DOCUMENT RESUME

ED 160 924 CG 012 859

AUTHOR	McKillip, Jack; Voss, Jacqueline R.
TITLE	Why Do We Need a Control Group? Why Should We Randomize? Some Answers for Evaluative Researchers.
PUB DATE	Mar 78
NOTE	12p.; Paper presented at the Annual Convention of the American Personnel and Guidance Association (Washington, D.C., March 19-23, 1978); Not available in hard copy due to marginal legibility
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	*Action Research; *Control Groups; *Evaluation Methods; Human Services; *Mental Health Programs; Program Evaluation; *Research Design; Research Methodology; Selection; *Social Science Research; State of the Art Reviews

ABSTRACT

There exists the practical possibility of randomly selected control groups for outcome assessments in mental health programs. Ethical considerations for randomization include distinguishing between innovation and "fooling around," and discriminatory implications of a first-come first-served selection procedure. Practical considerations include reviewing the differences between volunteers and non-volunteers and the utility of rigorous evaluation information. The strategies for implementing control groups in the research design, using administration and staff, include: (1) use of minimal program group; (2) separation of action and evaluation units; (3) inclusion of the lottery procedure on the informed consent form; and (4) allowance for particularly needy applicants. For dealing with participants the strategies include: use of a lottery for selection; informing potential participants about lottery; contacting and informing participants personally; when possible, providing subsequent services to the no-program group. (BN)

WHY DO WE NEED A CONTROL GROUP?

WHY SHOULD WE RANDOMIZE?

Some Answers for Evaluative Researchers

Jack McKillip

&

Jacqueline R. Voss

American Personnel and Guidance Association Convention

Washington, D. C., March, 1978

Please address comments to Jack McKillip, Department of Psychology, Southern Illinois University at Carbondale, Carbondale, Illinois 62901

WHY DO WE NEED A CONTROL GROUP?

WHY DO WE NEED TO RANDOMIZE?

Some Answers for Evaluative Researchers

Jack McKillip and Jacqueline R. Voss
Southern Illinois University; University of Washington

Our presentation focuses on the outcome assessment of social service programs, whether educational, therapeutic or growth-oriented. Such assessments ask if participants were better off after the program than they would have been if no or some different program had been offered. The assumption underlying our presentation is that this question cannot be answered absolutely. Evaluative data, like all social science information, rely on (at least implicit) comparisons in order to have meaning. Whether or not a program's 25% cure rate is impressive or disappointing depends on cure rates for similar programs. Especially in times of increased competition for social welfare dollars, statements such as "if only one is helped" cannot be taken seriously.

Accepting the implicit comparative nature of evaluative data, the best comparison for judging the postprogram status of participants is provided by the status of individuals as similar as possible to programs participants excepting that they have not participated in the program (although they may have participated in some alternative program). The question cannot be put simply: "Did the program do good?" but rather "Did the program do better tha[illegible]?" or "How did the program co[illegible] to some other program?" The bes[illegible] groups are those random[illegible] from the same population as those randomly selected for the program.

While most evaluative researchers would accept this presentation as far as it goes, many remain unconvinced of the practical possibility or necessity of randomly selected comparison groups for outcome assessments (e.g., Oetting, 1976; or Cronbach, 1977). For example, one survey of federally funded outcome evaluations of social service programs (Bernstein & Freeman, 1975) found that only one third utilized randomly selected comparison groups, while 43% did not include any comparison group. Drawing on our experience in implementing a number of evaluations (McKillip, 1978; McKillip & Kamens, 1977; Voss, 1978), it is our purpose to present information which will increase the practical possibility of the use of rigorous evaluation designs. Recognizing that the choice of design in evaluative research is not dictated soley by scientific concerns but, as in all other aspects of program evaluation, is strongly influenced by political factors, we review both ethical and practical arguments which we have found helpful in gaining acceptance of rigorous designs and then provide some "how to" suggestions.

Ethical Arguments

An extremely useful source in this area is Robert Boruch's (1976) collection of counter arguments to common objections to randomization. We have been quite successful in overcoming ehtical objections by taking an aggressive rather than a defensive posture using the following arguments.

When new programs or program alterations are introduced, human service organizations need to distinguish between innovation and fooling around (Gilbert, Light & Mosteller, 1975). While change is a positive organizational characteristic, it must be planned in ways which allow assessment of its value, e.g., unbiased comparison to other possible changes and/or to the established program. The medical literature provides a number of examples of innovation which are initially attested to by glowing clinical reports, but later upon more rigorous evaluation, are found to be useless or even harmful (Gilbert, et al., 1975). Innovation need not be slowed but needs to be done in a responsible manner. (A more whimsical approach can only lead to charges of quackery and to attempts at outside regulation). If the faith of program administrators and staff in the value of the change is not absolute and if plausible arguments for negative (side) effects can be devised, the thrust of this approach can be quite forceful.

