Increasing Compliance to Medical Treatment Regimens: A Meta-Analysis of Program Evaluation

Emil J. Posavac, James M. Sinacore, Sarah E. Brotherton, Michael C. Helford and Robin S. Turpin

Eval Health Prof 1985; 8; 7
DOI: 10.1177/016327878500800102

INCREASING COMPLIANCE TO MEDICAL TREATMENT REGIMENS

A Meta-Analysis of Program Evaluation

EMIL J. POSAVAC
JAMES M. SINACORE
SARAH E. BROTHERTON
MICHAEL C. HELFORD
ROBIN S. TURPIN
Loyola University of Chicago

Research articles evaluating the effectiveness of programs to increase compliance with medical treatment regimens were quantitatively integrated to assess the impact of these programs on the behavior of patients. A total of 58 studies involving two or more groups with 133 measures of compliance were identified and analyzed. The mean effect size was .47, indicating that the typical program participant complied better than 68% of the members of the control groups. The advantage of the program groups dropped as the amount of lifestyle changes required by the treatment regimen increased. Overall, the most successful interventions involved improving the facility providing care and helping patients to incorporate the treatment regimen into their daily routine. It is suggested that published evaluations of compliance programs would be more useful and more likely to contribute to an accumulation of knowledge if more careful descriptions of the interventions, including costs estimates, were included in reports of the program evaluations.

AUTHORS' NOTE: Requests for reprints should be addressed to Emil J. Posavac, Department of Psychology, Loyola University of Chicago, 6525 North Sheridan Rd., Chicago, IL 60626.

EVALUATION & THE HEALTH PROFESSIONS, Vol. 8 No. 1, March 1985 7-22

One of the continuing problems facing medical care providers is the level of patient noncompliance with medical treatments. Numerous observers have examined the problem and their conclusions suggest that 30%-75% of patients do not comply adequately with the treatment that is suggested for the ailment presented to the physician (Haynes et al., 1979; Sackett and Haynes, 1976). Not surprisingly, noncompliance differs depending on the type of treatment prescribed and the definition of compliance used, and it differs across educational levels of patients, illnesses, and types of providers. However, these differences do not negate the general finding that noncompliance with medical advice is often seen as a threat to the health of patients and even to the validity of conclusions drawn from experimental studies of the effectiveness of newly developed medications (Goldsmith, 1979).

Health-care journals frequently publish material written to aid the provider to motivate patients to take prescribed medications, return for follow-up visits, stay on diets, exercise appropriately, and so on. A relatively small number of the authors of these articles have undertaken the task of evaluating the effectiveness of steps suggested to increase compliance with medical treatments. Just as no single basic research finding is definitive, no evaluation is the final word on the effectiveness of the program evaluated. Thus, it is useful to review evaluative research and to integrate findings if programs to increase compliance are to be widely adopted. The best approach to research review and quantitative integration, however, is not clear (Cohen, 1977).

Meta-analysis is an approach to research integration that offers a quantitative tool to utilize the power that a multitude of studies possesses (Glass, 1978; Wortman, 1983). Glass points out that when taken together, many reports each having low statistical power can pinpoint a consistent effect. The central concept in meta-analysis is effect size. The effect size of a dependent variable, X, is defined as: $ES_x = (\bar{X}_e - \bar{X}_c) / s_{xc}$, where $\bar{X}_e$ is the mean of the experimental group, $\bar{X}_c$ is the mean of the control group and s_{xc} is the standard deviation of the control group. The difference between the means of the experimental group and the

control group is the amount the program improved the patient's behavior on variable X. Thus, effect size is like a z-score for the mean of the experimental group relative to the distribution of the dependent variable among the people making up the control group. If it is assumed that the dependent variable is normally distributed in the control group, the effect size can be converted into a percentile rank indicating where the experimental group's mean would be found relative to the control group distribution. This formula has been used in a number of meta-analyses (see, for example, Devine and Cook, 1983).

