Treatment May Influence Self-Report and Jeopardize Our Understanding of Outcome

TED NIRENBERG, PH.D.,[a,b,c,]* RICHARD LONGABAUGH, ED.D.,[b] JANETTE BAIRD, PH.D.,[a,c] AND MICHAEL J. MELLO, M.D., M.P.H.[a,c]

[a]Department of Emergency Medicine, Alpert Medical School, Brown University, Providence, Rhode Island

[b]Center for Alcohol and Addiction Studies, Department of Psychiatry and Human Behavior, Alpert Medical School, Brown University, Providence, Rhode Island

[c]Injury Prevention Center, Rhode Island Hospital, Providence, Rhode Island

ABSTRACT. Objective: Standardized measures of self-reported alcohol use are the predominant method by which change in alcohol use following interventions is evaluated. This study examined whether the invariance of the test–retest pretreatment Alcohol Use Disorders Identification Test (AUDIT) was affected by the treatment experience. In this study, the intervening exposure was to motivational interviewing (MI) versus community service (CS), the treatment-as-usual control group. **Method:** Analyses were conducted on a subsample of court-referred 16- to 21-year-olds recruited into a randomized controlled trial examining the effects of MI on alcohol use and police charges for risky driving and/or drinking. Youths were randomized to CS or MI. A subsample of 478 participants, who at baseline completed the AUDIT in reference to alcohol use for the 6 months before their conviction, later repeated the AUDIT at treatment completion, in reference to the same 6-month baseline period. **Results:** At completion of treatment, participants receiving CS had a significant decrease in baseline AUDIT scores, whereas those in MI reported no significant change. The difference between the two groups was significant ($p = .02$). Also, of those who reported no drinking before treatment, after receiving MI, 33.5% changed their response and acknowledged pretreatment drinking, compared with only 8.3% in CS. **Conclusions:** These results suggest that treatment received may differentially affect a standardized measure of self-reported risky drinking. This effect may be attributable to the treatment experience and/or the experience of the control group. Possible explanations for the effect are explored, including more honesty because of a trusting therapeutic alliance and a response shift bias. Differential change in self-report might affect treatment outcome assessment. Depending on the treatment contrasts, research that relies on pre- to post-treatment changes in self-report may be underestimating treatment effects. (*J. Stud. Alcohol Drugs, 74,* 770–776, 2013)

DESPITE THE RELIANCE ON self-report measures within most randomized controlled trials that examine the efficacy of alcohol interventions, the validity of self-reported alcohol use has been a topic of debate for several decades (Babor et al., 1987). Alcohol use is a sensitive topic, and even when subjects are asked about it under conditions of assured confidentiality or anonymity, such self-report may result in an underestimation of use.

On the one hand, investigators have argued that overall self-report can be as accurate as—or even more accurate than—measures such as biological tests (Babor et al., 2000). However, others have indicated that the validity and reliability of self-report may wax and wane depending on the characteristics of the data collection setting and the circumstances that surround the need for the requested information and person characteristics (Brener et al., 2003; Del Boca and Darkes, 2003). Babor et al. (1987) highlighted the need to get beyond a polarized debate on whether self-report of alcohol use is valid. Rather, they presented possible variables that could be a source of bias and techniques that can be used to reduce such bias.

The present study addresses an unstudied variable, the effect that the intervention itself may have on the accuracy of self-report. It may be that as patients feel more comfortable with and trusting of the treatment process, they may become more candid in acknowledging problem behaviors. Also, as patients become more knowledgeable about the target behavior, they may alter their self-report. If one or both of these occur, it may have ramifications for many treatment outcome studies that use self-report measures of alcohol use as the primary measures of treatment outcome. If we find that a treatment differentially affects one's self-report of problematic behaviors, this may have an undetected effect on treatment outcome data. This differential effect may be greatest in studies comparing outcomes of treated versus untreated groups. Patients randomly assigned to a treatment group, in contrast to those in the no-treatment control group, may provide a more candid or accurate report of their drinking behaviors over time, and through increased openness or awareness of how much they are drinking, report more alcohol use. The control group, not exposed to this process, may underreport their alcohol use. If so, the effects of treat-

ment on drinking would be confounded with the effects of treatment on accuracy of self-report.

