A Behavioral Assumption for the Analysis of Missing Data: The Use of Implanted Zeroes

David J. Weiss
Department of Psychology, California State University
Los Angeles, CA 90032

Standard analysis of variance techniques for handling empty cells use the existing scores to estimate the missing data. Developed by agricultural researchers, these techniques are sensible enough when a few empty cells are randomly distributed throughout the design. In such cases, the absent data may be regarded as the result of accidental, scientifically uninteresting events. In behavioral research, however, missing data may arise from systematic causes. The present concern is with participants who voluntarily withdraw during the course of the research. In such cases, a distinctive, predictable pattern of empty cells will present itself. This pattern is informative and may be regarded as an important component of the data. The zero-implantation method is advocated as a way of integrating the actual scores with the information furnished by attrition. It is based upon the behavioral assumption that those who withdraw from a treatment will no longer derive benefit from it. In this method, ordinary repeated-measures analysis is carried out, with zeroes used as the post-withdrawal scores for dropouts. Response scales with the ratio property are required.

The statistical procedures upon which psychologists rely were developed primarily by non-psychologists. Fisher and his colleagues, who developed analysis of variance, were agricultural scientists. The experimental unit for the agronomist was usually a particular plot of land, and the primary dependent variable was yield. For the psychologist, the analogies to experimental organism and quantifiable behavior were irresistible.

The plot and the subject have important elements in common. They show individual differences which may be analyzed but are usually not of primary interest. Extracting these differences requires that each unit be amenable to repeated measurements. Each plot and each subject may be

---

Author's Note: This paper is based upon an invited presentation given at the Diamond Jubilee Meeting of the Indian Science Congress Association, Pune, January 1988. Tables 1 and 2 have been reproduced from Weiss (1987a) with permission from Springer Publishing Company. I wish to thank Burton Alperson for discussions which helped me to clarify the exposition.

Journal of Social Behavior and Personality, 1991, Vol. 6, No. 4, 955-964.

regarded as having been drawn randomly from a larger population. Random assignment of plots or subjects to the experimental treatments has the happy effect of eliminating bias. Differences among individuals may reduce the power of the experiment, but they are unlikely to favor one treatment over another.

The power of this analogy is undeniable, and it has allowed researchers to conquer the variability which makes behavior so difficult to study. It is fair to say that in the forty years since the strong advocacy of Garrett & Zubin (1943), analysis of variance has become the primary statistical technique used to compare the effectiveness of experimental manipulations.[1]

Perhaps it is heretical to suggest that the analogy is imperfect. Nevertheless, an important difference between the plot and the subject seems to have escaped notice. This difference has vital statistical consequences. While plots of land are fixed, human participants in behavioral research are free either to contribute, or to withhold, their scores. The difference shows up when one is confronted with the problem of missing data.

Missing data are considered a normal, if unfortunate, happenstance in experimental work. I wish to consider the situation in which the investigator has some of the scores scheduled to be collected from an individual, but not all.[2] The situation is not merely a matter of unequal group sizes, but rather one in which some cells of the design are empty. Statistical analysts deal with this problem in a routine way. The absent values are estimated from a subset of the data which are present. This approach may be credited to Yates (1933), though the basis probably has roots in antiquity. More recent algorithms differ in the details (Glenn & Kramer, 1958; Beale & Little, 1975), but the fundamental logic remains that the scores which are available afford the best prediction of the scores which are not. In formal terms, the observed scores and the missing scores are considered random variables having the same distribution. Even the non-algorithmic approach, i.e., simply ignoring the missing data, makes an essentially equivalent assumption. The position implicitly taken is that the observations we do have accurately represent those we would have had if all subjects had completed their assignments.

Underlying this standard approach is surely the idea that the missing data are unavailable because of an accident of some kind. The usual example given is that through recording error, either human or mechani-

---

[1] *A historical account of the ascension of ANOVA has been given by Lovie (1981).*

[2] *The problem of estimating a participant's missing score when only one measurement per subject has been scheduled is not considered here. For three quite different proposed solutions see Welch, Frank, and Costello (1983); Weiss (1987a); and Rubin (1987).*

cal, a few numbers were lost. What is important here is that there is no bias, that is, no reason why those particular scores should have been the ones lost. This is a natural assumption when plots of land are the sources of the scores. The accident was as likely to occur for one plot's data as for another's. The pattern of missing values must not be related to the experimental treatments.