For ongoing programs, a point of discussion can be the (usual) first-come-first-serve (FCFS) entry process. Especially where the potential demand exceeds the number of openings, we have been successful in arguing that the FCFS selection procedure shows a bias toward those in the know, with ready access to normal communication channels to the most assertive, while those eligible but less well connected or newer to the established system tend to be excluded. It is specifically this type of criticism that was raised most recently about services provided by community mental health centers by Nader's Raiders (Chu & Trotter, 1974). When wide advertisement combined with outreach efforts results in an excess of applications for the program affair and specifiable selection procedure is random selection of program participatns from all eligibles with those not selected serving as a no-program comparison grour

Practical Arguments

When participation in programs is voluntary, the volunteer will differ systematically from the non-volunteer. People do not buy into a program randomly but only on the basis of a particular need or attraction. Such people are not static beings and thus will probably mature or improve to some extent even without a program. An example of this process is provided by Voss's (1978) evaluation of a sex-education workshop which included a randomly chosen no-program comparison group equal in size to the number of applications the program received above the enrollment ceiling. While program participants showed significant increases in knowledge over a one month period, so did the no-program group!! In this case the bias caused by using a comparison group of non-participant, non-volunteers would have resulted in the illusion of a program effect when none existed. A similar bias, but in the opposite direction, results when program participants are deteriorating rather than improving (in the absence of a program), e.g., drug abuse or alcohol treatment programs. In this case the use of a comparison group from a different population could easily cover-up a real program effect.

A somewhat similar point is that volunteers usually differ from non-volunteers in levels of performance, so that for both groups, post-program status needs to be compared to pre-program status in order to gauge program effects. For example, McKillip and Kamens (1977) found that volunteers for a health education program had lower blood pressure, ate less junk food and engaged in more exercise than non-volunteers. In this case, mere comparison of post-program status for the two groups would lead to the inference of a program effect where no such inference is justified. (See also Garrard, 1972; and Zuckerman, Tushup, & Tinner, 1970.)

The strongest practical argument for rigorous evaluations rests on the utility of the resulting information. As the importance of the decision to be made increases, where there are doubts as to the likelihood of program success, or where viable alternative approaches exist, rigorous designs are most useful. There is little argument with the assertion that the best, the clearest, the most persuasive information on program effects come from the use of experimental designs. Where this type of information is needed rigorous evaluations are most practical.

A number of related points can be made. As Rossi (1972) points out, in our "modified" welfare state, with a high level of both mobility and literacy, strong human service effects are hard to find. Further, the more difficult it is to show an effect, the more necessary it is to use a rigorous design. In evaluation of the Salk polio vaccine a quasi-experimental design comparing treated volunteers with an undifferentiated non-treated group underestimated the effectiveness of the vaccine by 50% compared to an experimental design using a randomly selected non-treated group (Meier, 1972). Secondly, if we do not follow the comparison group as closely as we do the program group, it may happen that the comparison group will have received a "proxy" program without our knowledge. This may be a problem for evaluations of all designs, but will probably be less serious where the comparison group is specified and measured both before and after the program. Finally, it has been a frequent experience in evaluations that comparison groups which are not chosen randomly introduce a negative bias, i.e., their use makes it more difficult to show a program effect. Both because of the tendency to select comparison groups of superior ability and because of a related tendency for measures to be more reliable for superior groups, a number of prominent evaluations have concluded that program effects did not exist when such effects probably did, e.g., Head Start and performance contracting.

HOW TO DO IT

Dealing with Administrators and Staff

Since a favorable ethical and practical atmosphere for a rigorous evaluation exists when there are more eligible applicants than the program can handle, utilization information is very important. While administrators may be receptive to increased advertising of the program, receptivity will be aided if underserved populations can be identified. This information will serve the additional function of directing advertising and outreach efforts.

