

Use of Exclusion Criteria in Selecting Research Subjects and Its Effect on the Generalizability of Alcohol Treatment Outcome Studies

Keith Humphreys, Ph.D., and Constance Weisner, Dr.P.H.

Objective: Researchers have not systematically examined how exclusion criteria used in selection of research subjects affect the generalizability of treatment outcome research. This study evaluated the use of exclusion criteria in alcohol treatment outcome research and its effects on the comparability of research subjects with real-world individuals seeking alcohol treatment. **Method:** Eight of the most common exclusion criteria described in the alcohol treatment research literature were operationalized and applied to large, representative clinical patient samples from the public and private sectors to determine whether the hypothetical research samples differed substantially from real-world samples. Five hundred ninety-three consecutive individuals seeking alcohol treatment at one of eight treatment programs participated. A trained research technician gathered information from participants on demographic variables and on alcohol, drug, and psychiatric problems as measured by the Addiction Severity Index. **Results:** Large proportions of potential research subjects were excluded under most of the criteria tested. The overall pattern of results showed that African Americans, low-income individuals, and individuals who had more severe alcohol, drug, and psychiatric problems were disproportionately excluded under most criteria. **Conclusions:** Exclusion criteria can result in alcohol treatment outcome research samples that are more heavily composed of white, economically stable, and higher-functioning individuals than are real-world samples of substance abuse patients seen in clinical practice, potentially compromising the generalizability of results. For both scientific and ethical reasons, in addition to studies that use exclusion criteria, outcome research that uses no or minimal exclusion criteria should be conducted so that alcohol treatment outcome research can be better generalized to vulnerable populations.

(*Am J Psychiatry* 2000; 157:588–594)

Most clinical research studies in psychiatry and other areas of medicine exclude some potential subjects on the basis of predefined criteria. Use of exclu-

sion criteria can optimize internal validity, make a study more feasible (e.g., by excluding patients who are nonadherent to treatment) (1), reduce cost (e.g., by excluding patients who would be difficult to follow up), and serve ethical functions (e.g., by excluding patients who might be harmed by study participation). At the same time, exclusion criteria can have negative implications for the generalizability (or “external validity”) of results to real-world practice settings. Because by definition patients meeting a study’s exclusion criteria differ from those not excluded, the results of the study cannot be assumed to apply to excluded patients. Further, clinicians may be less likely to administer an intervention if the patients they typically treat differ from the subset of patients who met predefined criteria in research that supported the intervention’s efficacy. Such concerns were a key reason for development of the National Institutes of Health guidelines encourag-

Received Dec. 14, 1998; revisions received July 8 and Sept. 7, 1999; accepted Sept. 28, 1999. From the Center for Health Care Evaluation, VA Palo Alto Health Care System; the Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, Calif.; the Alcohol Research Group, Public Health Institute, Berkeley, Calif.; and the School of Public Health, University of California at Berkeley. Address reprint requests to Dr. Humphreys, Center for Health Care Evaluation, VA Palo Alto Health Care System (152-MPD), 795 Willow Rd., Menlo Park, CA 94025; keithh@odd.stanford.edu (e-mail).

This research was supported by National Institute on Alcohol Abuse and Alcoholism grant AA-09750 and by the VA Mental Health Strategic Health Group.

The authors thank John Finney, Ph.D., Lee Ann Kaskutas, Dr.P.H., Rachel Korch, M.A., and Mija Lee, M.A., for assisting with data collection and analysis, and Rudolf Moos, Ph.D., and Andrew Winzberg, Ph.D., for commenting on an earlier draft of the manuscript.

ing inclusion of women and members of ethnic minority groups as subjects in clinical research (2); the guidelines are intended to ensure that both the burden and the benefits of treatment research are fairly distributed throughout society.