The objectives of the present study were (a) to test the effect of motivational interviewing (MI) delivered in a group format versus treatment as usual on assumed invariance of self-reported hazardous drinking and (b) to explore why the effect occurs. We hypothesized that an MI group in contrast to a community service (CS) control group would report relatively more pretreatment hazardous drinking at the end of the intervention.

Method

Subject population

Participants were 16–21 years old and were court referred as part of their community service sanctions for high-risk driving and/or alcohol/other drug charges. All participants provided informed consent/assent to participate in the research study, completed an assessment battery, and were randomized into a CS control group or one of two MI groups. To test the hypothesis of the present study, the two MI groups were combined. A consecutive sample of 478 participants, who at baseline completed the Alcohol Use Disorders Identification Test (AUDIT) in reference to the 6 months before their conviction, completed the AUDIT again at treatment completion in reference to the same 6-month baseline period. Participants were asked to provide accurate reporting and were assured confidentiality (i.e., the separation of identifying information and obtained information was stressed). They completed the assessment in privacy on a computer. Also, the protections provided by the obtained National Institutes of Health certificate of confidentiality were explained; participants were clearly told that the information provided by them would not and could not be communicated to the judicial system. All study procedures were approved by the hospital's institutional review board.

Interventions

Community service. As part of their court-mandated sanctions, all youths had to complete 20 hours of community service, which was developed and administered by research staff. Participants received an introductory session preparing them for their community service and an educational session on the various types of motor-vehicle crashes, related injuries, and high-risk driving behaviors. Other than drinking and driving, there was no focus on hazardous drinking. Participants completed their community service at their designated location, attended a final wrap-up session focusing on their experience, and then completed post-program assessments.

Motivational interviewing. Because our previous study did not find any major differences between the two MI

conditions (Nirenberg et al., 2013), in the present study the data from these groups were combined. Both MI groups were based on an adapted motivational intervention. The only difference between the two MI groups was the type of community service they participated in (either typical community service or observation in the hospital trauma center). The counselors who facilitated the MI groups were trained in MI, which stressed (a) the pivotal role of the participant in the decision to change behavior, (b) the locus of control for change resting with the youth, and (c) the nonjudgmental role of the counselor. Both groups received four group sessions (3 hours each), one individual session (1 hour), and a community service experience (6 hours). MI participants (a) received an interactive educational session on the various types of motor-vehicle crashes and related injuries as well as high-risk driving behaviors, (b) received an introduction to the concept of decisional balance in relation to alcohol use and high-risk driving behaviors, and (c) were asked to share the pros and cons of alcohol use and high-risk driving behavior.

The counselors discussed stages and processes of change, provided feedback regarding norms of peer alcohol and other drug use, and used reflective listening and summarizing techniques to establish the pros and cons of alcohol use and high-risk driving behaviors. They used the community service experiences to assist participants in focusing on future goals and values and how current alcohol use and high-risk driving might affect those goals. If appropriate, participants prepared a change plan. Participants were seen individually to give them an opportunity to discuss any issues they had not felt comfortable discussing in front of the group. Additional description of the CS group and MI group content is provided in Nirenberg et al. (2013).

Data collection

Data included in the present analyses were collected before study randomization and again after study group assignment at the end of the final intervention session. These time points were approximately 4–6 weeks apart.

