Author's response to reviews

Title: Are alcoholism treatments effective? The Project MATCH data

Authors:

Robert B Cutler (r.cutler@miami.edu)
David A Fishbain (d.fishbain@miami.edu)

Version: 4 Date: 2 December 2004

Author's response to reviews: see over
Author's response to reviews

Title: Are alcoholism treatments effective? The Project MATCH data

Authors:

Robert B Cutler (bcutler@med.miami.edu)
David A Fishbain (d.fishbain@miami.edu)

Version: 2 Date: 23 November 2004

Author's response to reviews: see over
Authors' Responses to Reviewers.

MS: 5544176094582555 -
Are alcoholism treatments effective? The Project MATCH data
Robert Cutler & David Fishbain

There were two positive reviews and one negative review. We respond to the negative review first. Reviewer's comments are in quotes.

Dr. Miller's review.

Authors' comment: Because this reviewer had a negative opinion of the article we include his comments in full and respond to what appear to be his strongest points.

Reviewer: "This manuscript reports a re-analysis of the Project MATCH data set, focusing on the question of whether the amount of treatment received was related to client outcomes. The authors represent this to be an index of whether treatment "worked." There is indeed a puzzle to explain here, which is why such a large amount of change occurs so early. Clients who enrolled in the study, completed assessment, and then dropped out before treatment still showed a within-group ES of 1.5 on both PDA and DDD, the two primary end points in the MATCH study. That is, by any standard, a large effect size. Those who completed all 12 sessions of treatment (in the two groups offered 12 sessions) showed a somewhat larger ES of 2.0 for PDA and 2.5 for DDD, but the puzzle remains because, as the authors point out, most of this change was already apparent by the first session of treatment. This phenomenon is not at all unique to MATCH, but is quite common in alcoholism treatment studies that measure drinking variables weekly. A majority of the reduction in drinking that will occur by the end of treatment is already present in the first few weeks of treatment, often the very first week. This was also true in the MET group, not included in the authors' main analyses, where clients received only two initial sessions and booster sessions at weeks 6 and 12. "By week one, study participants had reduced their drinking close to their final levell." This is a fascinating puzzle, often overlooked in treatment outcome research.

"My principal concern is that rather than focusing on why this rapid change may occur, as it did in MATCH and many other studies, the authors seem to be trying to justify that alcoholism treatment doesn’t work, and yet suggest that the phenomenon they describe is somehow unique to MATCH. They fail to support either of these two points, and I would counsel that they instead constructively address the puzzle of what is going on in treatment outcome research (MATCH and more generally) to produce such large changes so early."

Authors' response: On the first point, the conclusions in the article are drawn from the data that are presented, but the reviewer does not suggest that the analyses are invalid. With regard to the second point, this reviewer notes later in his comments that we state that the phenomenon is not unique to MATCH; that it may be present in other outcome studies such as the ones shown in Table 1 (in the reviewer's second to last paragraph of his review: "The authors make this point themselves, noting the similarity of findings
[to] the Rand report (no control over the content of treatment as usual) and [to] more recent multisite randomized clinical trials."

Reviewer: "Beyond this general issue of focus, there are some small and large problems with the authors’ analyses and conclusions. I have numbered them, and to each number added the letter a, b, or c designating what I regard to be: (a) major compulsory revisions, (b) minor essential revisions, and (c) discretionary revisions.

"1a. From the abstract onward, there is a somewhat negative bias regarding the MATCH study. For example, the results of MATCH are represented in the abstract and elsewhere as “disappointing.” Disappointment in the MATCH findings would, of course, depend on what one had hoped to find. The MATCH study is widely recognized as setting new standards of methodological rigor, which should increase confidence in its findings. By what criterion, then, were the results “disappointing”? The basis for this evaluative claim is not specified. Were the authors personally disappointed by the study’s findings? Its methodology? Why? If not, whose disappointment are they describing?

Authors’ response: In reviewing the final monograph of the MATCH series for the journal Addiction (Lead Review, Addiction, 97, 1477-1478, 2002) J. Rehm (of the Addiction Research Institute Zurich, Switzerland, Centre for Addiction and Mental Health, Toronto, Canada and University of Toronto, Canada) wrote "Why an edited monograph in 2001…? The main reason is that Project MATCH showed disappointing results overall in regards to its main goal—providing evidence for matching patients to treatment programs. As always, when the hypothesized results are not as expected, further analyses are conducted to understand the ‘why’ of the negative results."