It is important to note the major benefits of the effect size calculation. First, evaluative studies are often hard to compare with each other because they lack a common metric. Is the improvement of three points equivalent to the improvement of 20 points in the next study? By converting findings into a unitless index, study findings are comparable. The second intent of a meta-analysis is to convert findings often expressed as a test of a hypothesis into a statement on strength of findings. Readers of research, especially applied research, are concerned about the strength as well as the reliability of findings. Focusing on tests of significance limits the reader to information on reliability and may interfere with evaluating the strength of the intervention. By converting results into a common metric that can be further converted into an easily understood percentile that indicates the strength of an effect, it is hoped that the clarity of findings will be improved. This concern is especially applicable to program evaluation, an applied social and management science. Whereas the basic researcher may be content with the conclusion that independent variable Y does effect dependent variable M, the applied researcher cannot accept an analysis that stops at this point because weak, even if reliable, effects ought not to become the focus of a service program. A different problem occurs when the use of small sample sizes so reduces the power of a research design that reliable findings are nearly impossible to obtain (Freiman et al, 1978). Effect size calculations summarized over many weak designs may isolate a useful finding that would otherwise have been overlooked.

PROCEDURE

SOURCE OF EVALUATIONS

Using the MEDLINE computer-based search procedure, the *Hospital Literature Index,* and Haynes et al. (1979), a search of the compliance research literature revealed that many articles are available to help clinicians improve the patient's compliance. There are, however, relatively few articles that assess the success of such suggestions to clinicians. Among those studies in which data were provided, some had to be omitted because the study focused on only one group's response to the treatment. Effect size can be calculated from pretest-posttest designs; however, such evaluations were not included in this review. Studies were also eliminated from the sample if they concerned primarily substance abuse, weight reduction, or general psychotherapy programs. The sample of evaluations described here concerned programs to encourage adherence to medical treatments or in a few cases programs to encourage people to seek treatment for a medical problem. However, programs aimed at the general public were not used. The search located 58 evaluations published through December 1982 that met these criteria. Because many of these 58 reports contained more than one experimental group, there were a total of 96 different interventions evaluated in this set of studies.

RESULTS

DESCRIPTIONS OF THE SAMPLE

Program descriptions. The characteristics of the interventions are presented in Tables 1 and 2. Two types of programs made up the largest portion of these evaluations. The single most common type of compliance program was an educational program (53%). A reorganization of the facility or its procedures was the intervention of nearly a fifth of the programs (19%). Ensuring the continuity of care through a change in the scheduling procedures would be one such organizational change.

TABLE 1
Nature of the Compliance Interventions Evaluated in the Studies Analyzed

	n	%
Providing information	51	53
Reorganizing facility	18	19
Including in daily routine	10	10
Teaching self-monitoring	8	8
Encouraging social support	6	5
Shaping adherence	2	2
Multiple	2	2
Total	96	99%

Table 2 shows that the largest portion of the programs (44%) were based on the mode of communication most common in medical settings: one-to-one communication. Only written material was used in 15% of the programs. Nine programs (9%), most of which involved a reorganization of care, did not utilize any media at all.

There was a wide variety in the identity of the persons who prepared the compliance programs. Given that many of the problems in medical compliance are directed to ambulatory patients who don't take all of their prescribed medications, it is not surprising that nearly a third of the programs were developed by clinic pharmacists (33%). Health researchers constructed over a fifth of the programs (22%), and together physicians and nurses prepared 40%. Two programs were prepared by health educators and one by an unspecified health-care provider. It can be safely assumed that the overrepresentation of health researchers is due

TABLE 2
Media Used in the Compliance Interventions Included in the Analysis

	n	%
One-on-one contact	42	44
Multiple media	21	22
Written material	14	15
Classes	5	5
Audio tape	1	1
None	9	9
Other	4	4
Total	96	100%

to their desire to publish the results of their work. It seems quite likely that physicians and nurses do, in fact, prepare an even greater proportion of compliance programs than health researchers do; however, care providers are less likely to perform experiments or to publish articles.

