

AN INVITATION TO DEBATE

How to have a high success rate in treatment: advice for evaluators of alcoholism programs

WILLIAM R. MILLER¹ & MARTHA SANCHEZ-CRAIG²

¹Department of Psychology, University of New Mexico, Albuquerque, New Mexico, USA;

²Addiction Research Foundation, Toronto, Ontario, Canada

Abstract

Two seasoned alcohol treatment researchers offer tongue-in-cheek advice to novice program evaluators faced with increasing pressure to show high success rates. Based on published examples, they advise: (1) choose only good prognosis cases to evaluate; (2) keep follow-up periods as short as possible; (3) avoid control and comparison groups; (4) choose measures carefully; (5) focus only on alcohol outcomes; (6) use liberal definitions of success; (7) rely solely upon self-report and (8) always declare victory regardless of findings.

Introduction

The documentation of favorable outcomes from residential and ambulatory treatment is an issue of increasing importance. With an expanding international market to which treatment programs can be exported, it is advantageous to be able to cite high success rates. There is also growing pressure for program evaluation from funding sources, insurers, consumers and regulatory agencies such as the US Joint Commission on Accreditation of Healthcare Organizations. In competitive advertising some treatment programs assert their superiority, and at least one has openly claimed in a national campaign to have "the number one success rate". When asked informally about their outcomes, the normative reply of treatment programs appears to be in excess of 80% success (Miller & Hester, 1986). Some scientists have even described

success rates of 60% or higher as the expected "industry norm" for alcoholism treatment (e.g. Maltzman, 1987).

The pressure for evaluation poses a nagging quandary for programs, evaluators and marketers. Research to date points to much more modest overall success, with favorable long-term outcomes hovering around 30% in unselected populations (Costello, Biever & Baillargeon, 1977; Polich, Armor & Braiker, 1981; Helzer *et al.*, 1985). An evaluation which too rigidly clings to scientific standards may thus yield outcome data that compare unfavorably with the treatment industry's rosy picture of success.

Therefore, prudent program evaluators will take certain precautions when assessing the outcomes of treatment. Some studies indicate that variations in treatment programs contribute relatively little to ultimate outcome (McCance &

Correspondence to: William R. Miller, PhD, Department of Psychology, University of New Mexico, Albuquerque, NM 87131-1161 USA.

McCance, 1969; Moos, Finney & Cronkite, 1990), and even that the course of recovery is similar for treated versus untreated cases (Vail-lant, 1983). About half of controlled comparisons have failed to find significant differences among alternative treatment approaches (Miller *et al.*, 1995), and many studies to date have found surprisingly little overall benefit from increasing the length, cost or intensity of treatment (US Congress Office of Technology Assessment, 1983; Annis, 1985; Miller & Hester, 1986; Holder *et al.*, 1991). Uncontrolled outcome studies, however, offer much greater flexibility in success rates, with reports ranging from nearly 0% to 100% (Miller & Hester, 1980).

There are lessons to be learned here. If treatment programs contribute relatively little variation, then what *does* account for the wide differences in success rates found in uncontrolled studies? A careful comparison of programs with higher versus lower rates of reported success suggests a number of methodological subtleties that can be harnessed by program evaluators to enhance outcomes reports. This article outlines briefly a set of guidelines for evaluators faced with the need to document high success rates from routine treatment programs.

(1) Choose your cases carefully

It has been shown consistently that the life-style and characteristics of clients before and after treatment are strong predictors of outcome status, often overshadowing differences among treatment approaches (Moos *et al.*, 1990). Prudent evaluators will therefore use this knowledge to include only clients having more favorable prognostic characteristics. With impressive consistency, higher success rates are shown by people who are more socially stable (i.e. have a job, home, family, high income) and whose problems (e.g. dependence, medical complications) are less severe. Such people are likely to show high success rates in whatever treatment is provided (McLellan *et al.*, 1983). Conversely, socially unstable and more severely dependent people with little to lose tend to do less well, and should be excluded whenever possible if a high success rate is desired. Another factor that can substantially boost success rates is persuasive social pressure on the client to change (e.g. physicians under threat of loss of license). Still another possibility is to screen clients for readiness to change,

excluding from your evaluation any who fall into what Prochaska & DiClemente (1984) have termed the "precontemplation" or "contemplation" stages.