Modified AUDIT. The AUDIT is a 10-item self-report instrument used to determine whether a test subject is drinking in a hazardous manner (Saunders et al., 1993). Drinking patterns and alcohol-related negative consequences are assessed. The total AUDIT score is the weighted sum of the 10 items. The AUDIT has good internal reliability, with Cronbach α's ranging from .80 to .94 and good sensitivity and specificity (Daeppen et al., 2000; Fleming et al., 1991; Saunders et al., 1993). Donovan (2009) found that lower quantities of alcohol use by adolescents may have effects equivalent to those produced by higher levels of use by adults. In this regard, for Item 3 of the AUDIT, we have substituted "four or more drinks on one occasion" as the heavy episodic drinking ("binge") item, in place of

the "six or more" used for adults. For the present study, this AUDIT heavy episodic drinking ("binge") question ("How often do you have six or more drinks on one occasion?") was changed to the following: "How often do you have four or more drinks on one occasion?" The first AUDIT question ("How often do you have a drink containing alcohol?") also was used separately to assess whether the respondent drank or did not drink at all during the 6-month pre-conviction period. There are three subscale scores that can be derived from the AUDIT items—alcohol consumption (Items 1–3), drinking behavior related to dependence (Items 4–6), and alcohol-related problems (Items 7–10) (Karno et al., 2000). The total and subscale scores are used in the analysis.

MI-Experience was adapted from an instrument used by McNally et al. (2005) and has been modified to a 19-item instrument that included ratings of counselor behaviors that have been shown to have promise as potential mechanisms of change in adapted MI. Participants rated the 19 items on the modified MI-Experience form on a Likert-type scale with a range from 1 (*strongly disagree*) to 4 (*strongly agree*). We conducted a factor analysis on these MI-Experience ratings. The resulting factor analysis with varimax rotation resulted in 11 items being retained with a loading of at least .40 to yield one therapeutic experience factor. The responses to items on this scale were summed to form the unweighted MI-Experience score used in the data analysis reported here. This instrument was administered to all participants at the first and last group session. The first session score was used to examine differences between both MI groups and CS.

Group Climate Questionnaire–short form (GCQ-S) is a 12-item scale that measures participants' assessment of the group climate (MacKenzie, 1983). Responses follow a 7-point Likert scale ranging from 0 (*not at all*) to 6 (*extremely well*). There are three subscales: engagement, avoidance, and conflict. Internal consistency estimates range from .88 to .94 (Ogrodniczuk and Piper, 2003). The GCQ-S was administered to participants after the first and last group sessions.

Demographic assessment was administered at baseline and included gender, age, ethnicity, and race.

Data analysis

Participant responses were imported from the remote secure DatStat data storage system into SAS Version 9.3 (SAS Institute Inc., Cary, NC) for all data analysis. Demographic characteristics of the two groups were compared using chi-square tests for categorical variables and t tests for continuous variables. A difference score between the participant scores on the modified AUDIT at the two time administrations was calculated. T tests were conducted on the difference scores between the two AUDIT administrations to determine differences in self-reported hazardous

drinking between and within groups. A chi-square test was conducted to look at group difference in the change in proportion of those who reported no alcohol use at the two administrations. Regression models were conducted to estimate whether the effect of treatment condition on change in AUDIT score between administrations was mediated by group engagement and/or MI experience.

Results

Participants

A total of 553 participants who completed the baseline AUDIT were included in the sample of participants who could complete both administrations of the modified AUDIT. Of 553 participants eligible to complete the end-of-treatment second administration of the AUDIT, a total of 478 participants did so—162 from the CS group (85% of those who completed the baseline AUDIT) and 316 from the MI group (84% of those who completed the baseline AUDIT). There were no significant differences between these groups in age, which averaged 17.9 years, $t(470) = -0.19$, $p = .85$ (CS: $M = 17.94$, $SD = 1.31$; MI: $M = 17.96$, $SD = 1.31$); ethnicity, $\chi^2(1) = 0.33$, $p = .56$ (non-Hispanic: CS = 95%; MI = 94%); race, $\chi^2(1) = 0.001$, $p = .94$ (White: CS = 90%; MI = 91%); or gender, $\chi^2(1) = 0.63$, $p = .73$ (male: CS = 68%; MI = 69%). The groups also did not differ on the type of police charge that led to referral to the Reducing Offenses of Adolescent Drivers (ROAD) program, $\chi^2(2) = 1.05$, $p = .59$ (driving charge: CS = 46% vs. MI = 48%; driving and substance-related charge: CS = 8% vs. MI = 6%; substance-related charge: CS = 46% vs. MI = 46%).