In experiments with human subjects, the standard assumption is sensible if there are a few missing scores, attributable to such seemingly random events as illness, scheduling difficulty, or equipment failure. But what happens if the missing scores are concentrated in some experimental conditions and not in others? Should their absence be regarded as accidental? If not, can the fact that scores are missing be of use in making inferences about the effects of the treatments?

A crucial difference between plots of land and human volunteers is that the plots stay where they are assigned and provide data on schedule. The volunteers, on the other hand, may drop out. Then the empty cells may not be distributed randomly over the experimental design, but may follow a distinctive, predictable pattern. An observation scheduled late in the experiment is more likely to be missing than an early one. Once a participant has missed a scheduled session, it is unlikely that there will be any subsequent scores for that individual (Weiss, 1987b). This ordered pattern of "missingness" has been termed *monotone* by Rubin (1987), working in the context of survey research.

People may withdraw for reasons not associated with the particulars of the research, which is an annoyance but not of great interest. However, they may also withdraw because they do not like the treatment to which they have been assigned. In such cases, the use of standard estimation techniques ignores crucial information about the treatments. Biased comparisons are a likely result (Howard, Krause, & Orlinsky, 1986).

Differential attrition rates convey information (Hansen, Collins, Malotte, Johnson, & Fielding, 1985). How shall statistical practice be revised to incorporate this valuable information which the subjects furnish through their reluctance to provide data? Where the analogy with agriculture breaks down, can a psychologically plausible assumption generate more effective analytic procedures?

Undoubtedly, no single solution will suffice for all cases. A theory of attrition is needed in order to develop a general statistical solution. For the situation in which participants withdraw from an unsatisfying treatment, a procedure has been proposed (Weiss, 1987a).

The research area in which the procedure presented here was developed is patient compliance. The paradigm is straightforward. Patients involved in a medical regimen are randomly assigned to an intervention condition which is designed to help them follow the program. The compliance of each patient is assessed at periodic intervals. The results allow comparing the efficacy of the various conditions. It is presumed that the medical regimen is effective. The comparative evaluation is of the behavioral interventions. Thus, the domain of compliance research is psychology rather than medicine.

As an example, let us consider patients for whom the medical regimen consists of taking daily pills for six months. We wish to determine whether a daily phone call is helpful, relative to no reminder, in following this assignment. Patients might be seen monthly and a determination made of the proportion of assigned pills taken during the previous period. (Not a trivial matter, e.g., Gordis, 1979, but I shall gloss over it here.) If all goes well, at the end of the study a comparison of the two groups can be carried out using standard analysis of variance. (Patients are nested under treatments and crossed with the six time periods.)

Suppose it turns out, however, that patients in the phone call group are far less likely to complete the six-month program. While in the study, these patients take pills at a high rate; but they tend simply to stop coming.[3] What inference can be made about the intervention? One might, of course, attempt to ask the patients about the reasons for their withdrawal, but that is likely to be a fruitless enterprise. Most dropouts cannot be reached, and those that can are not likely to be frank (Caldwell, Cobb, Dowling, & de Jongh, 1970). Would a patient be likely to report that she found the phone call annoying or demeaning? More likely, one would hear that a relative took sick, or a work schedule changed, and it was no longer convenient to report for the monthly assessment. A researcher could go mad attempting to interpret this kind of data. Why bother? Let the data speak for themselves. When people withdraw from a treatment, they are telling the researcher in the most direct way that the treatment is ineffective in achieving its goal. In the context of smoking cessation, Shewchuk and Wynder (1977) have similarly argued that dropouts should be regarded as treatment failures.