Descriptors of programs were examined to learn how they were conceptualized. Only 11 of 58 program developers mentioned any theoretical basis to the program design. These reports mentioned, for example, shaping through reinforcement of desired behavior, basic research on social support, or the necessity of systematic knowledge of one's own behavior as inspirations for the development of the programs. Other programs were based on common sense assumptions that better knowledge of treatment or regimen or increasing the continuity of care would lead to higher rates of compliance.

Program participants. Evaluations of compliance programs are largely conducted with outpatients (69%). The median number of patients in an evaluation was 100. Among the 23 studies that included a gender breakdown, the median percentage of female patients was 55% versus 45% for male patients. Because women make more use of medical facilities than men do, this imbalance does not imply that the people taking part in these compliance evaluations were unrepresentative of patients in general. The median age of the patients was 54 years.

Experimental designs. The studies were chosen only if there were at least two groups involved, one with the program and one without. On these two or more group evaluations, 69% were experimental designs with random assignment to groups, 26% utilized existing natural groups, and 3% used self-selected groups. One writer failed to mention how the groups were formed. There was considerable attrition from some studies; the median attrition rate was 13%. It should be noted that attrition from medical treatment was the index of compliance in 15 studies. For these studies attrition did not weaken the research design and the attrition rates of these studies are not included in the rate just reported.

One final but important characteristic of the evaluations concerns an assessment of the degree the programs were actually implemented as planned. Twelve of the interventions consisted of changes that did not require the documentation of implementation. Among the 46 reports concerning programs requiring some teaching, training, distribution of written material, and so on, only 12 (26%) contained a specific comment on procedures used to examine the degree the program was implemented. Given that the evaluation literature contains many comments on the necessity of making special efforts to document that programs are actually carried out (see, for example, Sechrest et al., 1979; Rezmovic, 1984), this rate of reporting on implementation seems low.

COMPLIANCE EFFECT SIZES

Overall impact. The effects of a number of 96 interventions were assessed by more than one measure of compliance. Thus, 133 measures of compliance were examined. Because some writers did not provide enough information to permit the calculation of an effect size, effect sizes of 126 measures of program success were analyzed. The overall mean effect size for these 126 measures was .47 with a standard deviation of .56. An effect size of .47 indicates that the average member of the program groups complied to the same degree as someone at the sixty-eighth percentile rank in the control groups.

Comparing types of compliance. The overall mean effect size is informative; however, a better understanding of the impact of these evaluations can be obtained if the data are subdivided by various characteristics of the evaluations. Table 3 divides effect sizes by type of compliance measure and type of intervention. As shown in Table 3 there were three major types of dependent variables (i.e., types of compliance measure): (1) actual treatment compliance, such as taking medication; (2) being available for care, such as keeping appointments; and, (3) behavioral change, such as losing weight. When mean effect sizes were calculated separately for these three types of compliance, striking differences were found. As the column means show, changing patients' level of compliance to medication prescriptions is markedly easier than getting patients to come in for care: $t\ (105) = 2.54$, $p < .01$. Although not statistically reliable, it is easier to improve the rate of seeking care than increasing the degree people will make lifestyle behavioral changes. There were only four evaluations on preventive programs; however, the data suggest that the compliance rate of people who are not experiencing symptoms is not very easy to change.

Comparing types of interventions. Table 3 also provides comparisons of the impact of different approaches to increasing compliance among patients. Among the more frequently used

TABLE 3
Mean Effect Size Divided by Intervention Type and Measure of Compliance

Type of Intervention	Compliance Measure				
	Treatment compliance	Seeking needed care	Life style change	Preventive care	Total
Providing information	.54(38)[a]	.31(19)	.11(8)	--(0)	.42(65)
Reorganizing facility	.79(7)	.39(12)	2.10(1)	.19(4)	.54(24)
Teaching self-monitoring	.24(7)	.66(4)	.04(4)	--(0)	.30(15)
Including in daily routine	.89(8)	.21(3)	--(0)	--(0)	.71(11)
Encouraging social support	.62(4)	-.04(2)	-.11(2)	--(0)	.27(8)
Shaping compliance	1.08(1)	.92(1)	--(0)	--(0)	1.00(2)
Multiple interventions	2.24(1)	--(0)	--(0)	--(0)	2.24(1)
Total	.62(66)	.35(41)	.19(15)	.19(4)	.47(126)

a. Number of effect sizes.