Reviewer: "2a. The authors assert in the abstract and elsewhere that “there were essentially no patient-treatment matches.” This assertion is simply incorrect. The secondary set of a priori hypotheses in particular yielded several large and/or enduring attribute/treatment interactions, as reported in: [references omitted.]
"One possible argument to defend the authors’ claim that there were “essentially” no ATIs is their point that “some 504 hypotheses were tested.” The above-cited analyses were all a priori hypotheses, and alpha criterion levels were adjusted for the number of hypotheses being tested."

Authors’ response: Project MATCH was a well-designed well-resourced well-powered study designed to confirm relationships that were thought to be fairly well established. The reviewer helps make the case that “there were essentially no patient-treatment matches” when he has to reach down into the secondary sets of analyses to find supportive evidence of matches, and then can only say that of the 2 out of 11 found, they were large "and/or” enduring. The expectation was, of course, that there would be clinically meaningful and enduring matches found in the primary hypotheses.

Although not addressed in the article, the failure to find robust interactions in MATCH was an indirect blow to the case for treatment effectiveness. Because overall evidence for successful treatment was historically weak, an alternative explanation was that some subgroups of patients benefited from specific aspects of treatment (Stockwell,

Reviewer: "3b. The correct designation is Project MATCH, not Project Match. MATCH is an acronym."

Authors' response: The text has been changed.

Reviewer: "4a. Central to the authors' exposé tone is the claim that MATCH findings were "interpreted post hoc as evidence that all three treatments were quite effective." As the Project MATCH Research Group, we were careful not to assert efficacy claims, recognizing that there was no untreated control group, and we explicitly cautioned in print against such interpretation. The only evidence that the authors provide of such post hoc interpretation is a verbal remark by then-Director of NIAAA, Dr. Enoch Gordis, quoted second-hand from a Science News article. While it is understandable that an Institute Director might, for political purposes, wish to put positive spin on the findings of a major trial, this comment from Dr. Gordis is presented as though the investigators misrepresented the data, or that this misunderstanding of MATCH findings had been widespread. In fact our study reports did not interpret the MATCH findings as supporting absolute efficacy of any of the treatments, and several published critiques (e.g., Schaler) made the same assertion as presented by these authors, that the MATCH treatments didn’t “work.” A great deal of weight is therefore being placed on a single verbal comment, creating a straw man for the authors to attack."

Authors' response: See the response to 5a below.

Reviewer: "5a. The authors take this claim a step further by citing my article with Bennett and Walters which, they assert, “suggested that treatment is extremely effective by way of presenting the Match results.” We made no such claim or suggestion in this article. We were careful to describe the typical course of outcomes following treatment, without drawing causal inference."

Authors' response: The three sentence conclusion in the Miller article abstract was "Conclusions: About one third of clients remain asymptomatic during the year following a single treatment event. The remaining two thirds show, on average, large and significant decreases in drinking and related problems. This substantial level of improvement in "unremitted" clients tends to be overlooked when outcomes are dichotomized as successful or relapsed." [Emphasis added.] That conclusion suggests a very positive answer to the question posed by the article title "How effective is alcoholism treatment in the United States?"

We used the NIAAA Director’s comment and the Miller paper to establish two points: (1) that treatment is asserted to be effective, and (2) that the MATCH data support that effectiveness. That gave us a subject to investigate and a valid tool with which to work (the MATCH data).

Reviewer: "6a. The authors state that “Evidence is accumulating that extensive therapy may be no more effective than brief treatment.” In support of this claim they cite two literature reviews. The first, my own with Wilbourne, is a review of evidence for the
efficacy of various alcoholism treatment approaches. That review does show strong
evidence from randomized trials of the efficacy of brief interventions. There is nothing in
our review, however, to support the equivalence of briefer and longer treatment. The
amount of efficacy evidence for each modality is summarized in a cumulative evidence
score (CES), but the CES does not logically permit comparisons between specific
treatment modalities. This is also not a major point of the Moyer review. There are,
however, randomized trials that do directly compare more extensive with less extensive
treatment, only two of which are cited - the classic Edwards et al study, and Jonathan
Chick's trial of brief versus extended treatment. Their point can be well substantiated by
other studies, including the replication of the Edwards study: ….