The statistical issue is how to weight the information furnished by the premature termination. How shall this information be combined with the information given by the actual scores? Consider the (fictitious) pill-taking proportions given in Table 1. The mean of the twenty-eight Group I scores is 72.68, while the mean of the twelve Group II scores in 83.5. But which group has derived more medical benefit? That is, which group has ingested more pills? Any scheme which uses the present data to predict

[3] *Lange, Ulmer, and Weiss (1986) observed this pattern of noncompliance with a different, rather ineffective, intervention.*

TABLE 1 Fictitious Pill-Taking Proportions

	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

the missing scores will conclude that the intervention for Group II was more effective. But that, in my view, is an incorrect conclusion.

The advocated method has a different basis for estimating the missing scores. If a present score reflects the amount of potential medical benefit extracted by the patient from the treatment, then so should a missing score. And if the patient has dropped out of the treatment, how much medical benefit can the treatment deliver afterward? The most plausible estimate of post-withdrawal benefit is surely none. Therefore the appropriate replacement for a missing score is a zero. Here a behavioral assumption is used to determine a statistical procedure. The *zero-implantation* method calls for the replacement of all post-dropout scores with zeroes, arguing that a score is an estimate of a behavioral reality.

With zeroes replacing the X's in Table 1, the group means yield a radically different, and I trust more appropriate, conclusion. The mean score for Group I drops slightly to 67.8, while the mean for Group II drops dramatically to 33.4. Statistical verification is given by the analysis of variance shown in Table 2. The key result is the F ratio for groups, significant at the .05 level. The analysis was carried out using a standard computer program for nested designs, with only one modification of the output necessary.

The degrees of freedom for the within-subjects term were reduced from the program's 40 to 20; 1 df is lost for each implanted zero. Degree of freedom adjustment is a customary price to pay for using estimated rather than real data. The corrected total df, 39, is what normally would be expected with 40 observed scores, reflecting the fact that implanted scores are not contributing in the way that real scores do.

TABLE 2 Analysis of Variance

Source	*df*	*SS*	*MS*	*F*
Groups	1	17784.8	17784.8	8.65*
Error	8	16444.8	2055.6	
Months	5	18755.1	3751.0	2.94*
Months x Groups	5	6015.1	1203.0	< 1
Error	20	25544.4	1277.2	

The use of implanted scores leads to two violations of the usual assumptions underlying analysis of variance. Firstly, the zeroes cannot be regarded as coming from a normal distribution; and secondly, their use will produce heterogeneity of variance. The result of these violations is low power. Type II errors are inevitable, which correctly reflects the paucity of actual data. In contrast, the well-known robustness of ANOVA protects against serious disruption of the Type I error rate.

From a qualitative perspective, at least, the zero-implantation method works sensibly. The more dropouts a condition generates, the lower the mean score will be. This reflects the behavioral reality that compliance in a condition with many dropouts, and especially with early dropouts, is low. When actual scores are comparable but dropout rates are markedly different across conditions, zero-implantation will be more likely than a standard analysis to report differential treatment efficacy.

When dropout rates are high, the proposed method has low power for the time period evaluations, as the df for that error term will be reduced. The analysis captures the fact that there is not much real information about temporal effects unless most subjects can be tracked through the entire schedule.

An attractive aspect of this analysis is that it does not require the researcher to interpret individual occurrences of withdrawal. The procedure may be applied in a purely mechanical fashion, and no harm will be done if in fact attrition rates are low in a particular experiment. In such cases, implantation will have little effect.

Consumer Protection Warning

It is important to note that zero-implantation is not a general solution to the problem of missing data in behavioral research. Indeed, it is an applicable method only in the specific (but not uncommon) circumstance that zero is a plausible estimate of the missing score. This requires a psychological, rather than a statistical, decision. In the pill-taking example, the decision is rather easy because the patient has no access to the medicine outside of the study. But if the behavior in question were privately feasible, if for example the study were designed to promote daily exercise, the researcher would have to make a more difficult evaluation concerning the score to be assigned.

