Research Brief

The differential attrition of persons from comparison groups severely restricts the inferences that can be made from results of evaluative research. This problem is particularly troublesome in the evaluation of medical technologies, such as coronary artery bypass graft surgery, since a substantial percentage of medical or control patients cross over to the surgical group. A procedure using worst case assumptions is developed that allows researchers to estimate the maximum effect of differential attrition, and therefore enhance the quality of their inferences. The article first illustrates the procedure, then concludes with a discussion of the generality of the estimation procedure to other instances in which differential attrition is a problem, and points out the limitations of the approach.

DIFFERENTIAL ATTRITION

Estimating the Effect of Crossovers on the Evaluation of a Medical Technology

WILLIAM H. YEATON
Institute for Social Research
PAUL M. WORTMAN
School of Public Health and Institute for Social Research
The University of Michigan at Ann Arbor
NAFTALI LANGBERG
University of Haifa

Among the various ways that research designs are comprised, perhaps the most troublesome is the differential attrition of subjects from comparison groups during the actual conduct of an evaluation or experiment (Cook and Campbell, 1979). For example, Boeckmann's reanalysis (1981) of the New Jersey negative income tax

AUTHOR'S NOTE: *The work on this report was supported by a grant from the National Center for Health Services Research (HS-04848-01). The authors wish to acknowledge the helpful comments of Dr. Charles Reichardt and two anonymous reviewers on earlier*

EVALUATION REVIEW, Vol. 7 No. 6, December 1983 831-840

experiment (Watts and Rees, 1977) suggests that differential attrition from the experimental and control groups of both black and Spanish-speaking minorities could account for differences otherwise attributable to the intervention. Likewise, Wortman (1978) has argued that the process of differential attrition was a plausible explanation of the negative effects found in McCord's (1978) 30-year follow-up of the relationship of counseling to subsequent delinquency in a randomized controlled trial.

As a result of this differential attrition process, a well-conceived randomized experimental design may drift toward a quasi-experimental design model with all of its inferential limitations (see, for example, Special Report, 1982). Despite the creative application of statistical procedures to adjust for the resulting nonequivalence between groups (Kenny, 1975; Magidson, 1977; Reichardt, 1979), there is no satisfactory statistical solution to the lack of adherence to the original design protocol.

The most common recommendation of methodologists is to analyze the data from randomized experiments according to the original assignment (Riecken and Boruch, 1974) or "intention to treat" (Peto et al., 1976). This approach is a tradeoff that preserves the design at the expense of a biased estimate of the treatment effect.

THE PROBLEM OF CROSSOVERS

In the assessment of medical technologies, researchers will often confront situations in which specific techniques are preferred by patients due to their association with secondary outcomes that are intrinsically desirable. In instances such as these, it will be particularly difficult to maintain the design protocol. This precise situation confronts researchers interested in the evaluation of the potential benefits of coronary artery bypass graft surgery (CABGS) for patients with coronary heart disease (Wortman, 1981).

Given the consistent finding that angina is relieved in patients receiving surgery (Special Report, 1981), it is ethically problematic to withhold a potential benefit from the medical group. Accordingly, any efforts to compare survival rate between patients who are operated on and those adhering to a medical regimen are greatly complicated by the

---

versions of this paper. Requests for reprints should be sent to Dr. William H. Yeaton, Center for Research on the Utilization of Scientific Knowledge, University of Michigan, Box 1248, Ann Arbor, Michigan, 48106.

fact that substantial percentages of medical patients cross over to the group of patients who receive surgery. Furthermore, the migration is unidirectional since it is not possible for patients to cross over from the surgical to the medical groups once they have received surgery.

