Title: Change Talk During Brief Motivational Intervention With Young Adult Males: Strength Matters.

Authors: Gaume J, Magill M, Mastroleo NR, Longabaugh R, Bertholet N, Gmel G, Daeppen JB

Journal: Journal of substance abuse treatment

Year: 2016 Jun

Volume: 65

Pages: 58-65

DOI: 10.1016/j.jsat.2016.01.005

Change talk during brief motivational intervention with young adult males: strength matters

Jacques Gaume [1-2] (corresponding author), Tel +41 21 314 41 05, jacques.gaume@chuv.ch

Molly Magill [2], Tel +1 401 863 6557, molly_magill@brown.edu

Nadine R. Mastroleo [2-3,] Tel. +1 607-777-9269, nmastrol@binghamton.edu

Richard Longabaugh [2], Tel +1 401 272 1757, richard_longabaugh@brown.edu

Nicolas Bertholet [1], Tel +41 21 314 03 27, nicolas.bertholet@chuv.ch

Gerhard Gmel [1], Tel+ 41 21 314 73 52, gerhard.gmel@chuv.ch

Jean-Bernard Daeppen [1], Tel +41 21 314 08 75, jean-bernard.daeppen@chuv.ch

1. Alcohol Treatment Center, Lausanne University Hospital
   Avenue de Beaumont 21bis – P2, 1011 Lausanne, Switzerland
2. Center for Alcohol and Addiction Studies, Brown University
   Box G-S121-5, Providence, RI 02912, USA
3. College of Community and Public Affairs, Binghamton University
   4400 Vestal Parkway East, Binghamton, NY 13902, USA

Submitted as a Regular Article to the Journal of Substance Abuse Treatment, July 2015.

Revision 2, January 2016.

Abstract

Client change talk (CT) during motivational interviewing (MI) has been described as a predictor of change in alcohol use. We examined the predictive validity of different strength levels of CT within a brief MI session for 174 young men from the general population screened as hazardous drinkers. CT was measured using the MI Skill Code (MISC 2.1) and categorized with positive (toward change) and negative (against change) valence and 3 strength levels (1=low, 2=medium, 3=high). Analyses included linear regression models predicting drinking at 3-month follow-up, while controlling for baseline drinking. Frequency of *overall negative CT* (i.e., sum of −1, −2, −3) significantly predicted poorer drinking outcomes. In a multivariate model entering frequency of CT utterances at each level of strength (i.e. +1, +2, +3, −1, −2, −3), the directionality of negative strength ratings was consistently in the expected direction, but only *CT-2* was statistically significant. In contrast, *overall CT positive* (i.e., sum of +1, +2, +3) was not a significant predictor of less alcohol use, but the multivariate model showed that the presence of *CT+3* significantly predicted less drinking at 3-month follow-up. Averaged strength summary score (i.e. on the scale from -3 to +3) was a significant predictor of better outcome, while percent positive CT was not. Moderation analyses showed that young men with lower baseline readiness to change or lower alcohol problem severity had *higher* follow-up drinking when they expressed more *CT+1* or *CT+2*, while the opposite pattern was observed with those reporting higher baseline readiness to change or higher alcohol problem severity. Mixed findings for varying levels of positive CT strength might explain previous studies showing poor predictive validity of positive client language in MI. Together with other studies in similar settings, these findings suggest the importance of advanced MI techniques to shape client language to soften negative change talk (also known as sustain talk) and elicit positive CT verbalized with high intensity.

Keywords

Motivational interviewing, brief motivational intervention, change talk, young adults, alcohol.

1. Introduction

Motivational Interviewing (MI) originated as a treatment for alcohol use, but subsequently, there have been several hundred studies published on its use across a variety of contexts. In fact, MI has now been the subject of more than 10 meta-analyses examining its efficacy (e.g. Burke, Arkowitz, & Menchola, 2003; Dunn, Deroo, & Rivara, 2001; Hettema, Steele, & Miller, 2005; Smedslund et al., 2011). A general conclusion is that MI is significantly more effective than no treatment and approximately equal to other, sometimes more extensive, evidence-based behavioral treatments. While there is support for efficacy, effect sizes can be small to moderate, and vary in magnitude by a variety of population and intervention factors. This indicates a need to better specify how MI is associated with behavior change as a way to develop guidelines for its optimal implementation. The predictive importance of specific intervention components, therapist skills, and client mechanisms has been considered.