The implications of exclusion criteria for the external validity and clinical utility of findings in various areas of research have not been systematically examined. The study reported here assessed whether exclusion criteria affect the generalizability of research findings in one important area of psychiatric research—treatment for alcohol problems. On October 1, 1997, approximately 625,000 individuals were receiving specialty alcohol treatment services in the United States (3). The large number of treated individuals underscores the importance of clinical alcohol research's being relevant to real-world practice.

Meta-analyses of alcohol treatment research by Finney and Monahan (4, 5) have indicated that 74.9% of outcome studies (254 of 339 studies conducted between 1980–1992) reported using exclusion criteria. This proportion is a conservative estimate because researchers do not always report the use of exclusion criteria; they may believe that exclusion procedures “go without saying” and do not need to be described, or they may be unaware that exclusion criteria have been implemented (e.g., when treatment staff surreptitiously discourage more severely disabled patients from participating in a clinical trial) (6). The nine most common exclusion criteria used in the alcohol treatment studies reviewed by Finney and Monahan were psychiatric/emotional problems, noncompliance/lack of motivation, serious medical problems, neurological impairment (e.g., organic brain syndrome), drug abuse problems, lack of success in prior alcohol treatment, residence far from the treatment facility, social instability (e.g., unmarried, unemployed), and residential instability.

To examine the effect of exclusion criteria on the generalizability of findings from alcohol treatment research, the present study operationalized eight of these commonly used exclusion criteria and then applied them to real-world samples of patients seeking treatment at representative alcohol programs in the public and the private sector. By comparing patients who were excluded or not excluded under each rule, we estimated how exclusion criteria changed the samples in question, in other words, the extent to which the criteria produced research samples that differed from real-world clinical samples.

METHOD

Patients

Participants were individuals (N=593) seeking treatment at one of eight alcohol treatment programs that were representative of public and private-for-profit programs in a northern California county (7). In terms of the demographic characteristics of the population and the treatment services available, this county is representative of

many counties in the United States. To be considered eligible for participation, patients had to be at least 18 years old and able to complete a structured, in-person interview (i.e., non-English-speaking patients and patients suffering from cognitive impairment or delirium were excluded). During the period of fieldwork, 690 of 739 consecutive admissions (93.4%) met inclusion criteria for the study. Of the 338 study-eligible individuals admitted to the public programs, 298 (88.2%) were asked to participate and consented. Of the 352 study-eligible individuals admitted to the private programs, 295 (83.8%) were asked to participate and consented. A chi-square analysis indicated that this difference in participation rate between public and private systems was not significant ($\chi^2=2.71$, $df=1$, $p>0.05$). Of the 97 eligible patients who did not agree to participate in the study, 57 (58.8%) were non-Hispanic Caucasian and 62 (63.9%) were male. These proportions are similar to those of the study sample, as described below.

Men (N=391) composed slightly under two-thirds of the sample (65.9%). At intake, 211 (35.6%) participants were separated or divorced, and 167 (28.2%) were married or living in a marriage-like relationship. About two-thirds of the sample (N=380, 64.1%) were unemployed. The most common racial/ethnic backgrounds were non-Hispanic Caucasian (51.9%, N=308) and African American (34.2%, N=203). The mean age of the participants was 39.0 years (SD=10.5).

Programs

Alcohol treatment programs were selected on the basis of size (more than one admission per week) and primary funding source. The six publicly funded programs included two detoxification units, two residential facilities, and two outpatient clinics. At the time of the study, public county substance abuse services in California were mainly funded by federal block grants and matching funds from the state and its counties. Almost all patients in the public programs either had no insurance or had publicly funded insurance such as Medicaid. The two private programs were large, for-profit, hospital-based units that offered both inpatient and outpatient services and were primarily funded by fee-for-service insurance. These programs drew clients from all parts of the county and were representative of private programs in the county.