approaches, helping patients to include the desired behavior in their daily routines was the single most effective intervention, mean effect size = .71. This effect size means that the mean level of compliance of program patients was greater than the level of compliance of 76% of the control group patients. Also quite successful were interventions based on changing the facility to provide greater continuity of care, less waiting, reminders of follow-up appointments, and so on. The mean effect size of .54 is based on 24 effect sizes including 18 rated as very nonreactive. Overall, the impacts of reorganizations of care were estimated using the dependent variables, which were among the least reactive of the various types of interventions.

Other variables related to effect size. In order to examine other characteristics of the evaluations that are related to effect size, a multiple regression analysis was conducted using effect size as the dependent variable and compliance measure rating of the reactivity of the measure (five levels), type of intervention, degree of experimental control (three levels), level of attrition (four levels), whether cost of the intervention was described (1, no; 2, yes), and whether theory was used in program design (1, no; 2, yes). Type of compliance measure was converted into two dummy variables; the four preventive care evaluations were dropped. Type of intervention was converted into four dummy variables. There were so few evaluations of shaping and multiple interventions that they were dropped.

A backward method of multiple regression was chosen. This method uses all independent variables and removes variables, one at a time, if they do not meet the criterion for retention. In an exploratory analysis it was felt that a liberal criterion for retention would be appropriate. Consequently, variables were eliminated only when their probability exceeded 0.10. Table 4 provides the regression Beta weights, t values, and probability levels. The analysis revealed which characteristics of evaluations were the most highly related to impact, as measured by effect size. The programs with largest effect sizes tended to (a) seek to influence treatment/medication compliance, (b) be evaluated

TABLE 4
Regression Weights and t Values from Regression Analysis

Independent variable	Beta	t	p
Treatment compliance[b]	.30	3.47	.0008
Experimental control	.28	3.23	.0017
Cost mentioned	.21	2.41	.0179
Reorganizing facility[c]	.19	2.09	.0390
Including in daily routine[d]	.16	1.83	.0699

a. df = .05.
b. Treatment compliance versus all other types of compliance measure.
c. Reorganizing facility versus all other interventions.
d. Including in daily routine versus all other interventions.

using self-selected volunteers for the program groups, (c) have costs mentioned, (d) involve a reorganization of care facilities or procedures, and (e) help patients to include compliance with medical assistance in their daily routines. The explicit use of theory, greater attrition, and the use of more reactive measures were related to larger effect sizes; however, the levels of correlation were insufficient to merit retention in the regression formula.

DISCUSSION

The overall mean effect size found in these studies of compliance (.47) is within the range of effect sizes found in evaluations of patient services. For example, Devine and Cook (1983) found a mean effect size of .49 on hospital length of stay among 34 evaluations of hospital-based interventions involving patient education and psychological support. Furthermore, Cohen (1977) suggested that an effect size of .50 was a "moderate" sized

finding. There are some tantalizing hints of more powerful programs in Table 3. The one program utilizing multiple interventions had the largest impact, and the two based on the behavioral principle of rewarding successive approximations of outpatient compliant behavior were very strong as well. One can only hope that additional programs such as these three are developed and evaluated. As the literature stands, there is not enough evidence to permit more than stating that these programs suggest a potential area for further investigation.

RELATIONSHIP OF EXPERIMENTAL CONTROL AND EFFECT SIZE

It is often maintained that stronger effects are obtained from the least well-controlled research. Gilbert et al. (1975) demonstrated this pattern in medical treatment research. Ramsey (1983) showed that attrition can have an impact on effect size because (a) volunteers for compliance studies had been more compliant in their past medical behavior than non-volunteers and (b) early dropouts from a compliance study were among the least compliant. She concluded that many compliance studies do not focus on "hardcore" noncompliant patients because they never get into the research.