Some programs are fortunate in having as their primary or sole clientele people with good prognoses. This is probably the case, for example, in employee assistance programs and in hospitals charging high fees, which inherently screen out the unemployed, socially unstable and more deteriorated cases. Evaluators of such programs can readily produce pleasing outcome reports by attending to only a few of the other points discussed herein.

All is not lost, however, for evaluators of less selective treatment programs. Several devices can be used to filter out cases likely to dampen success rates. Perhaps the most important is to include in the evaluated sample only those who comply with instructions, complete treatment and dutifully attend a protracted course of after-care. Patients who comply faithfully with treatment procedures, even placebo medications, show more favorable outcomes (e.g. Fuller *et al.*, 1986), as do those who complete the full course of treatment. As a corollary, carefully exclude from your study all drop-outs from treatment or follow-up. Also consider excluding those who have been treated previously. Another helpful approach is to use less aggressive follow-up strategies, because harder-to-reach cases typically tend to show poorer outcomes. A simple approach is to exclude from analysis those who do not respond to early follow-up attempts.

Still another possibility is to use available data on client-treatment matching to choose your cases for evaluation. For example, it is advisable when testing programs for early intervention, where clients are given the opportunity to choose between abstinence and moderate drinking as their goal, to recruit as many women as possible. Women tend to achieve better outcomes than do men in these programs, particularly when the involvement of therapists is kept to a minimum (Sanchez-Craig *et al.*, 1989; Sanchez-Craig, Spivak & Davila, 1991).

An example of what can be achieved by careful sample selection is found in a report by Wallace and his colleagues (1988). They admitted to their study only patients: (1) who "had restorative potential"; (2) whose "medical, psychiatric, detoxification, and self-care problems, if any, were sufficiently managed so that the patient

could participate fully in the rehabilitation program"; (3) who were married, living with their spouses and had no plans to separate, (4) who had the economic means to pay for an expensive residential treatment program; and (5) who completed at least 21 days of treatment. It was from this carefully chosen pool that study subjects were then "randomly selected". Combining cases reporting continuous abstinence, those with a single slip, and those currently but not continuously abstinent, they were able to report "that 72% of the located patients currently were abstinent at the time of follow-up" at 6 months. This illustrates in passing yet another helpful strategy: using as the denominator the number of *located* cases when calculating your success rate.

(2) Avoid long follow-up periods

One of the simplest and most important ways to increase the success rate of your program is to shun long-term follow-up. The expected "industry standard" rate of 67% success represents the average outcome for a broad range of treatment programs when follow-up is restricted to short intervals (Emrick, 1974). If one allows follow-up to extend for as long as 12 months, however, outcomes are much less rosy (Costello *et al.*, 1977). The principal reason for this, as all program evaluators should know, is the well-established relapse rate for addictive behaviors (Hunt, Barnett & Branch, 1971). Very long follow-ups of 10 years (Pendery, Maltzman & West, 1982) or even longer (Vaillant, 1983) have yielded the most dismal pictures regarding treatment outcome. Program evaluation therefore should be kept relatively close to the end of treatment, preferably 3 months or less, and under no circumstances extending beyond 6 months. An upper-bound estimate of the expected impact of this methodological adjustment is to increment success rate from Costello's 12-month average of 26% success to Emrick's short-term rate of 66%, an absolute gain of 40%. A more conservative estimate would be an increase of 30%, roughly doubling absolute success rate.

The obvious brevity of follow-up can be mitigated somewhat by reporting the assessment period as extending from the date of *intake* rather than from treatment termination or discharge. Suppose, for example, that a typical client comes for intake evaluation, actually begins treatment 2 weeks later, and remains in the program for 16

weeks. A 6-month follow-up at week 26, then, actually occurs only 2 months after treatment termination, well within the safer window of sobriety.