Authors' response: The Reviewer is suggesting possibly better references, but not
disputing the point. We feel that the original references are adequate.

Reviewer: "7b. The authors state that “the Match Data set was made available to
qualified researchers for a limited time after the study had been completed. The
meaning of this is unclear, because the data are still available and being used for
secondary analyses.”

Authors' response: Here we were describing how we obtained an official copy of the
data. That source is now disbanded. Another source may have opened up. The
phrase "for a limited time" has been deleted.

Reviewer: "8b. “Data” is a plural noun. Search and change singular verbs."

Authors' response: The change has been made.

Reviewer: "9a. The authors assert that the group attending only one session “had worse
scores than the other treatment groups (those who attended 2-12 session) on 536 of
550 measures. The basis for this assertion appears to be visual comparison of means,
and no between-group statistical tests are provided in support of the claim, even for the
primary outcome measures. "

Authors' response: This was a minor statistical point added for descriptive
completeness and not germane to the subject of the article. The important point, that
the one treatment session group had worse scores than the zero treatment group, was
accompanied by between groups tests of the primary variables in addition to
comparisons between the means.

Reviewer: "10a. The authors want to conclude that “participants who reduce their
alcohol consumption are more likely to enter or remain in treatment and those who
continue drinking are more likely to drop out of treatment (p.14). To support this claim,
they cite the Rand study that showed (as typically occurs) a significant correlation
between length of stay/retention and treatment outcome, which the Rand authors
acknowledged could be due to selection bias – those who are more motivated stay in
treatment longer and also do better. However, this is precisely the opposite of what the
authors are reporting from MATCH data - that length of stay was not substantially
related to better outcomes. Indeed, their central argument is that clients did just as well
if they only completed intake and no treatment, or just one session. That is an exception to the usual finding, that retention is highly correlated with outcome. Thus they seem to want to explain their principal finding, that length of treatment did not predict outcome, by arguing that length of treatment does predict outcome because of self-selection factors. 

**Authors’ response:** Tables 2 and 3 show the correlations between outcome and attendance. Table 2 is the view typically presented, and it suggests that attendance predicts outcome. Table 3 suggests a different interpretation. It shows that drinking levels early in the study predict later treatment attendance. With regard to the Rand report, treatment quantity was not correlated with prognosis in the inpatients (p.217).

**Reviewer:** "11a. Ironically, the authors make the very same logical inference error for which they (mistakenly) critique the MATCH investigators: interpreting correlational data to infer causality. If one cannot infer from correlations between attendance and outcome that a treatment “worked,” then surely one cannot infer from non-significant correlations (including their comparisons of 0, 1, and 12 sessions) that treatment did not work (thus proving a null hypothesis). 

**Authors’ response:** Interestingly, the null hypothesis issue was discussed in relation to MATCH in the journal Addiction. (Hall, W. Patient matching in treatment for alcohol dependence: is the null hypothesis still alive and well? Comments on Project MATCH: matching alcohol treatments to client heterogeneity. Addiction, 94, (1), 52-53, 1999.) Although the issue there was the existence of interactions and the issue here is evidence of treatment effectiveness, many of his comments are relevant. Hall wrote "In declining to accept the null hypothesis, the Project MATCH investigators invoked the old shibboleth that one cannot prove the null hypothesis, a special case of the more general claim that it is 'impossible to prove a negative'. This is one of these widely believed statements that happens to be false...." He goes on to state that the critical issues are the quality of the evidence (or lack thereof) and the existence of plausible rival explanations. In view of the methodological rigor of Project MATCH, the data are probably some of the best that has ever been collected in the alcoholism field. Plausible rival explanations would probably require that some of the data are incorrect.

**Reviewer:** "12c. The authors observe correctly that MATCH sites differed in retention and drop-out rates, and wonder whether these differences “may have been due to characteristics of the therapists or participants. Part of the answer can be found in: Project MATCH Research Group (1998). Therapist effects in three treatments for alcohol problems. Psychotherapy Research, 8, 455-474. 

**Authors’ response:** None.

**Reviewer:** "13c. Similarly, the authors wonder in several places whether pretreatment patient characteristics may influence treatment retention and outcome. The effects of patient attributes on outcome are reported in the main MATCH reports."