I should like to emphasize that, although the scoring decision may be difficult, one cannot avoid it by using a standard approach to missing data. That too is a decision, one which states in effect that the dropout will maintain the same "exercise level" after withdrawal that was achieved while under supervision. A similar kind of decision is required in follow-up studies when some participants cannot be reached. Bélisle, Roskies, and Lévesque (1987) assigned a score of zero as the estimate of the number of post-program exercise periods completed by untraceable registrants.

The important restriction is that zero must represent the worst possible performance in the setting. The problem can be seen by considering a study designed to reduce obesity which uses weight loss as the measure of efficacy. Is the best guess about a dropout that weight is being maintained? Hardly. Assigning zeroes might well flatter the manipulation which generated the withdrawal. Technically, response scales for which zero truly represents the bottom are known as ratio scales.[4] Indices which record the occurrence of a directed behavior will generally have this property. Acceptable examples include minutes of exercise, sessions attended, assignments completed and the like. It may not always be possible to find a practical measure which has the ratio property, and in such cases no recommendation for coping with attrition is offered.

Treatments from which the volunteer drops out because he is pleased with the current outcome present an apparent difficulty.[5] For example, a patient might withdraw from psychotherapy because he feels sufficient mental health has been attained. From the researcher's point of view, however, this happy patient is a dropout and his data should be analyzed as such. The resolution to the difficulty is to focus on precisely what the

---

[4] *Strictly speaking, the "zero" of the scale need not actually be represented by the number zero, so long as it represents the lowest value on the scale, i.e., the absence of what is being measured. Thus -273 is the zero of the Kelvin scale of temperature. Most of the zeroes we encounter in psychology are not "zeroes" because our contrived measures are not ratio scales; one does not have zero aptitude or zero attitude. Scale typology has been discussed by Suppes and Zinnes (1963).*

[5] *A distinction between "dropouts" and "terminators" has been raised by Morrow, Del Gaudio, and Carpenter (1977).*

score means. The implanted zero measures the amount of benefit that the volunteer is getting from the program after withdrawal. The locus of the researcher's problem is the experimental design itself. The design calls for participants to be crossed with time periods. Thus it is presumed that a fixed duration of treatment, an interval during which benefit is available, is appropriate for all patients. While some patients may need even further care, for this analysis to be valid all of them must be in a position to benefit from complying throughout the experimental period.

The researcher must also fix an appropriate inter-measurement interval. The more frequent the measuring occasions for a given duration of treatment, the more impact the early dropouts will have on the between-groups comparison. Phrasing the same idea from another perspective, it may be said that the experiment is more sensitive to differential attrition rates if there are many measurements scheduled.

CONCLUSION

Dropout rate has been used as a dependent variable by many researchers, and it is common to find variation across treatments (Weiss, 1987a). However, the integration of attrition and observed performance to achieve additional insight into treatment effectiveness has not previously been featured in a formal method. Previous workers considering attrition have viewed it as a separate issue, with statistical assessment varying in sophistication from counting dropouts vs completers (Linn, Shane, Webb, & Pratt, 1979) to encoding length of stay as a measure (Lasky, 1962). Certainly an approach which separates attrition and response magnitude may be fruitful in some situations, particularly if there is theoretical interest in attrition. It is therefore advisable for investigators to report the components separately.[6]

In applied settings, though, the pragmatic issue of treatment efficacy demands joint consideration of both aspects, performance and attrition. A given treatment may show its effect on one aspect or the other, and a proper evaluation of the customary research hypothesis must incorporate both. The proposed method accomplishes the integration in a natural way, constructing a unidimensional variable which may be thought of as overall benefit gained by the patient from the compliance intervention. This variable has obvious utility in assessing the effectiveness of an intervention.