In a synthesis of results from 25 controlled trials of CABGS (Wortman and Yeaton, 1983), the crossover rates from the medical to the surgical group were found to be quite sizeable, ranging up to 45.0% in randomized controlled trials, with a mean rate of 21%. Compounding this problem is the systematic nature of the attrition. It is typically those medical patients with the worst prognosis, that is, those with the most severe angina and imminent danger of heart attacks who become crossovers (Murphy, Hultgren, Detre, Thomsen, and Takaro, 1977: 1470). As these researchers have noted:

> Medical nonadherers are frequently assumed to be treatment failures. Although this result was true in approximately 54 per cent of our patients (unstable angina, 19 per cent, and progressive angina, 35 per cent), patient or physician preference prevailed in some cases.

ESTIMATING THE EFFECT OF CROSSOVERS

Any inferences that are made from controlled trials of CABGS must weigh the extent to which patients in the medical group have crossed over to receive surgery. Clearly, the effect of the loss of the most ill patients from the medical group is to raise the survival rate in the medical group. Whether crossovers are dropped from the study or included, as most evaluation methodologists recommend, the extent of the increase is not readily apparent (Wortman, 1981).

In fact, the most common biostatistical research practice is to consider crossovers as an endpoint, that is, as no longer in the study, at the time they receive surgery. This would bring the medical group survival rate closer to that for the surgical group, if one assumes that surgery is beneficial, an assumption consistent with the data. Consequently, the crossover problem will underestimate any potential benefit due to surgery. If one follows the recommendation to include crossovers in their originally assigned group, then the effectiveness of surgery will also be underestimated (again assuming it is beneficial). Neither method then can overcome the effects of differential attrition and treatment diffusion to produce an unbiased estimate of effect.

The worst case approach uses a general strategy of determining the maximum degree of influence attributable to a particular factor and

thus the factor's likelihood of contributing to the difference obtained. In this way it resembles the sensitivity analysis used by economists in cost-effectiveness and cost-benefit analyses (Weinstein and Stason, 1982) to ascertain the consistency of conclusions under various conditions such as extreme and intermediate values of the discount rate in determining present costs. If conclusions are preserved under worst case assumptions, one can place considerable confidence in the validity of the results. If conclusions are contingent upon the values assumed, one can judge "up front" the plausibility of the value that is needed to maintain consistent conclusions.

Though the very nature of crossovers makes it impossible to determine exactly the effect of such attrition on survival rate, it is possible to estimate the maximum influence that crossovers would have on the mean survival rate of the medical group. By calculating this maximum influence attributable to crossovers, researchers would be in a greatly improved position for defending inferences about differences between medical and surgical groups. Since this procedure is likely to overadjust for attrition, it could provide convergent evidence if it agrees with the more traditional estimates, in their direction if not their magnitude.

This estimation procedure necessitates some very specific assumptions, however. We will assume that only those patients in the medical group with the worst prognosis cross over to receive surgery, and that these patients are considered as an endpoint at the time they cross over. With regard to the distribution of survival rates of medical patients, this assumption implies that the tail of the distribution is truncated at precisely the point that will eliminate the exact percentage of patients who cross over. This means that the area under the distribution curve that is eliminated will coincide with the percentage of patients who cross over. We further assume a distribution of composite measures of health status that reflects the probability of survival for medical patients after the time patients in the surgical group have received CABGS. We also hypothesize that the measure is standardized normal (so that the mean equals zero and standard deviation equals one), allowing us to use standard formulae to calculate the mean of the truncated distribution. Though truncated, standard normal distributions are also employed by economists (Stromsdorfer and Farkas, 1980), they are commonly used to correct statistically for attrition bias in ANOVA and regression models (Hansman and Wise, 1979) rather than to form a basis for a worst case solution.

