The Journal of Compliance in Health Care, Vol. 2, No. 1, 1987

Dropout in Patient Compliance Studies: A Suggested Analytic Procedure

David J. Weiss, Ph.D.

Department of Psychology, California State University, Los Angeles

In studies in which patient compliance is experimentally manipulated, differential attrition rates across conditions can bias the results. Standard statistical techniques for unequal cell sizes assume dropouts have occurred randomly. To the extent that certain treatments are more likely to prompt withdrawal, however, the usual analysis of the degree of compliance of those who remain will be incorrect. The zero-implantation method proposed here provides scores for dropouts that may be integrated with obtained compliance scores. The basis of the method is the assumption that dropping out of treatment represents the ultimate in noncompliant behavior; dropouts receive compliance scores of zero. When that assumption is justified, this simple method provides sensible comparisons of the treatments.

The situation to be considered here is an experimental manipulation of patient compliance in which participants are measured repeatedly. There are two or more treatment groups, and the treatments are expected to affect compliance differentially. The degree of compliance is measured periodically with a numerical dependent variable, such as the number of prescribed pills properly taken each month. Although this kind of study has a longitudinal character because of the periodic measurements, the researcher's emphasis is usually not on progress over time of compliance but rather on the overall performance of each participant as it reflects on the experimental condition to which he or she was randomly assigned. At the end of the study, a straightforward

analytic procedure (an analysis of variance on the monthly scores in which participants are nested under groups and crossed with months, for example) will provide statistical verification of the efficacy of the treatment variable.

When participants drop out of a study, though, a simple matter becomes complex. The situation is not too bad if dropouts occur at random across treatments. The scores that are present may be analyzed with standard techniques for handling missing data in repeated-measures designs (Federer, 1955; Yates, 1933). Such techniques basically predict the missing scores from those which are present; the major problem for the experimenter, aside from the minor problem that the data analysis is somewhat harder to carry out, is that the power of the experiment is reduced. The dropouts make it less likely that the experiment will detect true differences attributable to the treatment variable (Halperin, Rogot, Gurian, & Ederer, 1968; Lasky, 1962).

In a typical compliance study, however, dropouts may not occur randomly. The treatment variable may differentially affect not only the dependent variable, the pill-taking, but also the participant's willingness to remain in treatment. An odious treatment may drive people away while not affecting the pill-taking of those who remain: dropping out is the participant's way of letting the researcher know that the treatment is not effective. Indeed, in many situations dropout may be viewed as the ultimate noncompliant behavior. As such, it should be incorporated into the data analysis.

There are, of course, cases in which this view of dropout does not apply. Patients may withdraw for personal reasons unrelated to the treatment or because the assessment procedure (rather than the treatment) inspires noncooperation. Such factors would not be expected to affect the treatments differentially. In other situations dropping out may not be the most extreme of noncompliant behaviors. The patient may overcomply, that is, take too many pills or take pills in the presence of dangerous symptoms, and the practitioner might consider this overcompliance as medically worse than taking no pills. Overcompliance presents a difficult scoring problem for the researcher because the obvious index of compliance, the number of pills taken, is not ordinally related to the extent of medically appropriate behavior. One may build into the scoring system a correction for pills improperly taken, with the severity of the penalty tied to the medical risk (Lange, Ulmer, & Weiss, 1986). This adjustment avoids the absurdity that a patient who is in the study but is severely overcomplying may score worse than a dropout. If this theoretical possibility occurs in practice, then the analysis to be proposed here would not apply. The proposal is designed only for studies in which no observed degree of compliance is deemed worse than dropping out.

There is a long-standing tradition of inquiry into the causes of dropout (Baekeland & Lundwall, 1975; Beck, Shekim, Fraps, Borgmeyer, & Witt, 1983; Caldwell, Cobb, Dowling, & de Jongh, 1970; Finnerty, Mattie, & Finnerty, 1973: Oldridge, 1984). The customary goal of such research is to determine predictors of dropout. Then the researcher can eliminate from future compliance studies persons who are unlikely to finish the experiment. This perspective is sensible enough if one regards the dropout as an inconvenience to the researcher, or if only one treatment condition is to be used in a study. But it is not sensible if the goal is to understand how the several treatments affect the ordinary patients to whom they will be applied after the study is over. In order to generalize, study participants ought to be like those ordinary patients, and thus selecting on the basis of being likely to finish the study will yield results of limited applicability.