(3) Avoid control and comparison groups

If you wish to make a favorable impression with outcome results, it is advisable to avoid control or comparison groups in your evaluation. When a program is compared with alternative treatment methods, briefer intervention or no treatment at all, it is distressingly common for no differences to be found (Bien, Miller & Tonigan, 1993; Miller *et al.*, 1995). This spoils the impression made by an evaluation, and it is an embarrassment easily avoided. In the absence of any frame of reference, a 70% success rate may seem impressive. It detracts substantially if your success rate is merely comparable (or worse!) to those from other approaches. Consider the published examples of this common evaluator error, shown in Table 1. In these and other cases, a perfectly reasonable success rate was clouded by the inclusion of a group with which it could be compared. Annoying distractions of this kind can be avoided simply by studying only those clients receiving the treatment you wish to promote. If you must use a comparison group, choose treatment drop-outs.

(4) Choose your measures carefully

There is a two-edged sword here. There is value in keeping evaluation short and simple. Follow-up interviews that tarry too long, explore alcohol consumption in great detail or ask too many questions are likely to uncover conditions that cloud an otherwise sunny picture of successful outcome. An exemplar for assessing abstinence is the question, "You're not drinking, are you?" If greater specificity is required because you are submitting your work for peer review in a scientific journal, ask, "You haven't drunk this week, have you?" Needlessly detailed inquiry is likely to reveal isolated pockets of slips that result in costly reductions in your reported rates of continuous abstinence.

On the other hand, the use of too few outcome measures can prevent random error from working in your favor. This might be termed a "Type III" error. In experimental design, one commits a Type I error by incorrectly rejecting the null

Table 1. Examples of the disappointing effect of control groups

Study	Results for treatment group	Type of comparison group included	Results for comparison group
Ditman <i>et al.</i> (1967)	32% success rate (no new arrests) among clinic-treated offenders at 12 months	Probation alone, without treatment	44% success rate (no new arrests)
Fuller <i>et al.</i> (1986)	18.8% continuously abstinent for 1 year with disulfiram	Placebo medication	22.5% continuously abstinent for 1 year
Miller <i>et al.</i> (1981)	80% of therapist-treated cases improved at 6 months	Self-help manual and minimal therapist contact	87% of cases improved at 6 months
Sanchez-Craig <i>et al.</i> (1991)	71% of therapist-treated cases problem-free at 12 months	Self-help materials and brief therapist contact	71% of cases problem-free at 12 months

hypothesis, and a Type II error by incorrectly failing to reject the null hypothesis. A Type III error is committed when you give yourself insufficient opportunity to commit a Type I error. It is useful to include a large number of measures in your evaluation, because if enough variables are included in analyses it is likely that *something* will show an effect. In this way, an embarrassing failure to find sufficient effects on more obvious measures can be offset by emphasizing those variables where changes did occur. Chick and his colleagues (1988), for example, failed to find expected differences on drinking measures between a group receiving only one session of advice and those given extended inpatient or outpatient treatment. They were still able to issue a positive report for treatment, however, because they had included a measure of alcohol-related problems on which a significant difference was found.

Another approach for reducing Type III error is to break your follow-up period into segments, conducting separate analyses for each period. This is illustrated by Keso & Salaspuro (1990), who compared a Minnesota-model inpatient program with traditional Finnish treatment not emphasizing Alcoholics Anonymous. They failed to find differences at any follow-up point on alcohol consumption measures, abstinence

survival curves or serum tests indicative of drinking. By breaking the follow-up year into three segments, however, they were able to find a single period for which a χ^2 test reflected a significantly ($p < 0.05$) higher percentage of cases abstinent for 4 months.

It is noteworthy that in both of the studies cited here the evaluators could have completely avoided the problem simply by omitting their comparison groups. As noted above, this removes the necessity of searching for significant differences. When comparisons are necessary, there is merit in amassing as large a sample as possible so that even relatively small changes when found become statistically significant.