Formally, if f is a density function defined by

$$f(x) = \frac{e^{-x^2/2}}{SQ\ RT(2\pi)}.$$

where x assumes any real number value, then for any given percentage p of crossovers, the area under the normal curve yielding this percentage p can be found by integrating the normal curve density function from minus infinity to that point z on the x-axis which yields p as the result of the integration:

$$\int_{-\infty}^{z} \frac{e^{-x^2/2}}{SQ\ RT(2\pi)} = p.$$

The mean of interest (the mean of the truncated distribution) will be:

$$\left(\frac{1}{1-p}\right)\int_{z}^{\infty} \frac{xe^{-x^2/2}}{SQ\ RT(2\pi)} = \left(\frac{1}{1-p}\right)\left[\frac{-e^{-x^2/2}}{SQ\ RT(2\pi)}\right]_{z}^{\infty}$$

$$= \frac{e^{-z^2/2}}{(1-p)\ SQ\ RT(2\pi)}.$$

Given various crossover rates p, one can use a table of standard normal deviates to determine the corresponding z-value on the abscissa. These two constants can then be substituted into the above result for the mean of the truncated distribution to ascertain the magnitude of shift of the mean.

Examples

For purposes of illustration, several p values and the corresponding means of interest are displayed below. When plotted, the relationship appears to be essentially linear.

p = .01	mean = .03
p = .05	mean = .11
p = .10	mean = .19
p = .15	mean = .27
p = .20	mean = .35
p = .21	mean = .36
p = .25	mean = .42

Thus, the crossover rate of 21% (when p = .21) found in the authors' synthesis of the results of controlled trials of CABGS (Wortman and Yeaton, 1983 would be associated with a mean shift of .36 (a 36% increase in the standardized mean value), the maximum change attributable solely to crossovers.

In some instances it will not be necessary to translate shifts calculated in standardized units to their equivalents in unstandardized terms. Measures of effect size (Glass, McGaw, and Smith, 1981) are calculated by dividing mean differences by an appropriate standard deviation and thus are directly comparable to results generated from mean shifts in the standard normal distribution. For example, given an effect size of .50 and a crossover rate of 21%, one can determine the maximum effect of crossover by simply adding .36, the mean shift, to .50, to adjust for the underestimated outcome measure. By comparing this adjusted value to the original value, one can estimate the extent to which a difference between groups is likely to be due to crossovers.

In other cases there will be no immediately obvious standard deviation value by which one can standardize results, but reasonable estimates may be available. For example, in the medical research cited above, survival rates were assumed to be reflected in the distribution of composite measures of health status, and these measures might be used to produce a standard deviation. Another measure of variability might be the standard deviation of the survival means of other similar studies. For example, given a medical group mean survival percent of 65 and a standard deviation of 10 found from a composite index of health status, the maximum effect of a 21% crossover rate would be 3.6 (.36 times 10). Therefore, the survival rate would be increased to 68.6 as a result of crossovers.

In practice, means and standard deviations are available after crossovers have occurred, and researchers will be interested in determining the adjusted mean before the effect of crossovers. In this case, one simply subtracts the product of the percent shift and the standard

deviation from the given mean. As an illustration, again assume a medical group mean survival percent of 65 and a standard deviation of 10 percent, values determined after crossovers. The adjusted mean would be 61.4 (65 −.36(10)), thus allowing the researcher to conclude that the difference between medical and surgical groups would be underestimated by a maximum of 3.6 percentage points as a result of crossovers, assuming that the mean in the surgical group is greater than the mean in the medical group. While in the case of attrition due to crossovers it is obvious that an adjustment must be made in the control group measure, the practice is consistent with the identification of distortion in research that uses historical (Sachs, Chalmers, and Smith, 1982) and other nonrandomized controls (Meier, 1978).

GENERAL COMMENTS

These findings suggest that high crossover rates can substantially increase the mean of the distribution of the control group of medical patients in which crossovers have been eliminated. Consequently, the benefit attributable to surgery would be substantially underestimated in controlled trials of CABGS. While from a statistical point of view mean shifts between 20% and 50% would be considered between small and medium effect sizes (Cohen, 1977), innovative surgeries typically produce modest benefits (Gilbert, McPeek, and Mosteller, 1977) that assume importance through their implementation with large groups of patients. For example, the evidence from randomized controlled trials suggests a benefit of CABGS of less than 5% (Wortman and Yeaton, 1983). Fortunately, the ability to detect these modest benefits is enhanced considerably by the above estimation technique, since the degree to which crossovers may alter a group mean and thus underestimate differences between groups can be determined easily.