MI is defined as "a collaborative conversation style for strengthening a person's own motivation and commitment to change" (Miller & Rollnick, 2013, p.12). A theorized mechanism of intervention efficacy is verbalized decision-making, where client statements argue toward (positive change talk [CT+]) or against (negative change talk [CT-], also known as sustain talk) making a change in the targeted behavioral outcome (Amrhein, 2004; Miller & Rollnick, 2004). Specifically, the MI skills of the therapist should increase client language toward change (CT+, e.g., statements of reasons, abilities, and commitments to change) and decrease their language against change (CT-, e.g., statements of reasons not to change or to maintain behavior, inabilities, and commitments against change), which should, in turn, mediate intervention outcome. This proposed theoretical model is called the Technical Hypothesis of MI (Arkowitz, Miller, Westra, & Rollnick, 2008; Miller & Rollnick, 2013; Miller & Rose, 2009), and thus far,

two primary approaches to language measurement have been proposed: 1) frequency of positive and negative CT utterances within a session, and 2) average session-level strength of CT utterances along a negative to positive continuum (CT strength).

To date, the MI Technical Hypothesis has received some, but not definitive, empirical support. Studies have examined individual paths of the technical model and client language has shown a direct effect on alcohol (Bertholet, Faouzi, Gmel, Gaume, & Daeppen, 2010; Gaume, Bertholet, Faouzi, Gmel, & Daeppen, 2013; Gaume, Gmel, & Daeppen, 2008; Vader, Walters, Prabhu, Houck, & Field, 2010), drug (Amrhein, Miller, Yahne, Palmer, & Fulcher, 2003; Baer et al., 2008), and other behavioral (Hodgins, Ching, & McEwen, 2009) outcomes among those receiving MI treatment. Beyond examining individual paths, there are three studies to date supporting the MI Technical Hypothesis using a statistical mediation framework (i.e., the impact of therapist MI skills on client outcome mediated by client change talk), one study with treatment-seeking adult alcohol users (Moyers, Martin, Houck, Christopher, & Tonigan, 2009), one study targeting marijuana use among high school students (Barnett et al., 2014a), and one study on nutritional enhancement with firefighters (Pirlott, Kisbu-Sakarya, Defrancesco, Elliot, & Mackinnon, 2012). Despite these three promising studies, a recent meta-analytic review of the MI process literature (Magill et al., 2014) suggests that some measures of client language may have more utility than others. Across 12 studies, it was shown that measures of client language strength predicted behavior change outcome (r =.12, 95% CI [.03, .21], p=.006), rather than positive CT frequency (r=.06, 95% CI [-.09, .21], p=.41).

1.1. Study context and purpose

Data for the present study were drawn from a randomized controlled trial (Gaume et al., 2014) testing the efficacy of a single, brief MI session (BMI) in contrast to an assessment only control to reduce heavy drinking among non-treatment seeking young men from the Swiss general population. At 3-month follow-up, the BMI group (N=179) showed significantly lower drinking than the control group (N=182) on a composite drinking score that included drinking days per week, drinks per drinking day, and heavy episodic drinking frequency (6 drinks or more on a single occasion), controlling for the baseline value of this variable (effect size=0.22, p=0.02) (Gaume et al., 2014). Subsequent work (Gaume et al., *in press*), however, showed that the MI Technical Hypothesis was operative only under particular conditions. Specifically, the average strength of CT mediated the effect of therapist MI-consistent (e.g., reflections, open questions) behaviors on reductions in client alcohol use at follow-up only when: therapists were experienced in MI (i.e., greater than 3 years) and working with young men at higher levels of alcohol severity (as measured by the alcohol use disorders identification test, AUDIT, Babor, Higgins-Biddle, Saunders, & Monteiro, 2001). The purpose of the present research was to consider these results more closely, and particularly with regard to the theoretical and predictive validity, as well as the clinical utility of the CT construct. This work is the first, to our knowledge, to examine the relative importance of client CT utterances at each magnitude of language strength. Specifically, the literature has relied on summary measures (e.g. frequency or percentage of CT statements, averaged CT strength), implying internal consistency of the raw indicators and linearity of the raw scales. We believe the following study might be of interest both at the measurement level (e.g. to offer researchers guidance as to which measure(s) have the

greatest predictive validity) and at the clinical level (e.g. to offer clinicians, clinical trainers, and supervisors, guidance as to which type of language should be of primary focus).

The primary aim of this study was to test possible variation in the predictive validity of patient utterance measures by varying strength ratings (see Figure 1). The hypotheses were a) that positive ratings would predict lower alcohol use and negative ratings a status quo or increased alcohol use; and b) that higher strength ratings would predict stronger effect (i.e. stronger decrease in alcohol use when CT+ strength is high; stronger increase in alcohol use when CT- strength is high).