The present review supported some of the observations made by Gilbert et al. (1975). The significant regression weight for quality of research design was due to the large effect size (1.06) found with the six compliance measures reported in the three studies based on self-selected groups. (Although the mean effect size found in quasi-experiments (.46) exceeded that found in true experiments (.42), this difference was not statistically reliable.) This review could not directly address Ramsey's (1983) points; however, her findings would suggest that the studies with the least attrition might have the lowest effect sizes because the least compliant patients would be more likely to be still in the study when compliance was measured. However, level of attrition was not linearly related to effect size. Among the evaluations that did

not use attrition as the index of compliance, effect size was the largest for those studies with the *least* attrition (i.e., 5% or less). Without disputing Ramsey's points, it may be that the evaluations with the lowest rates of attrition were also the best designed and implemented and, therefore, the most likely to show an effect of the intervention.

ADVANTAGES OF USING EFFECT SIZE INDEXES

The advantage of using effect size in evaluative reviews has come to be widely recognized (see Light, 1983; Wortman, 1983). The value to research users, although less widely discussed, is illustrated in the comparisons described in this article. Clinicians and health researchers seeking to increase compliance need to recognize that the degree of success of a program will greatly depend on what behavior it is designed to affect. Or, conversely, clinicians setting out to change the lifestyle of patients had better be ready with a strong program or they will fail. Unfortunately, as Sechrest et al. (1979) lament, there are no standard ways of measuring the "dosage level" of a program designed to affect behavior. Taking the program in isolation, neither the clinician nor the evaluator can know whether a program provides a strong treatment or merely a weak one.

There are some ways to meet the program planners' need to estimate the strength of a planned program. The reports reviewed here contain remarkably few descriptions of the interventions and very few discuss how the programs were planned or implemented. Only 24% of the reports mentioned anything about costs. To develop a research-supported knowledge base for designing compliance interventions requires more thorough and complete reports of evaluations.

Even without such extended descriptions a consistent use of effect size and the resultant percentile rank could have a salutary effect on the realism of clinical practitioners as they plan programs. The authors' experiences as well as that of others (e.g., Posavac and Carey, in press; Rosenthal and Rubin, 1982; Weiss,

1972) suggest that program developers overestimate the impact that interventions will have. This may be a result of the need to "sell" the program before implementation or simply an inaccurate estimate of the possible effect that an intervention may have. The degree of overlap between treatment and control patient groups indicates that even among the more well-conceived and well-implemented programs, impact is often modest. Note that even though the programs on the whole are effective in helping people to comply, there is still much overlap between the compliance rates of the patients in the programs and those not in the programs. Specifically, 32% (i.e., 100% minus 68%) of non-program patients did indeed comply more than the average person who received special attention.

Moreover, the greater the degree to which the program depends on changing the motivations of individuals, the harder it will be to implement the intervention and the harder it will be to effect change (Caplan and Nelson, 1973). This review concurs with the psychological literature (Katz, 1974) in suggesting that a brief, orally-delivered presentation of information will not have consistent and long-lasting effects. Likewise, the mere provision of written material has limited impact (Morris and Halperin, 1979). Thus, changing the life styles of patients who smoke, are overweight, or display a Type A behavior pattern will be quite difficult in the context of typical encounters between physicians or nurses and patients. Choosing to develop a program administered to patients individually will be very hard to implement because medical care staffs have little additional time to devote to individual patients. Furthermore, such approaches are often expensive. More success is likely to be achieved through work with the environment of the patient. Deyo and Inui (1980) remark that organizational factors are quite effective in improving the rate of meeting follow-up appointments and are under the control of the facility. Their comments are supported by the present review. Helping patients incorporate the regimen into daily routine is perhaps less obviously an environmental intervention, but a crucial issue seems to be the provision of appropriately frequent cues to encourage compliance. By relating the regimen

to daily routines compliance becomes natural—not something the patient must keep remembering. Neither does the patient need to recall the reasons for complying.