**Authors’ response:** None
Reviewer: "14a. It is not clear what statistical test is being described on page 11 by "There were no significant differences between the zero treatment dropout group and the full treatment group from week 1 to follow up in percent days abstinent . . or in drinks per drinking day." Are these t tests of change scores? If so, between what two assessment points? "

Authors' response: The statistical results were presented in the next sentence of this two sentence paragraph. These were t-tests of difference scores from week 1 to follow-up. Follow-up was defined in the methods section.

Reviewer: "15c. The authors fail to note that on at least one time-honored outcome measure - the percentage of patients maintaining complete abstinence - patients in the Twelve-Step Facilitation treatment fared significantly better at all follow-up points than did patients in the other two conditions - a substantial advantage of about 10 percentage points that endured for the full 3-year follow-up. On this common measure, at least, the three treatments definitely were not equal in outcome."

Authors' response: We confined analyses to the MATCH primary outcome variables. An exceedingly large number of different alcoholism outcome variables have been defined, and there are often different ways that the same variable can be calculated.

Reviewer: "16a. There is also a logical flaw in inferring that if treatment gains were present at Week 1 and were maintained at Week 12, then treatment had no effect. Most of the gains in PDA and DDD that are reported in clinical trials are the result of people abstaining altogether, at least for periods of time. It is the nature of alcoholism treatment that in the early weeks of help-seeking (and after detox, as in the aftercare arm of MATCH), people tend to quit drinking and stay quit for a while. The authors seem surprised by this, that improvement "was almost instantaneous." Attaining initial abstinence is common and not all that difficult. The challenge of treatment is to help people keep from returning to drinking ("relapse prevention"), and if there has been a high rate of abstinence at the outset of treatment, helping patients to stay there could be seen as a good outcome of treatment. This pertains particularly to the aftercare arm of MATCH, where all patients entered the trial from intensive treatment and therefore were almost all abstinent at Week 0. (The baseline assessment period for this aftercare group was the period prior to admission to intensive treatment, not immediately prior to randomization.) This does not, of course, explain why the self-selected people in the zero sessions group also maintained their gains."

Authors' response: We disagree with the suggestion of a logical flaw. Outcomes of the zero and full treatment groups are highlighted and compared in the study. By definition the maximum treatment effect would be seen in the differences between the two over the weeks of the study and into follow-up.

Reviewer: "17c. It is unclear why the authors did not take advantage of the continuous retention data available for all cases in MATCH, and instead chose to eliminate from their analyses two-thirds of MATCH patients. One explanation offered is the desire to keep it simple for readers, but are readers really likely to be persuaded by analyses
predicated on one-third of cases? If indeed their findings hold up when including the entire sample, why not do so?"

Authors' response: Data on all cases were presented in Tables 2 through Table 5. Only Table 6 and the illustrations highlight the zero, one and 12 groups. Additionally the data on all groups are given in the Additional data table that accompanies the article.

Reviewer: "18a. The above problems are compounded in the Discussion section, where conclusions become still more expansive. For example, on page 12: "Ineffective treatment would be the most parsimonious explanation (#11a above) for the rather surprising (#1a) main findings of Project Match, that there was no match between patient characteristics and different types of treatment (#2a), and that all three treatments were equal (#15c). They claim that their analyses show “that current treatments are not effective” (p. 15), that MATCH “treatment was not particularly effective” (p. 16), and “that three of the best treatments currently available for addiction (sic) were not very effective” (p. 18). This is blatant use of correlational data to infer causality, and worse to prove a null hypothesis. This error is not mitigated by the use of waffle verbs like “suggest.” This line of argument gives the appearance of biased interpretation in order to justify a point that the authors want to make. Again, they fault Gordis for inferring causality from relational data in MATCH, but then predicate their entire argument on the same flawed logic, claiming that low correlation shows lack of efficacy. "

Authors' response: Generally speaking, if a large number of participants receive a carefully designed treatment and do no better than similar participants who do not receive that treatment, a reasonable conclusion is that the treatment is not particularly effective.
See the response to #11a.

Reviewer: "19a. At various points in the article, the authors claim that analysis of the relationship between treatment retention and outcome was “overlooked” in MATCH. In fact, these analyses were done, and the authors even cite one such report by the MATCH group on page 15. What was “overlooked” was the authors’ particular interpretation of MATCH findings."