The extent of dropout in a compliance study can be dramatic. In a recent investigation in which patients with inactive tuberculosis were scheduled for a year of daily pill-taking (Lange, Ulmer, & Weiss, 1986), more than half the patients dropped out soon after treatment began. Such a high dropout rate is not unusual in studies of chronic disease (Hecht, 1974). It is the possibility of such large-scale attrition that necessitates provision for dropouts in the design of a compliance study.

Many researchers have employed dropout rate as a dependent variable (Bigger, 1976; Hagan, Foreyt, & Durham, 1976; Krebs, 1971; Linn, Shane, Webb, & Pratt, 1979; Lange, Ulmer, & Weiss, 1986). It is common to find differences in dropout rates across treatments. The question is how to integrate the dropouts with the periodic scores that are available. It is possible for an analysis based on dropout rate to agree or to disagree, in terms of the effectiveness of the treatments, with one based on a conventional compliance measure. If both analyses in a study point to one treatment's superiority, then there is no great problem for the researcher (although specifying the extent of the superiority may be a problem for the statistician). But if there is disagreement between the two measures, what advice can the researcher provide the practitioner about which treatment to use?

A PROPOSED SOLUTION

Instead of regarding dropout and pill-taking as two aspects of compliance, it is proposed to unify the elements by taking seriously the idea that dropping out is the ultimate noncompliant behavior. As such, the participant who drops out is viewed as complying not at all. In other words, the dropout complies to the extent of zero for the duration of the

experiment. The data analysis should allow the dropping out to reflect adversely on the condition that generated it, and the earlier the event occurs, the less effective the treatment should appear. From the perspective of the analyst, dropping out is a final event; later resumption or possible treatment elsewhere is ignored.

These assumptions about dropout (and it should be stressed that they are assumptions) imply a simple analytic procedure. The periodic scores for all participants are entered into a standard analysis, such as the repeated-measures analysis of variance; and when a participant drops out, all of his or her subsequent scores are entered into the analysis as zeroes. For the dropout, a zero is entered for every designated measurement interval until the experiment is concluded or until all of the participants have dropped out.

The zero-implantation scheme is simple to implement and its consequences seem sensible. Comparison of the means across groups will be affected. The fewer actual scores there are for a participant, the more adverse the impact on the treatment from which the dropout occurred. The zeroes will ensure a low mean compliance score for the dropout; and the earlier the withdrawal, the lower the mean will be. Statistical power, the ability of the experiment to detect a true difference associated with the treatment, will also be affected. The analysis appropriately has difficulty confirming a treatment effect when there are few actual measurements. Because the means for individuals who have dropped out will be low relative to those who complete the treatment, the variability between participants will be high when there is considerable attrition. Since this variability is a yardstick against which the treatment effect is measured (it is the core of the between-subjects error term in the analysis of variance), high variability will make it difficult to show that the treatments affect compliance differently.

In general, temporal effects are likely to be of lesser importance to the researcher than is intertreatment comparison. This is fortunate, because when there are many dropouts, such information is not likely to be available. The researcher might be interested in subtle, systematic changes in compliance over the course of the study. However, such patterns among actual scores will be obscured by the gross changes from typical scores to zeroes as dropouts occur. A temporal effect is almost certain to be observed; as more patients drop out, the average compliance over time does indeed decrease. The zero-implantation method accurately detects this decrease, but it is not of substantive interest. It is not surprising that when participants are in a study for different lengths of time, temporal information is difficult to extract from the data (Weiss, 1985).

The difficulty with temporal information also has a technical counterpart. This information is captured by the analysis of variance in

two sources: "time periods" and the interaction "time periods by groups." The statistical power of the tests of these sources is low because of the dropouts. The implanted zeroes are not independent; obviously, they lack the variability of real data. Because these scores are not free to vary as ordinary scores do, it is necessary to apply the usual statistical adjustment for estimated data. This adjustment consists of removing 1 *df* for each estimated score (Yates, 1933). In the present case the appropriate removal is from the *df* for the error term used to test both temporal sources (the within-subjects error). Consequently, if many zeroes have been implanted then the mean square for that error term, having been produced by dividing the sum of squares by a small *df*, will be large. This means low power for tests involving that error; but as has been observed, such tests are not likely to be informative in any event when dropouts are prevalent. From a practical perspective, then, the *df* price for zero implantation is minimal. The crucial error term, the one against which the treatment effect is tested, is not affected. The *df* for the between-subjects error term depends only on the number of subjects and treatments, not on the number of scores per subject.