Procedure

Trained research staff independent from the treatment programs recruited participants on-site. Research staff described the study to eligible patients and made it clear that the decision of whether to participate in the study would have no impact on treatment services they would receive. After the study was fully explained, written, informed consent was obtained from the 593 patients who decided to participate. For each completed interview, \$20 was paid either to the participant or to the residents' fund at the treatment program.

The widely used Addiction Severity Index (8) served as the core assessment instrument. The Addiction Severity Index produces continuous composite scores (range=0–1) for alcohol, drug, and psychiatric problems. It has excellent concurrent reliability and has been shown to have discriminant and concurrent validity across a variety of substance-abusing populations (9). The Addiction Severity Index was supplemented with questions from the Diagnostic Interview Schedule for Psychoactive Substance Dependence (10) that assessed the presence or absence of nine drug dependence symptoms (e.g., tolerance, withdrawal symptoms, disruption of important daily activities) in the past 30 days. Finally, participants were asked to report the dates and types of any prior alcohol and psychiatric treatment episodes.

Operationalization and Analysis of Exclusion Criteria

Using the data gathered, eight of the nine commonly used exclusion criteria were operationalized: psychiatric/emotional problems, noncompliance/lack of motivation, medical problems, drug dependence, lack of success in prior alcohol treatment, residence distant from the treatment facility, social instability, and residential instability (table 1). The effect of an exclusion criterion for neurological impairment on the characteristics of samples could not be evaluated

TABLE 1. Operationalization of Exclusion Criteria Widely Used in Clinical Research on Alcohol Treatment

Criterion	Operationalization
Psychiatric/emotional problems	Any of the following: inpatient psychiatric treatment in the past year, current treatment with psychotropic medication, hallucinations not directly attributable to substance use in the past 30 days, serious thoughts of suicide not directly attributable to substance use in the past 30 days
Noncompliance/lack of motivation	Patient rating of treatment for current substance abuse problem less than "considerably important" on a 5-point Addiction Severity Index item
Medical problems	Addiction Severity Index medical composite problem score two or more standard deviations higher than the norm for the general population (11)
Drug dependence	Three or more NIMH Diagnostic Interview Schedule (10) drug dependence symptoms in past 30 days
Unsuccessful prior alcohol treatment	Treatment for substance abuse prior to current treatment episode
Residence distant from treatment facility	Residence outside of the county in which the treatment program was located
Social instability	Both unmarried and unemployed
Residential instability	Not living in a house or apartment at intake

here because neurological impairment was an exclusion criterion in the present study. Operationalizations were chosen to be the same or similar to widely employed operationalizations in the literature, within the constraints of the data of the present study.

For each dependent variable, the eight exclusion rules were applied individually within the group of public sector subjects, and then within the group of private sector subjects. Thus all tests and comparisons were made separately within service systems rather than across them. For each exclusion rule, the proportion of patients excluded was calculated, and then the characteristics of the excluded and the included patients were compared. For categorical dependent variables (race, sex, and income), statistical significance was assessed using a chi-square test. For continuous variables (drug, alcohol, and psychiatric problems), statistical significance was assessed with independent samples *t* tests. For chi-square analyses, a threshold of $p < 0.01$ was used to judge significance, whereas for the more statistically powerful *t* test analyses, a threshold of $p < 0.005$ was employed. In practical terms, statistically significant results in this study indicate that when a particular exclusion criterion is applied in a real-world treatment system, patients who are excluded from the resulting hypothetical research sample are significantly different on the variable of interest (e.g., race, sex) than patients who would be included in the research sample.

RESULTS

The proportion of patients excluded under each criterion was as follows: noncompliance/lack of motivation (10.7% [N=32] of public sector patients; 7.5% [N=22] of private sector patients), residential instability (44.0% [N=131], public sector; 16.9% [N=50], private sector), medical problems (22.5% [N=67], public; 39.3% [N=116], private), residence distant from treatment (19.8% [N=59], public; 50.8% [N=

150], private), psychiatric/emotional problems (34.6% [N=103], public; 50.5% [N=149], private), drug dependence (55.4% [N=165], public; 55.3% [N=163], private), social instability (72.5% [N=216], public; 40.7% [N=120], private), and unsuccessful prior alcohol treatment (74.8% [N=223], public; 57.6% [N=170], private). Overall, large proportions of patients in both systems were excluded under most rules. Thus, in a treatment outcome study conducted using these exclusion criteria, many or most of these treated patients would not be eligible to participate.