AUDIT

At the completion of their treatment, participants in the MI group reported a mean increase in baseline modified AUDIT of 0.15 ($SD = 5.22$), whereas those in the CS group had a mean decrease of -0.97 ($SD = 4.77$). These changes resulted in a significant between-group mean difference between the group's baseline and treatment completion reports of AUDIT scores, $t(476) = -2.36$, $p = .02$ (Table 1). Although a within-group mean difference in the modified AUDIT group was found for the CS group, $t(159) = -2.85$, $p = .01$, there was no significant change for the MI group, $t(317) = 0.41$, $p = .68$. Using Cohen's d and adjusting for the correlation for within-group means, the effect size for this within-group change in AUDIT score was .23 for the CS group and .02 for the MI group. Also, of those who reported no drinking before randomization, after receiving MI, 33.3% changed their response and acknowledged drinking (21 of 63 changed their response) compared with only 8.3% in the CS group (3 of 25 changed their response), $\chi^2(1) = 5.95$, $p = .02$.

TABLE 1.   Mean modified AUDIT scores by group and at each administration

Variable	Community service (n = 160) M (SD)	Motivational interviewing (n = 318) M (SD)	Statistic (476 df) t	p
Baseline				
modified AUDIT	6.60 (5.82)	6.02 (5.60)	1.17	.24
Post-completion				
modified AUDIT	5.63 (6.02)	6.17 (5.87)	-0.93	.35
Difference between baseline and post-completion				
modified AUDIT	-0.97 (4.77)	0.15 (5.22)	-2.36	.02

Note: AUDIT = Alcohol Use Disorders Identification Test.

To better understand how the two groups differed in their changing scores, we compared the difference in scores for the three AUDIT subscales that measure alcohol consumption, alcohol dependence, and alcohol-related problems. As seen in Table 2, the difference in subscale change scores between groups was only significant for the alcohol consumption subscale (AUDIT-C), $t(476) = -4.21$, $p < .0001$. Within-group change for the AUDIT-C was also significant, with the CS significantly decreasing their score on the drinking subscale, $t(159) = -3.84$, $p < .001$, whereas the MI group directionally increased their score, $t(317) = 1.90$, $p = .06$, between the two administrations. Again, we calculated effect size for these within-group changes using Cohen's *d;* this was 0.31 for the CS group and 0.14 for the MI group.

Further regression analyses showed the importance of the AUDIT-C subscale in explaining the overall between-group AUDIT differences, as the AUDIT-C items accounted for 73% of the variance in the total AUDIT score. As this subscale was the major factor in accounting for the difference in change in the two groups, we examined each of the three items that make up the alcohol consumption score.

There was a significant decrease in the CS group report of frequency of drinking compared with MI, $t(476) = -2.58$,

$p = .01$ (mean change: CS = -0.27, $SD = 0.89$; MI = -0.03, $SD = 0.96$); a significant increase in the MI report of number of drinks on an average drinking day, $t(476) = -4.01$, $p < .001$ (mean change: CS = -0.13, $SD = 1.12$; MI = 0.31, $SD = 1.15$); and a significantly greater decrease in the CS report of heavy drinking days, $t(476) = -3.38$, $p = .001$ (mean change: CS = -0.33, $SD = 0.89$; MI = -0.02, $SD = 0.94$).

Post hoc analyses were conducted to determine if gender, age, and type of referral police charge were significant covariates of the effect of group on AUDIT difference. Age was dichotomized to those participants age 18 or younger and those older than 18 when enrolled in the ROAD study. Analyses of covariance were conducted. Age, $F(1, 476) = 0.32$, $p = .55$; gender, $F(1, 476) = 1.46$, $p = .23$; and referral police charge, $F(1, 476) = 1.63$, $p = .20$, were not significant covariates, and there were no Gender × Group, Age × Group, or Referral Charge × Group interactions that predicted the change in AUDIT scores.