Authors' response: We are not aware of any other published analyses that show, for example, that zero treatment is similar to full treatment, or that most of the treatment effect occurs before the treatment is administered.

Reviewer: "20a. Out of the blue in the discussion section comes the claim “that part of the effect is not real; many active alcoholics underreport drinking.” There is a large literature on the validity of self-report of alcoholics in outcome studies, largely indicating that self-report yields reliable reports of drinking that are at least as high as those resulting from confirmatory measures. None of this literature is cited, and instead the authors simply imply that alcoholics lie. Furthermore, to produce the “unreal” result that the authors claim, the alcoholics in MATCH would have had to under-report their drinking significantly more at follow-up (after treatment) than at baseline. MATCH included objective measures (breath, blood tests) as well as collateral reports to confirm
patient self-report - as extensive a check on self-report as done in any clinical trial of alcoholism treatment. The measures all supported the validity of self-report. In support of their claim that outcomes can be explained by client under-reporting, the authors offer only the Rand finding “that 30% of the collateral informants were unable to provide information” (which presumably means that 70% did), and that improvement can and does occur without treatment.

Authors’ response: Collateral informants typically report more drinking by alcoholics than do the alcoholics themselves (see, for example, the comparison between patient report and spouse report in the Edwards et al. paper cited in the article, or the Rand report, page 208, that found that 25% of the drinkers consistently underreported by at least 1 ounce of ethanol per day). The use of collaterals and other checks such as breath alcohol levels and liver function tests serve to partially detect and deter underreporting. Instead of "the validity of self-report" a more accurate phrase would be "the limited validity of self-report."

Reviewer: "21a. In the discussion, the authors’ unsubstantiated claim that MATCH results were misrepresented as proving treatment efficacy is further exacerbated in denouncing “Exaggerated claims of treatment effectiveness.” While their discussion points on the potential perils of exaggeration are plausible, what does this have to do with the MATCH study? "

Authors’ response: The article does not suggest that the MATCH investigators misrepresented their findings. What is has to do with Project MATCH was that it was the high quality of the work done by the MATCH teams that made these results possible, and the high quality treatment that provided a subject to investigate.

Reviewer: "22a. After repeated statements that the treatments in MATCH and alcoholism treatments more generally are “ineffective,” the authors oddly allow that “We are not suggesting that alcoholism treatment should be discontinued or even reduced.” Is that not a logical extension of their argument? And what exactly, then, is the “profound influence on alcoholism research and treatment” that the authors immodestly claim would result from “widespread acceptance” of their interpretation of the MATCH data?"

Authors’ response: If we accept the implications of these data, that these expensive types of treatment are not particularly effective, we would be forced to look in new directions for new solutions, and more likely to embrace them when they appear.

Reviewer: "23c. With all these problems noted, the authors still have an interesting phenomenon to explain: why so much gain is observed before or by the first session, not only in MATCH but in treatment outcomes more generally. The observed effect size of change from baseline to first session in MATCH is large indeed. One explanation offered is “that participants who reduce their alcohol consumption are more likely to enter or remain in treatment.” Presumably this would be true not only of randomized clinical trials, but of alcoholism treatment more generally. Dan Anderson, President of Hazelden, quipped long ago when asked how to have a highly effective treatment program, “Be the place where people decide to go once they have finally decided to quit
drinking.” The authors seem to argue, then, that once a person walks through the door of a treatment center, the “effects” of treatment after that are superstitious, having little or nothing to do with treatment itself. This does not explain, however, hundreds of randomized clinical trials demonstrating significant effects of specific interventions above and beyond presentation for treatment, intake assessment, and treatment as usual (Miller & Wilbourne review). The authors take their assertion even further to claim that “Enrolling in the trial shows that the alcoholic has crystallized a decision to reduce or abstain from drinking.” That may be, but this claim is unsubstantiated, and inconsistent with the pretreatment variability reported by MATCH patients on measures of readiness for change (e.g., URICA)."

Authors’ response: The official MATCH results themselves fail to validate the findings of patient-treatment matches that were well supported in the literature. The issue of variability treatment readiness is not addressed in this article.