Race

Of the 298 patients treated in public programs, 117 (39.3%) were white, 137 (46.0%) were African American, and 44 (14.8%) had other racial/ethnic backgrounds. For six of the exclusion criteria, excluded and included patients did not differ on race. However, under the drug dependence criterion, a significantly higher proportion of African Americans were in the excluded group ($\chi^2=38.16$, $df=2$, $p<0.0001$). Specifically, in the excluded group (N=165), 25.5% of the patients (N=42) were white and 61.8% (N=102) were black, whereas in the included group (N=133), 56.4% of the patients (N=75) were white and 26.3% (N=35) were black. In addition, under the exclusion criterion of distance from treatment, the tendency for blacks to be disproportionately excluded approached significance ($\chi^2=6.45$, $df=2$, $p=0.04$): 59.3% (N=35) of the 59 excluded patients were black versus 42.7% (N=102) of the 239 included patients.

Of the 295 patients treated in the private sector programs, 191 (64.7%) were white, 66 (22.4%) were black, and 38 (12.9%) had other racial/ethnic backgrounds. In the private sector group, exclusion criteria also disproportionately affected black patients. There were significant differences in the racial characteristics of excluded and included patients for the psychiatric/emotional problem criterion (44 of 149 excluded patients [29.5%] were black versus 22 of 145 included patients [15.2%]; $\chi^2=10.25$, $df=2$, $p<0.01$), the drug dependence criterion (50 of 163 excluded patients [30.7%] were black versus 16 of 131 included patients [12.2%]; $\chi^2=15.50$, $df=2$, $p<0.001$), the social instability criterion (42 of 120 excluded patients [35.0%] were black versus 24 of 174 included patients [13.8%]; $\chi^2=18.48$, $df=2$, $p<0.0001$), and the residential instability criterion (18 of 50 excluded patients [36.0%] were black versus 48 of 245 included patients [19.6%]; $\chi^2=8.54$, $df=2$, $p<0.01$).

Overall, in both service systems, but particularly in the private sector, the application of exclusion criteria tended to reduce the proportion of African Americans (and, obviously, increase the proportion of Caucasians) in hypothetical research samples. In other words, exclusion criteria often created treatment research samples that had different racial characteristics than do real-world treatment samples.

TABLE 2. Mean Composite Drug, Alcohol, and Psychiatric Problem Scores of Patients in Public Sector and Private Sector Alcohol Treatment Programs Who Were Excluded From or Included in Hypothetical Research Samples by Using Common Exclusion Criteria^a

