North Carolina at Chapel Hill, addresses the issue of increasing fruit and vegetable consumption in African-American church members in rural counties of North Carolina. The study began in the summer of 1994 and continued through 1996. Between the baseline and follow-up surveys, a one-eighth sample of baseline respondents was interviewed to obtain interim results in the fall of 1995. Follow-up data collection results show the actual impact of the intervention programs. Data were collected by the Survey Research Unit in the Department of Biostatistics, using computer-assisted telephone interviewing (CATI).

Baseline interviews were completed with 3,737 church members (RR: 79%). In the interim study, 459 of the 565 selected members completed the telephone interview (RR: 83.8%). Respondents who completed the baseline, excluding those who became ineligible, were contacted again for the follow-up survey. Out of the 3,737 baseline respondents, attempts were made to contact 3,674 people for it, and 2,519 completed it (RR: 77.3%). For the purposes of this paper, only the baseline and follow-up data will be used.

Before baseline data collection, the ten counties selected for the study were paired according to certain demographic variables, including geographic location, total population, and percent African-American population. Within each pairing of counties, one county was randomly assigned to intervention treatment, and the other to control. Churches assigned to the intervention were given a list of suggestions for church activities involving consumption of fruits and vegetables, and were allowed to choose those that best fit their needs and resources.

Control churches were instructed to continue with normal activities including any nutrition programs in which they were currently participating, but not to implement any of the Five-A-Day intervention programs. However, some of the control churches wanted to participate in intervention activities anyway, and this may have had an influence on both their responses to the surveys or the actual activities they participated in over the course of the study. This influence may have manifested itself as social desirability bias, or one of the above discussed forms of sample contamination.

2. METHODS

2.1 Design and Sampling

The ten counties selected for the study were paired and one from each pair assigned randomly to intervention or control treatment. Within each county, churches were divided into two strata according to membership size (< 100 members and >= 100 members). Within each stratum, the churches were selected with probability proportional to size (PPS). In each county, three churches were selected from the first stratum, and two from the second , which resulted in fifty churches overall in the study. After randomization, 25 churches were assigned to intervention and 25 to control. There was no sampling of church members for the baseline data collection. All 5,596 members of the 50 churches selected for the study were contacted initially.

At the completion of baseline data collection, weights for each respondent were calculated in two steps. The first step, a preliminary weight, consisted of the inverse of the probability of selection of the church, which was calculated taking into account the size stratum of the church. A multiplicative weight adjustment computed in the second step was the inverse of the within-county response rate for each church. The adjusted weights were the product of the preliminary weight and the multiplicative adjustment.

2.2 Classification of Respondents

The first step in assessing the extent of crossover was to develop a crossover scale and classify each respondent at a certain level on it. Several questions were evaluated for their usefulness in this scale. Ten questions plus a geographical proximity indicator described later were selected for it. The questions are:

1. "Have you received any nutrition information from members of the nutrition action team?"
2. "Are you aware of your church having any lay health advisors?"
3. "Have you talked about health or nutrition questions with a lay health advisor?"
4. "How much did attending cooking classes or educational sessions cause you to try to eat more fruits and vegetables?"
5. "How much did reading printed materials such as pamphlets, posters, and bulletin boards cause you to try to eat more fruits and vegetables?"
6. "How much did talking with other church members cause you to try to eat more fruits and vegetables?"
7. "How much did having grocery store produce promotions cause you to try to eat more fruits and vegetables?"
8. "How much did using cookbooks or taste testing recipes cause you to try to eat more fruits and vegetables?"
9. "How much did hearing about fruits and vegetables from radio, television, or newspaper stories cause you to try to eat more fruits and vegetables?"
10. "How much did having fruits and vegetables served at church functions cause you to try to eat more fruits and vegetables?"

The response categories for questions 1 through 3 were yes, no, don't know and refused. For questions 4 through 10, they were not at all, a little, some, a lot, does not apply, don't know and refused.

In order to assign a crossover score to each respondent, the response to each question in the scale was evaluated. Since the questions referred to activities that should have taken place only in intervention churches, control group members who indicated that they had been exposed to the activities mentioned were considered to have some evidence of crossover. Conversely, since the activities mentioned in the crossover scale questions focused on activities in intervention churches, intervention subjects who responded negatively to the questions were assumed to have experienced some crossover from control subjects.