There is an important pitfall to avoid here in methodologically orientated reports. Be careful not to use the so-called Bonferroni correction, whereby the level at which "significance" is inferred (usually an α level set at $p < 0.05$) becomes more conservative, dividing α by the number of tests run. Like many others, Keso & Salaspuro could not have reported their significant effects had they applied this statistical nuisance. Fortunately, most journal reviewers seem not to notice the absence of a Bonferroni correction in the case of multiple dependent measures, or at least overlook it in hopes that reviewers of their own articles will practice simi-

lar grace. Alternatively, to avoid wasting readers' and reviewers' time, one can simply report only those variables on which significant differences were found, and omit mention of other measures included in the evaluation.

(5) Focus only on alcohol

Considerable discretion should be observed in asking about the use of other drugs of abuse. Some clients very neatly resolve alcohol problems while persisting in at least occasional use of other drugs. The requirement of abstinence from all drugs can disqualify many cases who otherwise might be classified as successes. If inquiry about other drugs is mandatory keep it short and simple, and by all means disregard legal drugs such as tobacco, caffeine and all over-the-counter medications.

(6) Use liberal outcome definitions

When distinguishing between successful and unsuccessful cases it is wise to be generous in your criteria for success. As noted above, it has been customary for treatment programs to quote success rates in excess of 80% as long as no data were required to substantiate the claim. In order for actual evaluation data to approximate such rates, even if the above methodological advice is followed, forgiving criteria for "success" must generally be used. The use of conservative criteria, such as continuous total abstention from all psychoactive substances, will almost always yield depressingly low success rates.

Where abstinence is the standard of success, avoid perfectionistic definitions. An approach that includes "mostly" or "essentially" abstinent cases will enhance your success rate. One time-tested procedure is to allow "abstainers" a certain number of slips or occasions of moderate drinking, since these are quite common (e.g. Vaillant, 1983). This permits one to classify cases as abstinent without overtly acknowledging occasional periods of moderate or even immoderate drinking.

Under most circumstances, however, there is every reason to include within your success rate not only abstainers but also those cases who show at least some improvement. This can be accomplished without opening the Pandora's box of "controlled drinking", simply by describing such cases as "improved". The widely recog-

nized "rule of thirds" applies to alcoholism treatment outcomes at short-term follow-up: about a third are abstinent, a third are improved, and a third are unimproved (Emrick, 1974). The implication is obvious: success rates can be doubled by including improved cases, either as a separate category or by broadening the definition of abstinence to include them. In one report, for example, the success rate (again, calculated with located cases as the denominator) varied depending on the definition: 23% with total abstinence only, 37% when problem-free drinking was included and 60% when the definition was expanded to include improvement with some continued impairment (Miller *et al.*, 1992). Similarly, in a 4-year follow-up Polich and his colleagues (1981) found that 21% of treated cases had been abstinent for at least 12 months. This success rate increased to 28% when abstinence of 6 months or longer was counted, to 34% when moderate and problem-free drinkers were included, and to 44% when all problem-free drinkers were counted. This study also demonstrates the wisdom of using a short follow-up window: if abstinence for the past 30 days had been used as the criterion 45% would have been successes, and adding these to the above categories raises the overall success rate to 61%. The most obvious blunder in both these studies was the long period of follow-up.

Another way to enhance success rates is to avoid harsh quantitative data altogether and instead base your success judgements on clients' self-ratings, or better still on their therapists' ratings. Clients tend to describe themselves in a favorable light when using global ratings, perhaps to spare their therapists' feelings. Therapists judging their own clients' outcomes are also understandably disposed to rate them positively.

(7) Rely solely upon self-report

An ancillary principle is that whereas clients and therapists are fairly accommodating in judging their own outcomes, other sources may not be. The evaluation of alcoholism treatment outcome has customarily relied upon the self-reports of those treated. Recent trends in research, however, have made it fashionable to check on the validity of self-report by obtaining biomedical verification or interviewing significant others. This is costly to the evaluator, not only in economic terms but in its annoying effect on success

rates. Fuller and his colleagues (1986), for example, excluded a case from their "abstinent" category if any drinking was reported at any of seven follow-up points by patient self-report, in interviews with friends and family, or was detected by blood or urine samples. This resulted in a 1-year abstinence rate of 19%. Obviously, any verification of self-report poses a serious risk for deflating your success rate.