Exclusion Criterion and Patient Group	Public Sector							Private Sector						
	N	Drug		Alcohol		Psychiatric		N	Drug		Alcohol		Psychiatric	
		Mean	SD	Mean	SD	Mean	SD		Mean	SD	Mean	SD	Mean	SD
Total sample	298	0.13	0.13	0.34	0.32	0.37	0.24	295	0.15	0.14	0.43	0.34	0.43	0.24
Psychiatric/emotional problems														
Excluded patients	103	0.16	0.13	0.41*	0.32	0.56*	0.19	149	0.18*	0.15	0.44	0.33	0.56*	0.20
Included patients	195	0.12	0.12	0.30	0.31	0.27	0.20	145	0.12	0.13	0.42	0.35	0.28	0.19
Noncompliance/lack of motivation														
Excluded patients	32	0.11	0.10	0.37	0.23	0.34	0.20	22	0.15	0.11	0.43	0.31	0.49	0.23
Included patients	266	0.14	0.13	0.33	0.33	0.38	0.24	273	0.15	0.14	0.43	0.34	0.42	0.24
Medical problems														
Excluded patients	67	0.16	0.14	0.48*	0.30	0.46*	0.25	116	0.15	0.15	0.48	0.35	0.49*	0.22
Included patients	231	0.12	0.12	0.29	0.31	0.34	0.23	179	0.14	0.13	0.39	0.33	0.38	0.24
Drug dependence														
Excluded patients	165	0.21*	0.11	0.34	0.31	0.43*	0.23	163	0.24*	0.11	0.29*	0.33	0.47*	0.24
Included patients	133	0.03	0.05	0.33	0.33	0.30	0.24	131	0.00	0.01	0.59	0.28	0.38	0.24
Unsuccessful prior alcohol treatment														
Excluded patients	223	0.14	0.13	0.36	0.33	0.38	0.24	170	0.16	0.14	0.46	0.34	0.46*	0.23
Included patients	75	0.12	0.11	0.28	0.27	0.35	0.24	125	0.13	0.14	0.38	0.34	0.37	0.25
Residence distant from treatment facility														
Excluded patients	59	0.14	0.12	0.47	0.30	0.38	0.22	150	0.15	0.15	0.43	0.33	0.44	0.23
Included patients	239	0.13	0.13	0.30	0.31	0.37	0.25	145	0.14	0.13	0.43	0.35	0.41	0.25
Social instability														
Excluded patients	216	0.13	0.13	0.34	0.32	0.37	0.24	120	0.16*	0.14	0.42	0.34	0.45*	0.24
Included patients	82	0.13	0.13	0.21	0.25	0.50	0.28	174	0.01	0.11	0.48	0.35	0.34	0.23
Residential instability														
Excluded patients	131	0.13	0.13	0.34	0.31	0.37	0.22	50	0.17	0.14	0.39	0.32	0.47	0.22
Included patients	167	0.13	0.12	0.34	0.32	0.38	0.26	245	0.14	0.14	0.44	0.35	0.42	0.24

^a Drug, alcohol, and psychiatric problems assessed by using the Addiction Severity Index (8). Comparisons of excluded and included patients under each criterion were conducted within each system (public and private) by using t tests (df=296 for public program comparisons and df=293 for private program comparisons). Total N=294 for some analyses in the private sector because of one subject with some missing data.

*p<0.005.

Sex

All but one exclusion criteria had no significant effect on the sex ratio of samples in either service system. The sole exception was the residential instability criterion, under which the proportion of males in public programs who were excluded (105 of 131 patients, 80.2%) was significantly higher than those who were included (101 of 167 patients, 60.5%) ($\chi^2=13.31$, df=1, $p<0.0001$).

Income

In the public sector, 289 of the 298 patients (97.0%) provided data on annual income. Of these, 184 (63.7%) had an annual income of less than \$10,000, 69 (23.9%) had an income between \$10,000 and \$35,000, and 36 (12.5%) had an income of \$35,000 or more. Of the 213 patients excluded under the social instability criteria, 157 (73.7%) had less than \$10,000 income, 40 (18.8%) had an income between \$10,000 and \$35,000, and 16 (7.5%) had an income greater than \$35,000. In contrast, the 76 patients included under this criteria had significantly higher incomes: 27 (35.5%) had less than \$10,000, 29 (38.2%) had between \$10,000–\$35,000, and 20 (26.3%) had greater than \$35,000 ($\chi^2=37.54$, df=2, $p<0.0001$). Similarly, under the residential instability criterion, the 126 ex-

cluded patients had significantly lower incomes ($\chi^2=8.91$, df=2, $p<0.01$) than the 163 included patients (e.g., 73.0% of excluded patients [N=92] had less than \$10,000 income versus 56.4% of included patients [N=92]). Under the criterion of noncompliance/lack of motivation, the tendency for excluded patients (N=32, 11.1%) to have lower incomes than included patients (N=257, 88.9%), approached but did not attain significance ($\chi^2=6.72$, df=2, $p=0.03$).