While a no-program comparison group has many advantages, it is not always the only or best comparison group. An attractive addition or alternative is a minimal program group, i.e., one receiving the low-cost purely informational components of the full program. The use of this group allows the staff to feel they are reaching more people, rather than withholding a needed program from them, but, at the same time, is not a (ethically dubious) placebo group which deceives participants by creating the impression of helping but not actually providing assistance. Combined with the no-program group, the use of a minimal program group allows administrators to gauge the overall effect of the full program as well as the benefit of its expensive components. McKillip (1978) discusses the advantages and disadvantages of variations on the minimal program. A final point is that the use of a no-program group makes clear to the staff that it is the program and not the staff which is being evaluated.

Three other strategies facilitated interactions with administrators and staff. First, action and evaluation units were kept separate, but maintained close contact. Secondly, inclusion of description of the lottery (randomization) for selection of program participants was included on the signed informed consent forms. This procedures eased administrative fears of retribution from those assigned to no-program groups. And thirdly, a number of program slots were kept open for those applicants whom the staff felt were particularly needy or deserving. Data from these participants was excluded from analyses of outcome measures.

Dealing with Program Participants

Perhaps because we work within a university setting we have encountered little resistance to the use of randomization from potential program participants (although considerable resistance came from program administrators and staff). We found the following procedures useful, however.

First, the term randomization is never used. Rather, program participants are selected by lottery when there are more applicants than available positions. We have found the concept of a lottery to be acceptable and widely understood. Perhaps this effect would be limited to states which have lotteries..

Secondly, we have informed potential participants at the time of application about the lottery and about the minimal or no-program groups. Including this information on the informed consent form has minimized resentment and the potential political hassle caused by randomization out of an attractive program (Wortman, Hendricks, & Hellis, 1976).

Third, as much as possible, we have dealt with applicants personally, especially those assigned to no-program groups. Those refused entry were informed by a phone call which included discussion of the rationale for the lottery. This procedure allows us to answer questions and clear up misunderstandings as soon as possible. Fourth, when possible, the promise of subsequent services to the no-program group may make program staff and group participants more comfortable.

By using this four pronged attack, i.e, lottery, fully informed consent, personal contact, and promise of subsequent services, we have been able to keep attrition from the no-program groups at the same level as that for the program groups.

REFERENCES

Bernstein, I.N., & Freeman, H.E. Academic and entrepeneurial research: The consequences of diversity in federal evaluation studies. New York: Russell Sage Foundation, 1975.

Boruch, R.F. On common contentions about randomized field experiments. In (ed.) G.V. Glass, Evaluation studies review annual, volume 1, 1976. Beverley Hills, CA: Sage Publications, 1976, 158-194.

Chu, T.D., & Trotter, S. The madness establishment: Ralph Nader's study group report on the NIMH, New York: Grossman Publishers, 1974.

Cronbach, L.J. Remarks to the new society. Evaluation Research Society Newsletter, 1977, 1, 1-3.

Garrard, J. Viatkus, A., & Chilgren, R. Evaluation of a course in human sexuality. Journal of Medical Education, 1972, 47, 772-778.

Gilbert, J.P., Light, R.J., & Mosteller, F. Assessing social innovations: An empirical basis for policy. In (eds.) G.A. Bennett & A.A. Lumsdaine, Evaluation and Experiment. New York: Academic Press, 1975, 39-193.

McKillip, J. Three flexible designs for impact evaluation of service programs. Mimeo available from the Department of Psychology, Southern Illinois University, Carbondale, IL 62901

McKillip, J., & Kamens, L. Evaluation of a positive health care program. Proceedings of the Second Annual Illinois Health Care Research Symposium. Springfield, IL, 1977.

Meier, P. The biggest public health experiment ever: The 1954 field trial of the Salk poliomyelitis vaccine. In (eds.) J.M. Turner, et al., Statistics: A guide to the unknown. San Francisco: Holden-Day, 1972.

Oetting, E.R. Evaluation research and orthodox science: Part I. Personnel and Guidance Journal, 1976, 55, 11-15.

Rossi, P.H. Testing for success and failure in social action. In (eds.) P H. Rossi & W. Williams, Evaluating social Programs. New York: Seminar Press, 1972, 11-49.

Voss, J.R. Program evaluation in sex education: The effectiveness of sexual awareness weekend workshops. Ph.D. dissertation, Southern Illinois University at Carbondale, 1977.

Wortman, C.B., Hendricks, M. & Hellis, J.W. Factors affecting participants reaction to random assignment in ameliorative social programs. Journal of Personality and Social Psychology, 1976, 33, 256-266.

Zuckerman, R., Tushup, R., & Finner, S. Sexual attitudes and experience: Attitude and personality correlates and changes produced by a course in sexuality. Journal of Consulting and Clinical Psychology, 1976, 44, 7-19.