More specifically, a respondent had a point added to his or her crossover scale for each "incorrect" answer to one of the selected survey questions. In the control group, an answer of "yes" to questions 1 through 3, and a response of "a lot", "some", or "a little" for questions 4 through 10 resulted in a point being added to the crossover score since the questions should have been answered negatively in that group. In the intervention group, a point was added to a respondent's score each time he or she answered "no" to questions 1 through 3, or either "none" or "does not apply" to questions 4 through 10. Since item 2 is a question of awareness, a response of "don't know" could be interpreted similarly to a response of "no" and was also considered to be an incorrect answer.

A geographical proximity indicator was also included in the crossover scale since the ten North Carolina counties included in the study were all located in the eastern half of the state. It was assumed that members of churches that were within 20 miles of a city in the opposite treatment group may have experienced at least some influence from members of the opposite group.

The possible range for the crossover score of any respondent is 0 to 11. After individuals were scored, they were assigned to one of the four crossover groups described above. Cut-off points to define "crossover" and "not crossover" respondents were selected so that the respondent had to answer at least half of the questions included in the crossover scale incorrectly. Four cut-off points were chosen to gauge the sensitivity of findings to how crossover is defined operationally. They were: 1) 0-5 points considered not crossover, 6-11 considered crossover; 2) 0-6 points considered not crossover, 7-11 considered crossover; 3) 0-7 points considered not crossover, 8-11 considered crossover; and 4) 0-8 points considered not crossover, 9-11 considered crossover.

2.3 Definition of Variables

Once respondents were classified as having crossed over or not, the primary outcome most useful for checking for an effect of crossover on the intervention was the mean of the differences between follow-up and baseline total daily fruit and vegetable consumption. This outcome was most useful since it would be used to test for behavioral change in assessing the efficacy of the intervention. It was assessed for the following six groups: overall intervention, overall control, not crossover intervention, not crossover control, crossover intervention, and crossover control. To do this, it was necessary to define and compute total daily fruit and vegetable consumption for each of the follow-up respondents.

Baseline and follow-up measures of total daily fruit and vegetable consumption as defined by published preliminary results from the Five-A-Day study were used in this analysis as well. Total consumption was calculated using the following seven questions that were asked on both the follow-up and baseline interviewing instruments:

1. "In the past month, how often did you drink 100% orange juice or grapefruit juice?"
2. "In the past month, how often did you drink other 100% fruit juices, not counting fruit drinks?"
3. "In the past month, how often did you eat green salad (with or without other vegetables)?"
4. "In the past month, how often did you eat french fries or fried potatoes?"
5. "In the past month, how often did you eat baked, boiled, or mashed potatoes?"
6. "In the past month, how many servings of other vegetables did you eat, not counting salad or potatoes?"
7. "In the past month, how many servings of fruit did you have not counting juice?"

Answer choices for all seven items were: never, 1-3 times per month, 1-2 times per week, 3-4 times per week, 5-6 times per week, 1 time per day, 2 times per day, 3 times per day, 4 times per day, and 5 or more times per day. For each of these seven items, all respondents received a point score towards total daily consumption according to his or her answer choice. An answer of never was zero points, 1-3 times per month was 2/30=0.07 points, 1-2 times per week was 1.5/7=0.21 points, 3-4 times per week was 3.5/7=0.50 points, 5-6 times per week was 5.5/7=0.79 points, 1 time per day was one point, 2 times per day was two points, 3 times per day was three points, 4 times per day was four points, and 5 or more times per day was five points.

The final outcome variable for each respondent was calculated as the sum of the point scores for all seven survey questions. It was calculated at baseline and follow-up for all respondents who answered both surveys. However, total consumption was not calculated for any respondents who had missing data for any of the defining questions. A derivative variable, the difference between follow-up and baseline total daily fruit and vegetable consumption, was also used.

2.4 Analysis Plan

The first analysis task was to test the difference between intervention and control of the overall mean difference in total daily fruit and vegetable consumption for significance, in order to determine the overall change in the outcome for all respondents, regardless of crossover. This result also gave an estimate for the overall effect of the intervention programs. The next task was to look at the means of the differences between follow-up and baseline total daily consumption for all crossover group by treatment group combinations, and to compare these to the means for overall treatment group. These tests were done to determine the change in the outcome within overall treatment group as well as within treatment by crossover group. If the intervention programs were at least partially successful, it was expected that for the intervention group, not crossover respondents would have a higher mean baseline to follow-up difference than overall respondents, and crossover respondents would have a lower mean difference than overall respondents. For the control group, it was expected that not crossover respondents would have a lower mean baseline to follow-up difference than overall respondents, and that crossover respondents would have a higher mean difference than overall respondents, since crossover control group members were influenced by the intervention regimen in some way.