Reviewer: "Another possible explanation is that there is something unique about randomized clinical trials, and that the process of presenting oneself, going through screening, surviving inclusion/exclusion criteria, accepting randomization, and enduring pre-treatment research assessment (7 hours in MATCH) has enough of an effect in itself that any impact of treatment (if any) is likely to be washed out. This is not supported by the few studies that have compared outcomes in randomized clinical trials with those for treatment as usual, including one of our own that included MATCH patients:" [references omitted]
"The authors make this point themselves, noting the similarity of findings from the Rand report (no control over the content of treatment as usual) and from more recent multisite randomized clinical trials.

"So what, then? Is it all (or mostly) making the decision to change? Are people going to AA, which really explains the positive outcomes? Are all of the efforts of treatment programs truly superstitious? Is it something about taking the public step of entering treatment or a trial? Is it that alcoholism treatment is different from interventions where no immediate impact is expected and the benefits instead emerge gradually over time (e.g., physical exercise, learning curves)? Is treatment just a form of or companion to a natural change process that occurs with or without therapists? What do the authors really think? The phenomenon cannot be explained away as an artifact of MATCH or randomized trials. Instead of the immodest axe-grinding tone of this article in current form, the authors could do better to describe, document, and reflect on a genuine puzzle in the literature, the solution to which might indeed change our thinking about how to address alcohol problems.

Authors’ response: This reviewer, as one of the MATCH Principal Investigators, undoubtedly has the full data set and a number of statisticians at his disposal. He does not challenge the accuracy of the analyses presented in the article, but instead seems to indicate that the interpretation of that data is too extreme. But he does not suggest an alternative interpretation of the data.
Dr. Moyer’s review:
Changes requested by Dr. Moyer were in the category of "Minor Essential." The reviewer’s comments are in quotes.

Reviewer: "It would be very useful to provide information on the number of individuals who fell into each of the number of treatment session groups (0-12, e.g., in Tables 4 and 5) and also to express this by treatment group (CBT, TSF, and MET). This would be instrumental in addressing whether there was differential treatment drop out at different points across treatment groups. This might provide further clues toward the subjective factors may have contributed to the self selection.
Also, it is not clear as to where there was attrition from data collection at each point during the planned treatment period and during follow-up. The degrees of freedom presented in Table 6 give readers some sense of the extent to which data are complete at follow-up for the 0 and 1 treatment groups relative to the 12 session group (around 87%)."

Authors’ response: In an earlier draft of the article we did have a table of Ns at pretreatment that partially addresses your point. It was a companion table to Tables 4 and 5, but we deleted it for purposes of brevity. It is shown below.

<table>
<thead>
<tr>
<th>Number of Sessions</th>
<th>Cognitive Behavioral (CBI)</th>
<th>Twelve-Step Facilitation (TSF)</th>
<th>Motivation Enhancement (MET)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=</td>
<td>N=</td>
<td>N=</td>
<td>N=</td>
</tr>
<tr>
<td>0</td>
<td>27</td>
<td>28</td>
<td>45</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>36</td>
<td>54</td>
<td>31</td>
<td>121</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>46</td>
<td>55</td>
<td>134</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>27</td>
<td>79</td>
<td>137</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>29</td>
<td>367</td>
<td>417</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>29</td>
<td>N=</td>
<td>44</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>23</td>
<td>N=</td>
<td>42</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>22</td>
<td>N=</td>
<td>46</td>
</tr>
<tr>
<td>8</td>
<td>29</td>
<td>29</td>
<td>N=</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td>35</td>
<td>N=</td>
<td>62</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>38</td>
<td>N=</td>
<td>78</td>
</tr>
<tr>
<td>11</td>
<td>70</td>
<td>62</td>
<td>N=</td>
<td>132</td>
</tr>
<tr>
<td>12</td>
<td>195</td>
<td>160</td>
<td>N=</td>
<td>355</td>
</tr>
<tr>
<td>Total</td>
<td>567</td>
<td>582</td>
<td>577</td>
<td>1726</td>
</tr>
</tbody>
</table>

The row total at the bottom shows that there were approximately equal numbers assigned to each of the three treatments, but an unusually large number of participants assigned to MET (N= 45) did not attend any treatment sessions (Chi square = 102.3 (2),
The high number of 0 attendees assigned to MET may be an example of the self-selection effect, or there may have been some other cause of which we are unaware.

We deal with this issue more later on, but to partially address the issue, the following has been added to the text, page 15, paragraph 1, line 13:

Table 3 shows that lower drinking levels early in the study predict later treatment attendance.