Explanatory variables

To try to further understand why these differential changes occurred, we conducted several post hoc analyses.

TABLE 2.   Group differences in AUDIT subscales change

Variable	CS M (SD)	MI M (SD)	Statistic
Pre- and post-ROAD mean difference in alcohol consumption subscale (AUDIT-C) (Items 1–3)	-0.73 (2.39)	0.26 (2.42)	$t(476) = -4.21$, $p < .0001$
Pre- and post-ROAD mean difference in drinking behavior related to dependence subscale (Items 4–6)	0.09 (1.61)	0.11 (1.82)	$t(476) = -0.04$, $p = .96$
Pre- and post-ROAD mean difference in alcohol-related problems subscale (Items 7–10)	-0.34 (2.50)	-0.22 (2.67)	$t(159) = 1.14$, $p = .37$

Notes: AUDIT = Alcohol Use Disorders Identification Test; CS = community service; MI = motivational interviewing; ROAD = Reducing Offenses of Adolescent Drivers.

TABLE 3.   GCQ-S and MI-Experience factor scores for groups at end of Session 1

Variable	CS M (SD) (n = 157)	MI M (SD) (n = 316)	Statistic
GCQ-S engagement	2.43 (0.99)	2.94 (1.06)	$t(471) = -5.09, p < .001$
GCQ-S conflict	1.29 (1.17)	1.40 (1.44)	$t(471) = -0.84, p = .36$
GCQ-S avoidance	2.59 (1.01)	2.63 (1.14)	$t(471) = -0.35, p = .73$
MI-Experience	33.37 (5.04)	36.15 (5.05)	$t(471) = -5.64, p < .001$

Notes: GCQ-S = Group Climate Questionnaire–short form; MI = motivational interviewing; CS = community service.

Treatment experience

After the first group session, the CS and MI group participants responded to the MI-Experience and GCQ-S. Table 3 shows the mean scores on the MI-Experience factor and the three GCQ-S factors for each group. The GCQ-S engagement and MI-Experience factors were significantly different between the groups, with the MI group being higher on both measures. We conducted bivariate correlations between these factor scores and the total AUDIT and AUDIT-C difference scores within each group and for the two groups combined. For the combined group, although MI-Experience and GCQ-S were significantly correlated ($r =.25$, $p < .00l$), only GCQ-S was correlated with the total AUDIT ($r =.11$, $p < .01$) and the AUDIT-C difference score ($r =.09$, $p = .05$). Within the MI group, MI-Experience and GCQ-S engagement were cor-

related ($r =.28$, $p < .001$). Whereas the GCQ-S engagement score was correlated with the total AUDIT difference score ($r =.13$, $p = .02$), the MI-Experience score was not ($r =.03$, $p =.59$). For the AUDIT-C difference scores, the correlations were also not significant (MI-Experience: $r = .02$, $p = .73$; GCQ-S engagement: $r = .10$, $p = .08$).

Within the CS group, the MI-Experience and GCQ-S engagement scores were also significantly correlated ($r = .21$, $p = .01$). However, in the CS group, neither the MI-Experience score nor the GCQ-S engagement score was correlated with either the total AUDIT difference score ($r = .03$, $p = .59$ and $r = .01$, $p = .98$, respectively) or with AUDIT-C difference scores ($r = .08$, $p = .33$ and $r = -.07$, $p = .39$, respectively).