If there are many early dropouts in all of the treatments, the means across conditions will be similar and the analysis will appropriately favor the null hypothesis. So long as one is willing to accept the cornerstone of the zero-implantation method, that dropping out is the ultimate in noncompliance, then each participant's summed scores reflect in an unbiased way his or her overall response to the treatment. The researcher need not attempt to ascertain the reasons for individuals' dropping out (although these reasons may be important in their own right for the design of future investigations) to decide how to process the data. The procedure is objective, and its applicability depends on a theoretically based, a priori decision by the researcher.

ILLUSTRATION

The tuberculosis study cited previously (Lange, Ulmer, & Weiss, 1986) may furnish an example of the zero-implantation methodology. Patients, who were low-income Hispanics treated at a community health center, were to take daily pill dosages for a year, returning to the clinic with their bottles once a month. They were randomly assigned, either to a group that received a medication calendar or to a control group that did not. The compliance scores were based primarily on the percentage of pills taken as directed. There were correction factors in the scoring algorithm, which reduced the score for taking excess pills and for missing appointments, but in practice little adjustment was needed. No monthly

composite score was ever below 32%, and most scores were much higher.

This experiment has the appropriate character for the zero-implantation scheme to be applicable. There were many dropouts. Because no observed score was negative, the basic assumption of the method was satisified, and one could justifiably regard dropping out as the ultimate act of noncompliance. However, the study was cut short, and the dropout rates were not appreciably different in the two experimental conditions. Accordingly, for illustrative purposes, simulated data for a six-month version is given in Table 1. Scores represent compliance on a monthly basis.

The mean of the twenty-eight Group I scores is 72.68, while the mean of the twelve Group II scores is 83.5. If instead of using group means one treats each patient as a unit and looks at the mean of the means, a similar result obtains. The mean of the means for Group I is 72.7; the mean of the means for Group II is 83.9. An analytic method that looks only at those scores that are present will conclude that there was greater compliance in Group II.

The proposed method, however, which implants the zeroes in a repeated-measures design with patients nested under groups, yields a radically different conclusion. The mean score for Group I is 67.8, and the mean score for Group II is 33.4. Statistical verification is given by the analysis of variance shown in Table 2. The key result is the F ratio for groups, 8.65, significant at the .05 level. This analysis shows that in spite of a higher mean monthly compliance for Group II in the raw scores, there is greater compliance in Group I. The superiority from a medical standpoint is obvious, as many more pills are being properly taken by Group I patients.

The ANOVA was carried out using a standard computer program for nested designs. Zeroes were input in place of the X's given in Table 1. Only one modification of the output was necessary. The degrees of freedom for the within-subjects error term were reduced from the program's 40 to 20; 1 df is lost for each implanted zero. The corrected total df, 39, is what one would expect with the 40 observed compliance scores.

A LIMITATION

For the zero-implantation to be meaningful, the scores used in the analysis should have the property that zero represents the least possible compliance. Technically, the number zero must be a true zero on the scale used for the dependent variable. This limitation rules out compliance indices that may be either positive or negative, such as number of

TABLE 1. Simulated Data

	Treatment Group I							Treatment Group II						
	Month							Month						
	1	2	3	4	5	6		1	2	3	4	5	6
Patient 1	83	72	91	91	76	83	Patient 6	87	83	86	77	81	X
Patient 2	57	84	52	63	74	82	Patient 7	81	X	X	X	X	X
Patient 3	96	80	33	52	68	66	Patient 8	78	87	84	X	X	X
Patient 4	85	91	73	42	X	X	Patient 9	87	X	X	X	X	X
Patient 5	87	59	65	72	83	75	Patient 10	82	89	X	X	X	X

X = dropout

TABLE 2. Analysis of Variance

Source	df	SS	MS	F
Groups	1	17784.8	17784.8	8.65*
Error (between)	8	16444.8	2055.6	
Months	5	18755.1	3751.0	2.94*
Months × Groups	5	6015.1	1203.0	<1
Error (within)	20	25544.4	1277.2	
	39			

* = significant at .05 level

pounds lost or decrease in blood pressure. Were one to apply the zero-implantation in such cases, it would be possible for a dropout to receive a higher periodic compliance score than a current participant. Many measures are acceptable: number of pills taken, minutes of exercise, sessions attended, assignments completed, and the like. In general, compliance indices that simply record the occurrence of a directed behavior have the desired property. When compliance is measured as an indirect, perhaps complex, function of the instructions (as in the case of exercise inferred from weight loss), careful choice of an index is required, and it may not prove possible to find a practical measure whose zero is a true minimum.