Reviewer: "It would be useful to comment on the extent to which the data from early dropouts from treatment is compromised by accompanying attrition from data collection relative to the group who attended all treatment sessions. For instance, the authors assert that, once in the trial, continued monitoring of drinking behavior by study personnel may have had motivational and therapeutic benefits (even in the absence of treatment); this explanation would only be applicable to participants who continued to allow themselves to be monitored."

Authors' response: The limited data we obtained from MATCH do not allow us to make these calculations. In MATCH, when a participant did show up, all missing data from all previous sessions were filled in. Additionally, the data we obtained only tell us how many sessions were attended, but not when they were attended.

The table below shows Ns for drinking data of the 0, 1, and 12 session attendees over time.

<table>
<thead>
<tr>
<th>Number of treatment sessions attended</th>
<th>0</th>
<th></th>
<th>1</th>
<th></th>
<th>12</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Week 0</td>
<td>100</td>
<td></td>
<td>121</td>
<td></td>
<td>355</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>57</td>
<td>57%</td>
<td>88</td>
<td>73%</td>
<td>340</td>
<td>96%</td>
</tr>
<tr>
<td>2</td>
<td>67</td>
<td>67%</td>
<td>98</td>
<td>81%</td>
<td>354</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>70</td>
<td>70%</td>
<td>101</td>
<td>83%</td>
<td>354</td>
<td>100%</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>75%</td>
<td>101</td>
<td>83%</td>
<td>355</td>
<td>100%</td>
</tr>
<tr>
<td>5</td>
<td>79</td>
<td>79%</td>
<td>101</td>
<td>83%</td>
<td>355</td>
<td>100%</td>
</tr>
<tr>
<td>6</td>
<td>79</td>
<td>79%</td>
<td>101</td>
<td>83%</td>
<td>355</td>
<td>100%</td>
</tr>
<tr>
<td>7</td>
<td>79</td>
<td>79%</td>
<td>101</td>
<td>83%</td>
<td>355</td>
<td>100%</td>
</tr>
<tr>
<td>8</td>
<td>79</td>
<td>79%</td>
<td>101</td>
<td>83%</td>
<td>355</td>
<td>100%</td>
</tr>
<tr>
<td>9</td>
<td>79</td>
<td>79%</td>
<td>101</td>
<td>83%</td>
<td>355</td>
<td>100%</td>
</tr>
</tbody>
</table>
The above table shows that about 84 percent of the 0 and 1 treatment session participants came in for follow-up assessment (for which they received a financial award). But, as you can see, it is not possible to calculate the relationship between ongoing attrition and drinking.

**Reviewer:** "In Table 6 the means for the pretreatment values presented for percent days abstinent (PDA) and drinks per drinking day (DDD) differ slightly in the context of the results presented for total improvement versus instantaneous improvement. This may be due to different numbers of participants represented in each comparison with pretreatment, but this is not made clear."

**Authors' response:** Yes, the df show that there was a different N. In order to make this more explicit we have added the following statement to page 8, paragraph 3, line 7:

Drinking data were incomplete for a number of the participants in the 0 and 1 session treatment groups. The highest level of missing data occurred at week one, where data were available for 57% of the 0 session group and 73% of the 1 session group. By follow-up, data were available for about 80% of participants in both these dropout groups.

**Reviewer:** "It would be important to place the results of this study in the larger context of treatment research in general, which has produced similar arguments. For instance, evidence that patients experience early responses before treatment might be expected
to have exerted its effects has been presented for cognitive-behavioral therapy for depression. Discussing the implications of this study in this larger scope would maximize its importance.

Authors' response:

The following has been added to page 17, paragraph 2:

There may be similarities to effects seen in other types of patients. Depressed patients sometimes report significant improvement after enrolling in clinical trials but before receiving therapy [21]. Recent time-course analyses in depression report sudden decreases in depression regardless of treatment condition [22]. These rapid responders were associated with better outcome at the end of the treatment and into follow-up [23].

Reviewer: "In this vein, it would be useful to discuss more broadly the various (positive and negative) reasons that participants may drop out of a treatment clinical trial (e.g., demoralization at being assigned to a treatment modality that they did not want, feeling so empowered by taking the step of seeking treatment that they believe that they do not need treatment) and how these might have played into the results found here."