In the private sector, 289 of the 295 patients provided income data. Of these, 83 (28.7%) had less than \$10,000 annual income, 84 (29.1%) had between \$10,000–\$35,000 income, and 122 (42.2%) had an income of over \$35,000. In general, the eight exclusion criteria disproportionately excluded low-income patients. The pattern of those results approaching or attaining significance can most easily be summarized by reporting the proportion of excluded and included patients under each criterion that lived in extreme poverty (i.e., had less than \$10,000 annual income): psychiatric/emotional problems (66 of 149 excluded patients [44.3%] lived in poverty versus 17 of 140 included patients [12.1%]; $\chi^2=37.25$, df=2, $p<0.0001$), medical problems (37 of 115 excluded patients [32.2%] versus 47 of 175 included patients [26.9%]; $\chi^2=6.59$, df=2, $p=0.04$), drug dependence (57 of 160 excluded patients [35.6%] versus 26 of 128 in-

cluded patients [20.3%]; $\chi^2=10.00$, $df=2$, $p<0.01$), unsuccessful prior alcohol treatment (65 of 167 excluded patients [38.9%] versus 18 of 122 included patients [14.8%]; $\chi^2=25.94$, $df=2$, $p<0.0001$), social instability (64 of 117 excluded patients [54.7%] versus 19 of 171 included patients [11.1%]; not a significant difference), and residential instability (26 of 50 excluded patients [52.0%] versus 57 of 239 included patients [23.8%]; not a significant difference).

To summarize, in both service systems and particularly in private programs, exclusion criteria significantly reduced the proportion of low-income patients who would have been eligible to participate in a hypothetical research study.

Drug, Alcohol, and Psychiatric Problems

Differences in the frequency of drug, alcohol, and psychiatric problems (as assessed by the Addiction Severity Index) among excluded and included patients under each exclusion criterion are presented in table 2. In public programs, patients excluded ($N=103$) under the psychiatric/emotional problems criterion had significantly higher scores for alcohol and psychiatric problems than included patients ($N=195$). Patients excluded under the medical problems criterion ($N=67$) had significantly higher scores for alcohol and psychiatric problems than included patients ($N=231$). Finally, patients excluded under the drug dependence criterion ($N=165$) had significantly higher scores for drug and psychiatric problems than patients included under this criterion ($N=133$).

Parallel effects were evident within the private sector. Excluded patients had higher scores for drug problems than included patients under the psychiatric/emotional problems, drug dependence, and social instability criteria. Similarly, excluded patients had higher scores for psychiatric problems than included patients under the psychiatric/emotional problems, medical problems, drug dependence, unsuccessful prior alcohol treatment, and social instability criteria. The only exception to the overall trend for exclusion criteria to differently affect more severely disabled patients was for the drug dependence criterion. When this criterion was applied in private programs, included patients had more severe alcohol-related problems than excluded patients.

Overall, the effects identified are of considerable magnitude, with the problem severity of excluded patients exceeding that of included patients by three or more standard deviations in some cases. The findings underscore the difficulty of excluding patient problems in only a single domain. For example, although it is not surprising that an exclusion criterion for psychiatric/emotional problems would reduce the frequency of psychiatric problems in a sample, the extent to which the same criterion can significantly alter the frequency of drug and alcohol problems of the samples is striking.

Sensitivity Analyses

A variety of alternative operationalizations were chosen for the exclusion rules to determine if the pattern of findings was significantly affected. These operationalizations ranged from substantially more restrictive (e.g., requiring that three of the four psychiatric problem indicators be present for the psychiatric/emotional problems criterion, requiring multiple prior treatments for the unsuccessful prior alcohol treatment criterion) to substantially more liberal (e.g., excluding all patients who had used drugs in the past 30 days, even if they did not have a diagnosis of drug dependence) than those applied here. These results (not shown here but available from the first author) indicated that, as would be expected, more liberal operationalizations resulted in larger proportions of patients being excluded than did more conservative operationalizations. However, across operationalizations, the pattern of disproportionate exclusion of African American, low-income, and more severely disabled patients continued to hold.