DISCUSSION

The justification for the proposed analysis, the zero-implantation scheme, is that it is plausible and practical. Is it not reasonable to assume that a dropout takes no pills? In the usual clinical setting, the monthly allotment of pills is not even available unless the patient shows

up, so zero is an accurate estimate of the number of pills taken by a dropout.

Plausibility is not an especially strong argument for a method of analysis. One disturbing possibility is that the dropout might somehow get medications from another source external to the study. Ideally one might employ a predictive, or criterion-related, validation scheme in which a participant's scores using the proposed analysis are shown to have high correlation with an established criterion. Unfortunately, treatment outcome is unlikely to be a one-to-one function of compliance in a therapeutic setting; thus no criterion is available. For this reason, even a countable index, such as the number of pills not in the bottle, might be challenged as to its validity as a measure of compliance (Gordis, 1979; Porter, 1969); conceivably, the patient could simply throw pills away. The approach proposed here presumes the validity of the pill count as a measure and attempts to refine the measure to cover a special situation. Content, or face, validity, that is, plausibility, is less objective. The reasonableness of the approach is stressed. One argues that the method responds sensibly to possible data outcomes and that the statistical machinations are consistent with standard practice. As researchers employ the method in their empirical investigations, a consensus may develop that satisfactory results are achieved and the method may enter the mainstream.

A practical implication of the key idea here, that dropouts may not be, and indeed are unlikely to be, random occurrences unrelated to the treatments, is that the customary experimental practice of replacing the dropout with a new participant is an error. Suppose that one of the treatments is particularly arduous, so that only the hardy, or especially compliant, participant survives it to the end. Then replacing the dropout, repeatedly if necessary, biases the results by concentrating those hardy subjects in that arduous condition. The very participant who dropped out of the difficult condition might well have finished the experiment had the random assignment worked out differently. If a researcher does not wish to use the zero-implantation scheme but instead insists on replacing dropped-out subjects, it is necessary to replace in cohort sets. Participants should be assigned to treatment groups in sets, one to each group, using a random permutation to determine the assignment. Then if any dropout occurs, the entire set should be replaced to avoid bias. Replacing only the subjects who drop out is defensible only if the attrition can be assumed to be unrelated to the treatments; and it is not easy to determine if there is a pattern in the withdrawals while the study is ongoing.

In this presentation, two closely related situations involving dropouts have not been considered. These situations encompass studies in which only one measurement per participant is planned. The first case is

one in which compliance is defined as remaining in the study for a designated period of time. In that case, no information is available other than the fact of early withdrawal or the date, if that is known. If length of time in the study is available for everyone, it may serve as a dependent variable to distinguish the treatments via analysis of variance. Commonly, only whether each person is a dropout or a completer is known, in which case the less powerful chi-square analysis evaluating the independence of treatment condition and final status make optimal use of the qualitatave data.

The second case is one in which each participant is measured once, at the time of leaving the project. This situation was considered in a useful article by Lasky (1962). He proposed analysis of covariance on the scores, with length of time in treatment as the covariate. As Lasky himself noted, this simple scheme has the drawback that the covariate as well as the score would be expected to be affected by the treatment. The violation of the customary independence requirement may lead to severe problems of interpretation (Sprott, 1970). This problem may be avoided with an alternative analysis that employs the same information, using score and duration as two dependent variables in a multiple-regression analysis. A more complex regression-based analysis that can be applied to the dropout situation has been proposed by Welch, Frank, and Costello (1983).

Although the regression approach is appealing because it seems to make maximum use of the available information, the zero-implantation logic can be extended to yield a simpler and perhaps more appropriate analysis. In the crucial comparison across treatments, each participant's score using the zero-implantation method is the total number of compliance responses (such as pills taken) while in the study. The temporal distribution of the responses does not affect the assessment of the treatments. If this sum is an acceptable measure, then it is proposed that in the case of one observation per subject, the single score should be analyzed. An important advantage of this strategy is that, unlike the regression approach, it is not disrupted if a dropout occurs prior to measurement; a zero is used as the score, which affords a properly negative impact on the treatment generating the dropout. As far as comparing treatments is concerned, there is no essential difference between the one-score and several-scores cases; one simply replaces the score(s) that would have been gathered after the dropout with zeroes.