Of course, the relationship between the rate of crossovers and the shift in mean survival rate found in actual reports of CABGS will not follow the idealized relationship described above. Distributions may only approximate the normal, and variances will change as a function of the range of diagnostic severity of patients in the medical group. To the degree that the distribution is negatively skewed or the variance is large, the shift in the mean will increase. Also, not all medical patients will cross over at the same point in time, as we have assumed in our calculations.

When there is a high incidence of crossovers early in the follow-up period of a controlled trial, the degree of bias attributable to crossovers will be maximal. The longer the delay period before patients begin to cross over from the medical group, the closer one approaches the case of an intact control group. Furthermore, we have assumed that only the worst medical patients cross over, an assumption not likely to be true in actual practice. However, the closer the mix of crossovers approximates the case in which only the worst medical patients cross over, the closer the mean shift will approximate the maximum shift shown in this report.

The problem of crossovers in the assessment of effectiveness of CABGS is illustrative of the differential attrition process that plagues evaluation research. The "solution" presented here is applicable to those instances in which the differential attrition process selects subjects or patients in the same manner as they were selected in this report. Specifically, if a differentially attrited subgroup of persons is homogeneous on some measure(s) of status (such as health in this report, occupation in the McCord (1978) study, and ethnicity in the Watts and Rees (1977) volume) that correlates with the outcome variable in question, then the findings of this report are relevant. Obviously, the degree of direct relevance will depend on the match of the groups resulting from the differential attrition process to the pertinent assumptions upon which our estimates are based: attrition of only worst case persons from one of the comparison groups and the shape of the relevant distributions. Other potentially important factors such as the strength of correlation between the status and the outcome measures may compensate for departures from worst case assumptions, however.

While the emphasis of this report has been on the accurate interpretation of research results in studies plagued by differential attrition, the findings may also be used in planning studies. Briefly, if one is armed with knowledge from past studies with regard to the expected rate of crossovers, precise estimates can be made of the degree to which the magnitude of difference between groups is likely to be altered. Accordingly, sample sizes can be either increased or decreased to reflect smaller or larger differences between groups, thus enhancing the power of experiments or diminishing their expected costs.

Despite the shortcomings associated with idealized data, the relationship between crossover rates and survival presented in this report will allow researchers to estimate more accurately the potential influence of crossovers, and thus to improve the quality of their inferences. Given the uncertainties in interpreting the results from flawed research studies, it is

important that investigators acknowledge the potential bias caused by such "threats to validity" (Campbell and Stanley, 1966). These problems are much more common in the applied field studies characteristic of program evaluation and medical technology assessment. Worst case assumptions can provide a bound for a treatment's impact and indicate the extent to which the estimate of effect is sensitive to bias.

REFERENCES

BOECKMANN, M. E. (1981) "Rethinking the results of a negative income tax experiment," pp. 341-355 in R. F. Boruch, P. M. Wortman, D. S. Cordray, and Associates (eds.) Reanalyzing program evaluations: Policies and Practices for Secondary Analysis of Social and Educational Programs. San Francisco: Jossey-Bass.

CAMPBELL, D. T. and J. C. STANLEY (1966) Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally.

COHEN, J. (1977) Statistical Power Analysis for the Behavioral Sciences (Rev. Ed.) New York: Academic Press.

COOK, T. D. and D. T. CAMPBELL (1979) Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

GILBERT, J. P., B. McPEEK, and F. MOSTELLER (1977) "Progress in surgery and anesthesia: benefits and risks of innovative therapy," in J. P. Bunker, B. A. Barnes, & F. Mosteller (eds.) Costs, Risks, and Benefits of Surgery. New York: Oxford Univ. Press.

GLASS, G. V, B. McGAW, and M. L. SMITH (1981) Meta-Analysis in Social Research. Beverly Hills, CA: Sage.