DISCUSSION

The study reported here was stimulated by our observation that clinicians and administrators in alcohol treatment programs frequently speculate openly about whether the patients in their programs are similar to the research subjects who provided the data that is supposed to guide clinical intervention. This study operationalized common exclusion criteria and applied them to a real-world sample of patients seeking alcohol treatment. The results suggest that, in terms of sex ratios, samples of real-world patients are similar to samples of research subjects. However, the results indicate that exclusion criteria can produce research samples in which African American, low-income, and severely disabled patients are underrepresented. The differences between excluded and included patients were more pronounced in private programs because of the relative heterogeneity of the caseloads in these programs, which included both well-off patients with private insurance as well as poor, disabled patients with fee-for-service alcohol treatment coverage through Medicare. However, the differences were still of significant concern in public programs. The study results have yet to be replicated in other service systems and in other parts of the country, but given the size of the study samples, the representativeness of the county and the programs, and impressive size of the effects generated by the exclusion criteria, it would be unwise to ignore these results in the interim, particularly because race, socioeconomic status, and problem severity have been shown to influence alcohol treatment outcome (12).

One might argue that our operationalization of the exclusion criteria was too liberal and tended to overstate effects. However, this result seems unlikely, given that most of the criteria were quite strict (e.g., having

medical problems of a severity more than two standard deviations higher than the mean for the general population), and sensitivity analyses indicated that alternative operationalizations produced similar results. Further, the exclusion criteria were applied singly. Applying criteria singly has a smaller effect on research samples than applying multiple exclusion criteria, which is done in most studies of alcohol interventions. To take a recent, well-known example, Project MATCH (13), the largest randomized trial of alcohol treatment ever conducted, excluded potential participants who were unwilling to complete an extensive assessment battery, had used intravenous drugs in the past 6 months, were dependent on any drug other than marijuana, were currently a danger to themselves or others, had probation or parole requirements that would interfere with protocol participation, were residually unstable, could not identify a collateral contact who would assist in locating them at follow-up, were acutely psychotic, had severe organic impairment, or had planned or current involvement in treatment other than that provided in the study (as well as, of course, patients who were unwilling to be randomly assigned to treatment). As a result, more than 60% of the patients presenting for alcohol treatment were excluded from that study, which is unsurprising given the data presented here.

The higher likelihood that African Americans, low-income individuals, and more severely troubled patients will be excluded from clinical alcohol research can have significant scientific, clinical, and ethical consequences. From a scientific viewpoint, exclusion criteria can enhance internal validity and thereby facilitate evaluation of treatment "efficacy" (i.e., how well a treatment works under ideal conditions) (14). At the same time, exclusion criteria weaken our ability to assess treatment "effectiveness" (i.e., how well a treatment works in the real world of day-to-day clinical practice) (14). For example, African Americans constitute about one-fourth of substance abuse patients in the United States (3), so to the extent that substance abuse treatment research does not include them, it becomes less useful as a method for assessing treatment effectiveness (and for that matter, informing health care policy decisions that affect African Americans). Further, whether a study is primarily aimed at assessing treatment efficacy or treatment effectiveness, the results presented here suggest that some exclusion criteria can make a study more difficult to complete. Because some criteria exclude many potential research participants, fieldwork time may need to be extended until enough nonexcluded patients are admitted to the recruitment site.