This recommendation yields a sensible integration of length of time in the study and actual scores, assuming that dropping out represents completely noncompliant behavior. If that assumption is unwarranted in an application, then a different weighting scheme is indicated. Regression-based analyses allow the data to determine the weights of the two kinds of information: the variable that discriminates among the

treatments more effectively gets greater weight. This responsiveness may be an advantage in some applications, but in compliance studies the researcher may not wish to afford the variables equal opportunity. Use of the zero-implantation scheme may thus be seen as a theoretical statement expressing the unique and final nature of dropping out.

REFERENCES

Baekeland, F., & Lundwall, L. (1975). Dropping out of treatment: A critical review. *Psychological Bulletin, 82,* 738–783.

Beck, N.C., Shekim, W., Fraps, C., Borgmeyer, A., & Witt, A. (1983). Prediction of discharges against medical advice from an alcohol and drug misuse treatment program. *Journal of Studies on Alcohol, 44,* 171–180.

Bigger, J.F. (1976). A comparison of patient compliance in treated vs. untreated ocular hypertension. *Transactions of the American Academy of Opthalmology and Otolaryngology, 81,* 277–285.

Caldwell, J.R., Cobb, S., Dowling, M.D., & de Jongh, D. (1970). The dropout problem in antihypertensive treatment. *Journal of Chronic Diseases, 22,* 579–592.

Federer, W.T. (1955). *Experimental design: Theory and application.* New York: Macmillan.

Finnerty, F.A., Mattie, E.C., & Finnerty, F.A. (1973). Hypertension in the inner city. I: Analysis of clinic dropouts. *Circulation, 47,* 73–75.

Gordis, L. (1979). Conceptual and methodologic problems in measuring patient compliance. In R.B. Haynes, D.W. Taylor, & D.L. Sackett (Eds.), *Compliance in health care,* (pp. 23–38). Baltimore: Johns Hopkins University Press.

Hagan, R.L., Foreyt, J.P., & Durham, T.W. (1976). The dropout problem: Reducing attrition in obesity research. *Behavior Therapy, 7,* 463–471.

Halperin, M., Rogot, E., Gurian, J., & Ederer, F. (1968). Sample sizes for medical trials with special reference to long-term therapy. *Journal of Chronic Diseases, 21,* 13–24.

Hecht, A.B. (1974). Improving medication compliance by teaching outpatients. *Nursing Forum, 13,* 112–129.

Krebs, R. (1971). Using attendance as a means of evaluating community mental health programs. *Community Mental Health Journal, 7,* 72–77.

Lange, R.A., Ulmer, R.A., & Weiss, D.J. (1986). An intervention to improve compliance to year-long Isoniazid (INH) therapy for tuberculosis. *Journal of Compliance in Health Care, 1,* 47–54.

Lasky, J.J. (1962). The problem of sample attrition in controlled treatment trials. *Journal of Nervous and Mental Disorders, 135,* 332–337.

Linn, M.W., Shane, R., Webb, N.L., & Pratt, T.C. (1979). Cultural factors and attrition in drug abuse treatment. *The International Journal of the Addictions, 14,* 259–280.

Oldridge, N. (1984). Compliance and dropout in cardiac exercise rehabilitation. *Journal of Cardiac Rehabilitation, 4,* 166–177.

Porter, A.M. (1969). Drug defaulting in a general practice. *British Medical Journal, 1,* 218–222.

Sprott, D.A. (1970). Note on Evans and Anastasio on the analysis of covariance. *Psychological Bulletin, 73,* 303–306.

Weiss, D.J. (1985). Snapshot analysis of variance: Comparing groups with unequal numbers of scores per subject. *Perceptual and Motor Skills, 61,* 420–422.

Welch, W.P., Frank, R.G., & Costello, A.J. (1983). Missing data in psychiatric research: A solution. *Psychological Bulletin, 94,* 177–180.

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture, 1,* 129–142.

Acknowledgments: I am indebted to J.D. Tate for suggesting the multiple regression analysis, and to Burton Alperson for a critical reading of an earlier draft.

Reprints: Requests for reprints should be sent to David J. Weiss, Department of Psychology, California State University, Los Angeles, CA 90032.

Received: April 1986
Revised: April 1986
Accepted: July 1986