Based on the results of these bivariate analyses, we conducted a series of sequential regression analyses to test whether GCQ-S engagement was a mediator of the relation-
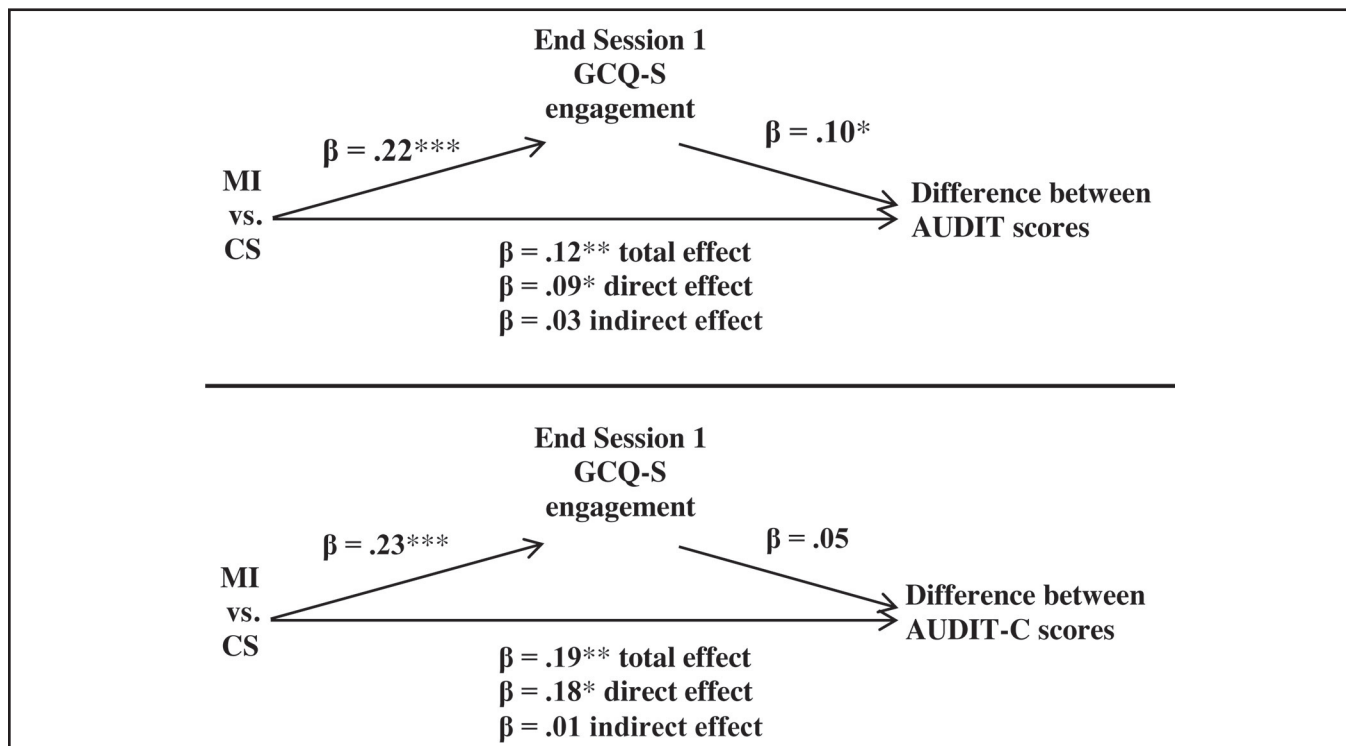

FIGURE 1.   Mediation model. GCQ-S engagement = Group Climate Questionnaire–short form engagement subscale; MI = motivational interviewing; CS = community service; AUDIT = Alcohol Use Disorders Identification Test; AUDIT-C = alcohol consumption subscale of AUDIT.
*$p < .05$; **$p < .01$; ***$p < .001$.

ship between MI versus CS assignment and change in total AUDIT or AUDIT-C scores (MacKinnon, 2008). There was a direct effect of group assignment on the difference between the AUDIT and AUDIT-C scores at the two administrations, with the CS group showing significant decreases. End of Session 1 GCQ-S engagement mediated approximately 25% of this effect, with greater engagement predicting an increase in total AUDIT scores. Although treatment condition was more strongly predictive of change in AUDIT-C, GCQ-S engagement did not explain any of this effect (Figure 1).