Clinically, the differences produced by exclusion criteria between research and real-world samples of patients with alcohol problems may help explain the commonly described "research-practice gap." Many treatment providers work in alcohol programs where the majority of patients are low-income persons with minority racial/ethnic background who have comor-

bid psychiatric and drug problems; in short, the very type of patient who is particularly likely to be excluded from alcohol treatment research. Practitioners may be understandably wary about applying findings from samples that have been "creamed" through exclusion criteria to omit such disadvantaged and troubled patients.

As for ethical consequences, the medical research community has already reached consensus that African Americans and more severely troubled individuals have a legitimate expectation that publicly funded medical science will be relevant to them (2). We do not believe that the tendency of exclusion criteria to reduce the number of poor and African American individuals in treatment samples has been an outcome intended by researchers. We are sure that researchers are committed to having their work be useful in the treatment of vulnerable populations; thus we have pointed out how exclusion criteria can inadvertently subvert this goal.

Obviously, exclusion criteria are sometimes necessary. However, we concur with Wells (14) that the scientific process of evaluating interventions is not complete until studies with minimal or no exclusion criteria are conducted, such as is more often the case in the burgeoning field of health services research. Funding agencies can help treatment researchers conduct such studies by recognizing that loosening or eliminating exclusion criteria may in some cases raise the cost of conducting outcome research (14) (e.g., more funds may be needed to locate more severely troubled participants at follow-up). Because it seems unlikely that alcohol treatment research is the only area of research in which exclusion criteria may compromise generalizability, we urge other psychiatric and medical researchers to undertake systematic analysis of these issues in their own fields of intervention research.

REFERENCES

1. Pablos-Mendez A, Barr RG, Shea S: Run-in periods in randomized trials: implications for the application of results in clinical practice. *JAMA* 1998; 279:222-225
2. Department of Health and Human Services: National Institutes of Health Guidelines on the Inclusion of Women and Minorities in Clinical Research. *Federal Register* 1994; 59: 11146
3. Substance Abuse and Mental Health Services Administration: Uniform Facilities Dataset (UFDS) 1997. Washington, DC, Office of Applied Studies, Substance Abuse and Mental Health Services Administration, 1999
4. Finney JW, Monahan SC: The cost-effectiveness of treatment for alcoholism: a second approximation. *J Stud Alcohol* 1996; 57:229-243
5. Monahan SC, Finney JW: Explaining abstinence rates following treatment for alcohol abuse: a quantitative synthesis of patient, research design and treatment effects. *Addiction* 1996; 91:787-805
6. Devine JA, Brody CJ, Wright JD: Evaluating an alcohol and drug treatment program for the homeless: an econometric approach. *Eval Program Plann* 1997; 20:205-215
7. Weisner C, Schmidt L: The Community Epidemiology Laboratory: studying alcohol problems in community and agency-based populations. *Addiction* 1995; 90:329-341

EXCLUSION CRITERIA AND GENERALIZABILITY

8. McLellan AT, Luborsky L, Woody GE, O'Brien CP: An improved diagnostic evaluation instrument for substance abuse patients: the Addiction Severity Index. *J Nerv Ment Dis* 1980; 168:26–33
9. McLellan AT, Luborsky L, Cacciola J, Griffith J, Evans F, Barr HL, O'Brien CP: New data from the Addiction Severity Index: reliability and validity in three centers. *J Nerv Ment Dis* 1985; 173:412–423
10. Robins LN, Helzer JE, Cottler L, Golding E: National Institute of Mental Health Diagnostic Interview Schedule, version III, revised. St Louis, Washington University, Department of Psychiatry, 1989
11. McLellan AT, Weisner C, Hunkler E: Data from general membership and treated samples of HMO members: one case of norming the Addiction Severity Index. *J Subst Abuse Treat* (in press)
12. Gibbs L, Flanagan J: Prognostic indicators of alcoholism treatment outcome. *Int J Addict* 1977; 12:1097–1141
13. Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *J Stud Alcohol* 1997; 58:7–29
14. Wells KB: Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *Am J Psychiatry* 1999; 156:5–10