To take into account the complex study design, weighted estimates of means and standard errors for the outcome were calculated assuming a with-replacement stratified cluster sample by using SUDAAN, Version 6.42 [Shah (1996)]. All of the overall means and crossover group by treatment group means were tested for significance using a t statistic, SUDAAN weighted analysis standard errors, and a significance level cut-off of 0.05.

3. RESULTS

Using the definitions and the scale for measuring crossover described above, some degree of crossover is clearly present in this study (Table 1). As expected, the percent of crossover respondents, both control and intervention members, decreased as the crossover criteria became more strict (i.e., as the cut-off point increased). The percentage of controls that were classified as crossover was higher than that of the intervention group for all four cut-offs, ranging from 53.7% to 15.6% while the corresponding intervention percentage of crossover ranged from 11.7% to 1.4%.

The mean difference of total daily fruit and vegetable consumption between follow-up and baseline for the intervention group was 0.62 servings. For the controls the mean was -0.15 servings, yielding an overall difference of 0.78 servings between the two groups ($p<0.0001$). Thus, even in the presence of crossover these numbers indicate that overall, intervention was better than control (Table 2).

As illustrated in Table 3, the not crossover intervention group had a larger mean difference than the crossover intervention group for all four cut-off points. In the control group, the mean difference for the crossover subjects was larger than that of the not crossover controls at all cut-off points, as expected.

Statistically significant *positive* mean differences were observed for the not crossover intervention subgroup at each cut-off point, and significant *negative* mean differences were observed at each cut-off for the not crossover controls (i.e., respondents in that subgroup were eating significantly fewer fruits and vegetables each day at the time of the follow-up interview than they were at baseline).

4. DISCUSSION

In this analysis, the mean difference observed in the overall intervention group was substantially higher than in the overall control group for all cut-off points, and the difference between these two was statistically significant, indicating that the intervention programs were successful. This was the expected result for the study, regardless of crossover. However, after respondents were assigned to crossover groups, the intervention programs appeared to be even more effective, since the not crossover intervention group showed more fruit and vegetable consumption than the overall intervention group, and all of the negative values in the control groups for all cut-off points showed as significant. These results also indicated that crossing over seemed to exist in these data, and that the crossover scale developed was a satisfactory method for measuring it.

As part of the interim survey data analysis, we searched for characteristics in each treatment group that could have influenced crossover. We found that attendance at church functions seemed to prevent crossover, and that less educated males were more likely to cross over. Finding these possible explanatory variables prompted us to continue the crossover analysis on the follow-up survey data.

Some related limitations of this study must be noted. One of the them is the general difficulty we had in operationalizing the concept of a "crossover," to delineate those in the Five-a-Day study samples who were to be considered as if they had been contaminated by the regimen of the opposite treatment group, from those who were to be treated as if they had not been contaminated. Our general approach was to accumulate a body of "circumstantial" evidence of crossover, based on answers to survey questions and geographic proximity to the opposite experimental group.

The survey questions used to define a crossover referenced either awareness of intervention activities or knowledge about fruit and vegetable consumption. For the purposes of this study, intervention respondents were considered as having crossed over, even if they were exposed to intervention activities, but answered survey questions as if they were not. Another issue is that there may have been social desirability bias, that we considered crossover, in the control group. Since higher fruit and vegetable consumption is a current secular trend in promotion of nutrition, indicating awareness of the intervention on the survey questions may have seemed like "correct" answers for some control group members. Our definition of crossover seemed to be reasonable for these particular data, but may not be for other types of surveys. Although the results followed the expected trends, much stronger results might be seen with more specific questions regarding the intervention programs, and a more standard way to measure responses for crossing over.

Another related measurement difficulty in this study was how to use the crossover scale to isolate crossover groups. This issue was addressed by doing the analyses using the four different crossover cut-off points, rather than trying to justify one of them. Although patterns of group differences were fairly consistent regardless of how we defined the crossover groups, crossover rates were clearly influenced by which cutoff we used for the scale. Thus we are unable to make a quantitative summary judgment about the extent of crossover in our study.

Beyond the matter of the extent of crossover is the impact it can have on findings. In the event where the crossover groups display results that are intermediate to the findings of the intervention and control, logic suggests the hypothesis that the likely effect of crossover or other forms of sample contamination in studies like ours is to *understate* intervention-to-control differences, thus making the intervention appear less effective than it actually is.