The applicability of zero-implantation may be extended to areas beyond the realm of compliance. Widening the sphere will make it easier

[6] *Separate reporting also maintains the archival relevance of the data should another analytic method become standard in the future. Empirical evidence may dictate an alternative scheme for weighting attrition.*

for an individual researcher to undertake the risk of using a new analytic technique. For example, training programs, in which the student or trainee has the freedom to withdraw during the course of evaluation, would seem amenable to the analysis. Consider a comparison of the effectiveness of interventions designed to help students succeed in an academic environment, such as that of Bloom (1971). Each student's quarterly GPA would be the primary measure. When a student drops out, zeroes would be implanted until scheduled graduation. Early dropouts will have a dramatically adverse effect, reflecting the fact that dropping out is an ultimate failure for a program emphasizing retention. In a similar manner, one might evaluate programs for training salespersons using monthly sales figures as the basic measure. Here an arbitrary decision would have to be made about the duration of the evaluation period.

Perhaps it is fitting that a final example be drawn from agriculture. The nutritional value of various crops as animal feeds is an important issue to the farmer. The researcher may evaluate the diet fed to ewes by periodically measuring the weight gain of their lambs (Kenney & Smith, 1985). Corresponding to a dropout is the death of a lamb, which is scored as an ultimate failure of the diet. The zero-implantation method combines the weight gains with the deaths in a natural way.

The hallmark of the technique is the integration of behavioral assumption and statistical procedure. This intimate connection calls to mind the basic tenet of functional measurement (Anderson, 1977) that theory and measurement go hand in hand. One must be aware that any statistical procedure, whether historically entrenched or novel, involves behavioral assumptions when it is applied to data. The researcher should choose the procedure with the assumptions in mind, rather than simply accept the default option offered by a computer package. The validity of zero-implantation, or of any missing data method, depends upon the generally unobservable level of performance subsequent to attrition. Post-treatment performance is an empirical question worthy of study in its own right (Lee & Owen, 1986).

REFERENCES

Anderson, N.H. (1977). Note on functional measurement and data analysis. *Perception & Psychophysics, 21,* 201-215.

Beale, E.M.L., & Little, R.J.A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society, Series B, 37,* 129-145.

Bélisle, M., Roskies, E., & Lévesque, J.M. (1987). Improving adherence to physical activity. *Health Psychology, 6,* 159-172.

Bloom, B.L. (1971). A university freshman preventive intervention program: Report of a pilot project. *Journal of Consulting and Clinical Psychology, 37,* 235-242.

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.

Garrett, H.E., & Zubin, J. (1943). The analysis of variance in psychological research. *Psychological Bulletin, 40*, 233-267.

Glenn, W.A., & Kramer, C.Y. (1958). Analysis of variance of a randomized block design with missing observations. *Applied Statistics, 7*, 173-185.

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.

Hansen, W.B., Collins, L.M., Malotte, C.K., Johnson, C.A., & Fielding, J.E. (1985). Attrition in prevention research. *Journal of Behavioral Medicine, 8*, 261-275.

Howard, K.I., Krause, M.S., & Orlinsky, D.E. (1986). The attrition dilemma: Toward a new strategy for psychotherapy research. *Journal of Consulting and Clinical Psychology, 54*, 106-110.

Kenney, P.A., & Smith, R.S. (1985). Effects of including lupins with cereal grain rations on the production of lambing ewes during drought. *Australian Journal of Experimental Agriculture, 25*, 529-535.

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.

Lee, C., & Owen, N. (1986). Community exercise programs: Follow-up difficulty and outcome. *Journal of Behavioral Medicine, 9*, 111-117.

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

Lovie, A.D. (1981). On the early history of ANOVA in the analysis of repeated measures designs in psychology. *British Journal of Mathematical and Statistical Psychology, 34*, 1-15.

Morrow, G.B., Del Gaudio, A.C., & Carpenter, P.J. (1977). The dropout and the terminator: A methodological note on definitions. *Journal of Clinical Psychology, 33*, 867-869.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Shewchuk, L.A., & Wynder, E.L. (1977). Guidelines on smoking cessation clinics. *Preventive Medicine, 6*, 130-133.

Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 1-76). New York: Wiley.

Weiss, D.J. (1987a). Dropout in patient compliance studies: A suggested analytic procedure. *Journal of Compliance in Health Care, 2*, 73-83.

Weiss, D.J. (1987b). *Use of the hazard function to explore dropout*. Paper presented at the Mathematical Psychology Meeting, Berkeley, CA.

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.