(8) Always declare victory

The political concept of positive "spin" is not without merit for treatment evaluators. Even seeming disasters can be turned into victories when a sharp-witted evaluator writes the discussion section of the report. Declare victory no matter how bad the outcome. Suppose, for example, that the worst occurred. Suppose that an evaluator made many of the errors against which we have warned, and consequently found the exact *opposite* of what was wanted and predicted. Woititz (1976), for example, included a comparison group, predicting that children of alcoholics attending Alateen would have higher self-esteem than those not involved in this program. Finding exactly the opposite, she concluded that "Thoughtful analysis of the data and an understanding of the alcoholic family pattern can help explain this result. Denial is a part of the disease both for the alcoholic and his family.... This researcher suggests that the non-Alateen group scores significantly higher than the Alateen group scores because the non-Alateen children are still in the process of denial" (pp. 53-55). Such interpretive saves are particularly useful when written into the executive summary, the abstract and the conclusions of a report, inasmuch as these are the only sections likely to be scanned by most readers.

Conclusions

Evaluators of alcohol treatment programs have a difficult job. From past decades during which no data were required to substantiate outcome claims, the field came to expect success rates of 67%, 80% or even higher as "standard". Such rates are difficult to produce without considerable creativity and caution on the part of the program evaluator. It is quite possible to double, triple or even quadruple success rates, given sufficient forethought and methodological pru-

dence. No program need suffer from low success rates. The ideal program evaluation should study only the treatment of interest as applied to highly compliant good-prognosis clients, reporting short-term outcome for easily located cases on carefully selected self-report measures with liberal definitions of success. Programs following these few simple principles have reported impressive track records. By taking the above-described steps, programs can join the ranks where "all have won and all must have prizes". The essence, in fact, was captured decades ago in a song lyric of Johnny Mercer:

You've got to accentuate the positive,
Eliminate the negative,
Latch on to the affirmative,
Don't mess with Mister In-between.

References

- ANNIS, H. M. (1985) Is inpatient rehabilitation of the alcoholic cost effective? Con position, *Advances in Alcohol and Substance Abuse*, 5, 175-190.
- BEN, T. H., MILLER, W. R. & TONIGAN, J. S. (1993) Brief interventions for alcohol problems: a review, *Addiction*, 88, 315-336.
- CHICK, J., RITSON, B., CONNAUGHTON, J., STEWART, A. & CHICK, J. (1988) Advice versus treatment for alcoholism: a controlled trial, *British Journal of Addiction*, 83, 159-170.
- COSTELLO, R. M., BIEVER, P. & BAILLARGEON, J. G. (1977) Alcoholism treatment programming: historical trends and modern approaches, *Alcoholism: Clinical and Experimental Research*, 1, 311-318.
- DITTMAN, K. S., CRAWFORD, G. G., FORGY, E. W., MOSKOWITZ, H. & MACANDREW, C. (1967) A controlled experiment on the use of court probation for drunk arrests, *American Journal of Psychiatry*, 124, 160-163.
- EMRICK, C. D. (1974) A review of psychologically-oriented treatment of alcoholism. I. The use and interrelationships of outcome criteria and drinking behavior following treatment, *Quarterly Journal of Studies on Alcohol*, 35, 523-549.
- FULLER, R. K., BRANCHY, L., BRIGHTWELL, D. R. et al. (1986) Disulfiram treatment of alcoholism: a Veterans Administration cooperative study, *Journal of the American Medical Association*, 256, 1449-1455.
- HELZER, J. E., ROBINS, L. N., TAYLOR, J. R. et al. (1985) The extent of long-term moderate drinking among alcoholics discharged from medical and psychiatric treatment facilities, *New England Journal of Medicine*, 312, 1678-1682.
- HOLDER, H., LONGABAUGH, R., MILLER, W. R. & RUBONIS, A. V. (1991) The cost-effectiveness of treatment for alcoholism: a first approximation, *Journal of Studies on Alcohol*, 52, 517-540.
- HUNT, W. A., BARNETT, L. W. & BRANCH, L. G. (1971) Relapse rates in addiction programs, *Journal of Clinical Psychology*, 27, 455-456.