Our current work on the crossover problem explores this very issue. Early results, in which we use various statistical models to quantify actual intervention-to-control differences had there been no crossover, support our stated hypothesis. However, other modeling approaches may need to be explored to further support our contention. For instance, Imbens and Rubin (1997) address the related issue of noncompliance in studies involving human subjects. Viewing crossover as a form of noncompliance, their approach may provide another way to estimate real differences and thus to quantify the crossover effect in studies like ours.

In spite of these limitations, the results of this study provide some insight into the problem of crossover in the design of community intervention trials. While the results we have presented strongly suggest that crossover existed in this study, as it may in other similarly configured studies, it is very difficult to define and measure accurately. Further research is needed to develop better methods and tools for handling crossover and for measuring its effect on reported findings. Successful development in this area hopefully will lead to more accurate and effective community intervention trials.

REFERENCES

Donner, A (1996), personal e-mail communication.

Farquhar, JW, et al. (1978), "The Community-Based Model of Life Style Intervention Trials," *Am J Epid*, 108, 103-111.

Imbens, GW, Rubin, DB. (1997), "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance," *Annl Stat*, 25, 305-327.

Koepsell, TD, et al. (1992), "Selected Methodological Issues in Evaluating Community-Based Health Promotion and Disease Prevention Programs," *Annl Rev Pub Hlth*, 13, 31-57.

Mickey, RM, et al. (1991), "Estimation of the Design Effect in Community Intervention Studies," NY: John Wiley.

Shah, BV, et al (1996), SUDAAN User's Manual, Ver 6.4, 2nd ed, RTI, RTP, NC.

Stokes, ME, Davis, CS, Koch, GG (1995), Categorical Data Analysis Using the SAS® System, Cary, NC, SAS Institute Inc.

Table 1. Weighted percentage of crossover respondents by treatment group

Cut-off	Points	Classification	Control	Intervention
1	0-5	Not Crossover	46.3 (n=617)	88.3 (n=1040)
	6-11	Crossover	53.7 (n=699)	11.7 (n=151)
2	0-6	Not Crossover	61.1 (n=800)	92.2 (n=1091)
	7-11	Crossover	38.9 (n=516)	7.8 (n=100)
3	0-7	Not Crossover	74.6 (n=967)	95.7 (n=1135)
	8-11	Crossover	25.4 (n=349)	4.3 (n=56)
4	0-8	Not Crossover	84.4 (n=1108)	98.6 (n=1174)
	9-11	Crossover	15.6 (n=208)	1.4 (n=17)

Table 2. Weighted mean difference from baseline to follow-up of total daily fruit and vegetable consumption: overall intervention - overall control

	Mean Difference	Standard Error	t-test Statistic	Significance Level	Sample Size
All Cut-Offs	0.78	0.12	6.25	<0.00001	2507

Table 3. Weighted mean difference from baseline to follow-up of total daily fruit and vegetable consumption

Overall Control	Mean Difference	Standard Error	t-test Statistic	Significance Level	Sample Size
All Cut-Offs	-0.15	0.08	-1.83	0.03	1316
Not Crossover					
Cut-Off 1	-0.29	0.10	-2.80	0.003	617
Cut-Off 2	-0.27	0.08	-3.19	0.0007	800
Cut-Off 3	-0.24	0.08	-2.95	0.002	967
Cut-Off 4	-0.17	0.10	-1.71	0.04	1108
Crossover					
Cut-Off 1	-0.04	0.10	-0.37	0.35	699
Cut-Off 2	0.02	0.12	0.18	0.43	516
Cut-Off 3	0.11	0.18	0.57	0.29	349
Cut-Off 4	-0.08	0.31	-0.25	0.40	208
Overall Intervention					
All Cut-Offs	0.62	0.09	6.79	<0.00001	1191
Not Crossover					
Cut-Off 1	0.65	0.10	6.36	<0.00001	1040
Cut-Off 2	0.65	0.09	7.15	<0.00001	1091
Cut-Off 3	0.65	0.09	7.55	<0.00001	1135
Cut-Off 4	0.64	0.09	7.13	<0.00001	1174
Crossover					
Cut-Off 1	0.42	0.18	2.27	0.01	151
Cut-Off 2	0.25	0.25	0.96	0.17	100
Cut-Off 3	0.01	0.53	0.03	0.49	56
Cut-Off 4	-0.90	0.74	-1.22	0.12	17