HAUSMAN, J. A. and D. A. WISE (1979) Attrition bias in experimental and panel data: the Gary income maintenance experiment. Econometrica 47: 455-473.

KENNY, D. A. (1975) "A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design." Psych. Bulletin 82: 345-362.

MAGIDSON, J. (1977) "Towards a causal model approach for adjusting for preexisting differences in the nonequivalent control group situation: a general alternative to ANCOVA." Evaluation Q. 1: 399-420.

MEIER, P. (1978) "The biggest public health experiment ever: the 1954 field trial of the Salk poliomyelitis vaccine." pp. 3-15 in J. M. Tanur, et al. Statistics: A Guide to the Unknown. Berkeley: Holden-Day.

McCORD, J. (1978) "A thirty-year follow-up of treatment effects." American Psychologist 33: 284-289.

MURPHY, M. L., H. N. HULTGREN, K. DETRE, J. THOMSEN, and T. TAKARO (1977) "Special correspondence: a debate on coronary bypass." New England J. of Medicine 297: 1470.

PETO, R., M. C. PIKE, P. ARMITAGE, N. E. BRESLOW, D. R. COX, S. V. HOWARD, N. MANTEL, K. McPHERSON, J. PETO, and P. G. SMITH (1976) "Design and analysis of randomized clinical trials requiring prolonged observation of each patient: introduction and design." British J. of Cancer 34: 585-612.

REICHARDT, C. S. (1979) "The statistical analysis of data from nonequivalent group designs," in T. D. Cook and D. T. Campbell (eds.) Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally College Publishing Company.

RIECKEN, H. W. and R. F. BORUCH (1974) Social Experimentation: A Method for Planning and Evaluating Social Intervention. New York: Academic Press.

SACHS, H., T. C. CHAIMERS, and H. SMITH, JR. (1982) "Randomized versus historical controls for clinical trials." The American J. of Medicine 72: 233-240.

Special Report (1981) "National Institutes of Health Consensus Development Conference Statement." New England J. of Medicine 304: 680-684.

Special Report (1982) "The anturane reinfarction trials: reevaluation of outcome." New England J. of Medicine 306: 1005-1008.

STROMSDORFER, E. W. and G. FARKAS (1980) "Methodology," pp. 32-41 in E. W. Stromsdorfer and G. Farkas (eds.), Evaluation Studies Review Annual, Vol. 5. Beverly Hills, CA: Sage.

WATTS, H. and A. REES (1977) "The New Jersey Income Maintenance Experiment," Vol. 2. New York: Academic Press.

WEINSTEIN, M. C. and W. B. STASON (1982) "Cost-effectiveness of coronary artery bypass surgery." Circulation 66: III-56-III-66.

WORTMAN, P. M. (1978) "Differential attrition: Another hazard of follow-up research." American Psychologist 33: 1145-1146.

WORTMAN, P. M. (1981) "Randomized clinical trials," pp. 41-60 in P. M. Wortman (ed.), Methods for evaluating health services. Beverly Hills, CA: Sage.

WORTMAN, P. M. and W. H. YEATON (1983) "Synthesis of results in controlled trials of coronary artery bypass graft surgery," in R. J. Light (ed.), Evaluation Studies Review Annual, Vol. 8. Beverly Hills, CA: Sage.

William Yeaton is an Assistant Research Scientist at the Center for Research on Utilization of Scientific Knowledge, University of Michigan. His research interests include evaluation research methodology and the development and assessment of techniques to teach people to be more critical consumers of research findings.

Paul M. Wortman is a Professor in the Department of Medical Care Organization, School of Public Health and Program Director in the Center for Research on Utilization of Scientific Knowledge at the Institute for Social Research, both at the University of Michigan, Ann Arbor. His research interests focus on evaluation research methods and medical technology assessment.

Naftali Langberg is currently Chairman of the Department of Statistics at the University of Haifa in Israel. His current interests include probability theory and applied problems in mathematical statistics.