In summary then, the present review illustrates that compliance can be enhanced most effectively by improving the procedures in health-care organizations—an environmental change—and by incorporating the regimen into the patient's daily routine—an intervention that is sensitive to the patient's life pattern and environment.

REFERENCES

CAPLAN, N. and S. D. NELSON (1973) "On being useful: the nature and consequences of psychological research on social problems." Amer. Psychologist, 28, (March): 199-212.

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

DEVINE, E. C. and T. D. COOK (1983) "Effects of psycho-educational intervention on length of hospital stay: a meta-analytic review of 34 studies," 417-432 in R. J. Light (ed.) Evaluation Studies Review Annual: Vol. 8. Beverly Hills, CA: Sage.

DEYO, R. A. and T. S. INUI (1980) "Dropouts and broken appointments." Medical Care 18, (November): 1146-1157.

FREIMAN, J. A., T. C. CHALMERS, H. SMITH JR., and R. R. KUEBLER (1978) "The importance of *beta,* the Type II error, and sample size in the design and interpretation of the randomized control trial." New England J. of Medicine 299, (September 28): 690-694.

GILBERT, J. P., R. J. LIGHT and F. M. MOESTELLER (1975) "Assessing social interventions: an empirical base for policy," pp. 39-193 in C. A. Bennett and A. A. Lumsdaine (eds.) Evaluation and Experiment. New York: Academic.

GLASS, G. V. (1978) "Integrating findings: the meta-analysis of research," pp. 351-379 in L. Shulman (ed.), Review of Research in Education, Vol. 5. Itasca, IL: Peacock.

GOLDSMITH, C. (1979) "The effect of differing compliance distributions on the planning and statistical analysis of therapeutic trials," pp. 137-166 in D. L. Sackett and R. B. Haynes (eds.) Compliance with Therapeutic Regimens. Baltimore: John Hopkins Univ. Press.

HAYNES, R. B., D. W. TAYLOR and D. L. SACKETT (1979) Compliance in Health Care. Baltimore: John Hopkins Univ. Press.

KATZ, D. (1974) "Factors affecting social change: a social-psychological interpretation." J. of Social Issues 30, 3: 159-180.

LIGHT, R. J. (1983) "Introduction," pp. 13-29 in R. J. Light (ed.) Evaluation Studies Review Annual: Vol. 8. Beverly Hills, CA: Sage.

MORRIS, L. A. and J. A. HALPERIN (1979) "Effects of written drug information on patient knowledge and compliance: a literature review." Amer. J. of Public Health 69, (January): 47-52.

POSAVAC, E. J. and R. G. CAREY (in press) Program Evaluation: Methods and Case Studies (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

RAMSEY, J. (1983) "Impact of attrition on studies of noncompliance." J. of Applied Psychology, 68, (January): 166-169.

REZMOVIC, E. L. (1984) "Assessing treatment implementation amid the slings and arrows of reality." Evaluation Rev. 8 (April): 187-204.

ROSENTHAL, R. and D. B. RUBIN (1982) "A simple, general purpose display of magnitude of experimental effect." J. of Educational Psychology 74, (April): 166-169.

SACKETT, D. L. and R. B. HAYNES (eds.) (1976) Compliance with Therapeutic Regimens. Baltimore: John Hopkins Univ. Press.

SECHREST, L., S. G. WEST, M. A. PHILLIPS, R. REDNER, and W. YEATON (1979) "Introduction," pp. 15-35 in L. Sechrest, et al. (eds.) Evaluation Studies Review Annual: Vol. 4. Beverly Hills, CA: Sage.

WEISS, C. H. (1972) Evaluation Research. Englewood Cliffs, NJ: Prentice-Hall.

WORTMAN, P. M. (1983) "Evaluation research: a methodological perspective." Annual Rev. of Psychology 34: 223-260.