Discussion

At the completion of treatment, participants receiving MI and CS showed a significant difference in their reporting of baseline AUDIT and AUDIT-C scores. Those receiving CS had a significant mean decrease in their baseline AUDIT and AUDIT-C scores, whereas MI participants did not. Also, of those who reported no drinking before treatment, after receiving MI 33.3% changed their response and acknowledged drinking, compared with only 8.3% in the CS group. Speculations are offered for these observed differences.

Therapeutic alliance

One possible explanation is the strength and/or type of relationship that is built as part of the treatment experience. In our examination of a mediational model, we found that this differential effect on overall hazardous drinking (total AUDIT score) may be partially explained by the participant's self-reported engagement in the counseling group. The less that participants reported that they felt engaged in the group, the more their self-reported pretreatment total AUDIT score decreased when retested after treatment completion. Bordin's model of therapeutic alliance emphasizes the collaborative nature of the relationship between patient and therapist as a central tenet of this construct (Bordin, 1979), a theme that is also echoed in MI (Miller, 2002). Within the confines of a research protocol, we have found that when a counselor builds more emotionally supportive relationships, the patient is more likely to complete treatment (Baird et al., 2007). Extrapolating to the present findings, such a therapeutic alliance may also increase the accuracy of self-report.

Babor et al. (1987), in their examination of possible influences on self-report, highlighted task variables and respondent variables. In terms of task variables, involvement in MI may influence the rapport with the interviewer, the demand characteristics of the situation, and even the clarity of the instructions. MI may also reduce some of the respondent variables such as a respondent's defensiveness and anxiety level.

Response shift bias

Response shift bias also may explain some of the effect that treatment may have on self-report. Response shift bias refers to the possible contamination that an experimental intervention might have on the respondent's interpretation of a self-report instrument (Aiken and West, 1990; Howard, 1980). For instance, although it is probable that a randomized control group and treatment group will respond in a similar manner at pretest, after the differential treatment experience the treatment groups may alter their response patterns. In the present study, MI group members had many opportunities to discuss their alcohol consumption during the four group sessions, whereas CS participants had limited exposure to didactic information, with very limited discussion.

Therefore, the finding that those participants in the MI group reported more drinking than the CS group at the end of treatment on three drinking questions (how often, how many drinks, and how often four or more drinks, summed by the AUDIT-C) might be accounted for by the programmatic exercises in which participants examined their drinking behavior. During these exercises, it was not unusual for participants to realize that they drank much more than they had thought. This newly acquired knowledge may have then influenced their responses. Notably, collaboration did not mediate any of the difference between MI and CS AUDIT-C reports. These findings lend weight to the notion that the focus on drinking in the MI group, rather than increased collaboration, accounted for the MI versus CS reporting differences in consumption.

Dissonance reduction

That the CS group generally reported less drinking at the end of treatment than before similarly suggests that the absence of focus on drinking in their CS experience might account for the decreased salience of the event that brought them to the court and the factors that led to it. Defensive responses such as denial, minimalization, and distortion might well go unchecked.

Potential implications

A shift in self-report of alcohol use following treatment, especially an MI treatment that is alcohol focused, may have important implications. Specifically, given that the MI and CS groups changed their reports of pretreatment heavy drinking in opposite directions (MI more, CS less), it is quite plausible to hypothesize that this differential reporting would continue and would affect post-treatment self-reports as well (with MI reporting more drinking and CS less, in relation to the actual amount consumed). If so, the influence of treatment condition on self-perceived and/or reported alcohol consumption may diminish the measured differential treatment effects. More comprehensive and parametric studies of this effect are necessary to evaluate generalizability.

The current study involves young adults who were court referred to a mandatory program. The accuracy of self-report

and the reported influence in the current study of treatment condition on self-report may vary depending on the treatment setting and circumstances. Although assured confidentiality, the court-referred individuals in the present study initially may have been less forthcoming than self-referred patients, and therefore the effect of treatment on their self-reports may have been greater.