- KESO, L. & SALASPURO, M. (1990) Inpatient treatment of employed alcoholics: a randomized clinical trial on Hazelden-type and traditional treatment, *Alcoholism: Clinical and Experimental Research*, 14, 584-589.
- MALTZMAN, I. (1987) Controlled drinking and the treatment of alcoholism, *Journal of the American Medical Association*, 257, 927.
- MCCANCE, C. & MCCANCE, P. F. (1969) Alcoholism in Northeast Scotland: its treatment and outcome, *British Journal of Psychiatry*, 115, 189-198.
- MCLELLAN, A. T., LUBORSKY, L., WOODY, G. E., O'BRIEN, C. P. & DRULEY, K. A. (1983) Predicting response to alcohol and drug abuse treatments, *Archives of General Psychiatry*, 40, 620-625.
- MILLER, W. R. & HESTER, R. K. (1980) Treating the problem drinker: modern approaches, in: MILLER, W. R. (Ed.) *The Addictive Behaviors: treatment of alcoholism, drug abuse, smoking and obesity*, pp. 11-141 (Oxford, Pergamon Press).
- MILLER, W. R. & HESTER, R. K. (1986) Inpatient alcoholism treatment: who benefits? *American Psychologist*, 41, 794-805.
- MILLER, W. R., GRIBSKOV, C. J. & MORTELL, R. L. (1981) Effectiveness of a self-control manual for problem drinkers with and without therapist contact, *International Journal of the Addictions*, 16, 1247-1254.
- MILLER, W. R., BROWN, J. M., SIMPSON, T. L. et al. (1995) What works? A methodological analysis of the alcohol treatment outcome literature, in: HESTER, R. K. & MILLER W. R. (Eds) *Handbook of Alcoholism Treatment Approaches: effective alternatives*, 2nd edn, pp. 12-44 (New York, Allyn & Bacon).
- MILLER, W. R., LECKMAN, A. L., DELANEY, H. D. & TINKCOM, M. (1992) Long-term follow-up of behavioral self-control training, *Journal of Studies on Alcohol*, 53, 249-261.
- MOOS, R. H., FINNEY, J. W. & CRONKITE, R. C. (1990) *Alcoholism Treatment: context, process, and outcome* (New York, Oxford University Press).
- PENDERY, M. L., MALTZMAN, I. M. & WEST, L. J. (1982) Controlled drinking by alcoholics? New findings and a reevaluation of a major affirmative study, *Science*, 217, 169-175.
- POLICH, J. M., ARMOR, D. J. & BRAIKER, H. B. (1981) *The Course of Alcoholism: four years after treatment* (New York, Wiley Interscience).
- PROCHASKA, J. O. & DiCLEMENTE, C. C. (1984) *The Transtheoretical Approach: crossing traditional boundaries of therapy* (Homewood, IL, Dow Jones/Irwin).
- SANCHEZ-CRAIG, M., LEIGH, G., SPIVAK, K. & KEI, H. (1989) Superior outcome of females over males after brief treatment for the reduction of heavy drinking, *British Journal of Addiction*, 84, 395-440.
- SANCHEZ-CRAIG, M., SPIVAK, K. & DAVILA, R. (1991) Superior outcome of females over males after brief treatment for the reduction of heavy drinking: replication and report of therapist effects, *British Journal of Addiction*, 86, 867-876.
- VAILLANT, G. E. (1983) *The Natural History of Alcoholism* (Cambridge, MA, Harvard University Press).
- US CONGRESS, OFFICE OF TECHNOLOGY ASSESSMENT (1983) *The Effectiveness and Costs of Alcoholism Treatment* (Washington, DC, Author).
- WALLACE, J., McNEILL, D., GILFILLAN, D., MACLEAN, K. & FANELLA, F. (1988) Six-month treatment outcomes in socially stable alcoholics: abstinence rates, *Journal of Substance Abuse Treatment*, 5, 247-252.
- WOITITZ, J. G. (1976) A study of self-esteem in children of alcoholics, Unpublished Doctoral dissertation (Rutgers University).

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.