Authors' response: The data set we obtained did not include termination reason, which would allow us to address this issue. The following has been added to page 17, para 1, line 6:

There are a large number of both positive and negative reasons why alcoholic participants drop out of clinical trials. Some of the positive reasons may include work commitments, pregnancy, re-location to another area and remission from drinking. Negative reasons include continued or increased drinking, abuse of some other substance, attitude towards the clinical staff or environment, physical illness, hospitalization and incarceration.

Reviewer: "It would be useful to discuss more thoroughly the ways in which individuals who drop out before experiencing any treatment sessions and those who drop out after engaging in one session might differ and how that may have led to the study findings. For example, for what reason would the baseline drinking measures be worse for the 1 treatment group than for the 0 treatment group?"

Authors' response:

Broadly speaking, dropouts tend to be the more seriously afflicted individuals with the worst prognosis. (e.g., Cutler RB, Fishbain DA, Cole B, Rosomoff HL, Steele-Rosomoff R. Identifying patients at risk for loss to follow-up after pain center treatment. Pain Medicine 2001, 2(1):46-51.) We hesitate to generalize from other types of patients to alcoholics because addiction may be fundamentally different. It is also difficult to evaluate the effect of baseline differences in clinical trials because inclusion/exclusion criteria are designed to produce a relatively homogenous sample.

The following has been added to page 15, para 1, line 6:
The higher drinking level of the 1 session dropout participants at baseline suggests that they may have been more dependent on alcohol than those in the 0 session dropout group. The relative higher level of dependence may have put these individuals under more pressure to do something about their drinking, explaining why they did not drop out prior to the first session.

Reviewer: "A critical question arising from these findings is why individuals, including those who preemptively drop out before receiving any treatment, following their agreement to be part of a treatment trial, show such impressive improvement so rapidly. The notion that enrolling in a trial or deciding to seek help, itself motivates and effects change has been raised in other contexts and could be expanded upon more to make this argument more convincing."

Authors’ response:
The following has been added to page 18, last line:
In order to enter the trial participants had to first achieve a level of abstinence or reduced intake. If a participant arrives at a site in an intoxicated state immediate action is required by staff, such as admission to a detox unit, or detainment in the waiting room until the breath alcohol level returns to normal. These rules would have applied to each participant in Project MATCH at the time of enrollment and would have contributed to the rapid improvement seen in the week one data.

Reviewer: "The authors also allude to non-treatment effects that can result in reduced drinking. This also could be expanded upon because to suggest that treatment is ineffective in reducing drinking, but that the motivation provided by the monitoring by study personnel is effective could use further substantiation.

Authors’ response: The following has been added to page 19 paragraph 1, line 10:
For example, in one report of a study with a 2 year follow-up [21], over half the participants indicated they felt liked the "caring, concern and help" contact of the follow-up telephone call, and in another [27], the telephone interviewers reported that they usually entered in a sympathetic interaction with the study participants. Such positive empathetic contact could be of therapeutic benefit.

Reviewer: "To address the point of selection effects, it might be illuminating to compare the mean outcome levels of PDA and DDD found here to those found for other groups of untreated alcoholics reported in the literature who did not choose to be untreated (i.e., were assigned to a no treatment control group) as the participants in this study did. This would provide a reference point to for the zero group’s improvement compared to other non-selected untreated individuals. "

Authors’ response: The following has been added to page 20, paragraph 1:
A major contribution of Project MATCH was the high quality of the follow-up data collected on those who dropped out. It is difficult to compare these data to that reported in the literature, much of which are collected under quite different circumstances. Of the zero treatment participants, 45% (35/78) reported being abstinent during the final follow-up interval (month 15). They reported a mean of 25.1 drinks per week. These outcomes appear somewhat better than for various
types of the no-treatment only conditions in the literature that were recently summarized [17]. Of some 17 studies than included placebo or no treatment conditions, with and without prior detoxification, a mean (for studies) was 21% abstinent, and the average participant was drinking 31 drinks per week [17].

Reviewer: "The commentary presented with each table should be integrated into a descriptive title and/or the text of the manuscript.

Authors’ response: These changes have been made.

Dr. Peele’s review:
No changes were requested

Final comment by the authors:
We would like to thank all three of the reviewer for the considerable effort they have devoted to this article. We think that each of them have made major contributions to our understanding of addiction.