Also, we examined the influence of treatment on the self-report on one assessment instrument (AUDIT). The effect might be affected by the type of questioning. For example, the use of a more specific/detailed instrument such as the timeline follow-back (Sobell and Sobell, 1992) might be less susceptible to this influence. Finally, the treatment group was an adapted MI group, and the control group was a didactic educational presentation. Are the observed differential changes specific to MI, with its focus on acceptance and support of client autonomy and a control condition that was directive in nature, or would this effect occur with almost any alcohol treatment when compared with almost any control group?

Therefore, in general, the effect may be influenced by type of population, assessment instrumentation, perceived ramifications of the data in terms of confidentiality, type of intervention, and control group conditions. Depending on the generalizability of the observed effect, implications for interpretation of treatment effects could be considerable. The effect size observed for the change in control condition AUDIT scores is in itself sufficient to markedly alter the magnitude of measured differences in drinking outcomes based on self-reported hazardous drinking. Research that relies on pre- to post-treatment changes in self-report may be underreporting differences and thus underestimating the effects of treatment.

References

Aiken, L. S., & West, S. G. (1990). Invalidity of true experiments: Self-report pretest biases. *Evaluation Review, 14,* 374–390.

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol, 61,* 55–63.

Babor, T. F., Stephens, R. S., & Marlatt, G. A. (1987). Verbal report methods in clinical research on alcoholism: Response bias and its minimization. *Journal of Studies on Alcohol, 48,* 410–423.

Baird, J., Longabaugh, R., Lee, C. S., Nirenberg, T. D., Woolard, R., Mello, M. J., . . . Gogineni, A. (2007). Treatment completion in a brief motiva-
tional intervention in the emergency department: The effect of multiple interventions and therapists' behavior. *Alcoholism: Clinical and Experimental Research, 31, Supplement s3,* 71s–75s.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16,* 252–260.

Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health, 33,* 436–457.

Daeppen, J. B., Yersin, B., Landry, U., Pécoud, A., & Decrey, H. (2000). Reliability and validity of the Alcohol Use Disorders Identification Test (AUDIT) imbedded within a general health risk screening questionnaire: Results of a survey in 332 primary care patients. *Alcoholism: Clinical and Experimental Research, 24,* 659–665.

Del Boca, F. K., & Darkes, J. (2003). The validity of self-reports of alcohol consumption: State of the science and challenges for research. *Addiction, 98, Supplement s2,* 1–12.

Donovan, J. E. (2009). Estimated blood alcohol concentrations for child and adolescent drinking and their implications for screening instruments. *Pediatrics, 123,* e975–e981.

Fleming, M. F., Barry, K. L., & MacDonald, R. (1991). The Alcohol Use Disorders Identification Test (AUDIT) in a college sample. *International Journal of the Addictions, 26,* 1173–1185.

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review, 4,* 93–106.

Karno, M., Granholm, E., & Lin, A. (2000). Factor structure of the Alcohol Use Disorders Identification Test (AUDIT) in a mental health clinic sample. *Journal of Studies on Alcohol, 61,* 751–758.

MacKenzie, K. R. (1983). The clinical application of group climate measure. In R. R. Dies and K. R. MacKenzie (Eds.), *Advances in group psychotherapy: Integrating research and practice* (pp. 159–170). New York, NY: International Universities Press.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

McNally, A. M., Palfai, T. P., & Kahler, C. W. (2005). Motivational interventions for heavy drinking college students: Examining the role of discrepancy-related psychological processes. *Psychology of Addictive Behaviors, 19,* 79–87.

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2nd ed.). New York, NY: Guilford Press.

Nirenberg, T., Baird, J., Longabaugh, R., & Mello, M. J. (2013). Motivational counseling reduces future police charges in court referred youth. *Accident Analysis and Prevention, 53,* 89–99.

Ogrodniczuk, J. S., & Piper, W. E. (2003). The effect of group climate on outcome in two forms of short-term group therapy. *Group Dynamics: Theory, Research, and Practice, 7,* 64–76.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption–II. *Addiction, 88,* 791–804.

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. In R. Litten & J. Allen (Eds.), *Measuring alcohol consumption.* New York, NY: Humana Press.