As this first analysis resulted in some unexpected findings, we added a secondary aim which was to test if baseline characteristics moderated the effects measured above. Indeed, one treatment, and subsequently one treatment theory, will not necessarily work the same across all circumstances. Therefore, another interpretation of mixed evidence for the MI Technical Hypothesis is that the theory is only operative under specific conditions (i.e., conditional indirect effects; Hayes, 2013). Regarding a behavioral intervention that targets decision- making, factors such as alcohol problem severity and motivation for change may impact the nature of this process, and subsequently, the empirical performance of the theoretical model. It has been shown that higher problem severity was related to higher BMI effects (Blow et al., 2009; Daeppen et al., 2011; Vasilaki, Hosier, & Cox, 2006). In the latter two studies, the authors concluded that the presence of more severe alcohol use patterns may help develop a discrepancy between the individual's current behavior and their future and desired goals, hence leading to higher readiness to change and resulting behavior change, which is consistent with MI theory (Miller & Rollnick, 2013). We thus added secondary analysis to investigate whether a priori hypotheses from the primary aim were verified under specific circumstances (i.e., when young men had higher

alcohol problem severity and readiness to change) and not under others (i.e., lower severity and readiness).

2. Material and Methods

2.1. Sample and inclusion procedures

Study participants were from a randomized controlled trial of BMI among hazardous drinkers within the Army Recruitment Center of Lausanne, Switzerland (Gaume et al., 2014). Switzerland has a mandatory two-day army conscription process for all males at age 19, and virtually all young men complete the assessments to determine eligibility for service in the Swiss military. This setting offers a unique opportunity to address a sample virtually representative of the general population of young men this age. Women are allowed to join the military service on a voluntary basis but were not included because of their scarcity and resulting non-representativeness. Inclusion for this study took place at the Recruitment Center of Lausanne, which serves the French-speaking part of Switzerland. At all research stages, participants were reminded that the research staff had no connection with the army and that all information was confidential.

Inclusion and study procedures are described in detail in the primary trial report (Gaume et al., 2014). Briefly, 1023 young men were randomly selected for possible participation. Of those, 192 (18.8%) were not included due to priority army assessment (i.e. additional physical, cognitive, or psychological evaluation). Of the remaining 831, 194 (23.4%) refused participation (185 due to concerns about follow-up procedures and 9 due to concerns that the BMI would be audio-recorded). Consenting participants filled out a self-administrated assessment questionnaire

with research staff assistance as needed. The Alcohol Use Disorders Identification Test–Consumption (AUDIT-C, Bush, Kivlahan, McDonell, Fihn, & Bradley, 1998) was used to screen for hazardous drinking (score greater or equal to 4, Reinert & Allen, 2007) and non-hazardous drinkers were excluded from the study (n=196, 30.8%). The remaining participants were allocated to the BMI condition (n=217) or to a control condition with no intervention (n=224). Within the BMI condition, 208 sessions (95.8%) were audio-recorded while the remaining 9 had technical problems. Follow-up procedures took place 3 months after baseline and were conducted by telephone with interviewers masked to intervention condition. Among the 208 young men with audio-recorded BMI sessions, 174 (83.6%) completed follow-up. We found no significant differences on the variables used in this study between those followed-up and those not (chi square test for categorical variables and Wilcoxon rank test for continuous, non-normally distributed variables).

2.2. Intervention

The experimental condition was a 20 to 30-minute BMI, exploring alcohol use, its related consequences, and upon participant agreement, a change plan discussion. No specific guidelines were provided regarding the content of individual interventions, but all counselors were familiar with the brief intervention format and with MI techniques and principles. The study counselors were 18 physicians and psychologists from the Alcohol Treatment Center at Lausanne University Hospital, Switzerland, selected for the parent trial (Gaume et al., 2014) to provide varied backgrounds as well as a range of clinical (from new residents/psychologists to faculty members) and MI (from beginners to recognized experts) experience. All clinicians, however, received four half-days of MI training (Baer et al., 2004) when they assumed their positions at the hospital;

experience with MI was 3.8 years on average (standard deviation=4.1), with a range between 0 months (included right after MI training) and 14 years.

2.3. Measures

The self-report assessment included the following drinking measures: usual number of drinking days per week and usual number of drinks (defined as 10 grams of alcohol such as 100 ml of wine, 250 ml of beer or 25 ml of spirits) per drinking day over the past 12 months (Rehm, 1998), and frequency of heavy episodic drinking episodes (6 drinks or more) over the past 12 months (measured on a 0-4 scale with 0=never, 1=less than monthly, 2=monthly, 3=weekly, and 4=almost daily). Additional self-report measures included alcohol problems severity using the Alcohol Use Disorder Identification Test (AUDIT; Babor et al., 2001); and readiness to change using the University of Rhode Island Change Assessment Scale (URICA) - DELTA Project Reduced Drinking Version (Soderstrom et al., 2007). Alcohol use measures were the same at 3-month follow-up except they were framed within a 3-month window instead of a 12-month window. The outcome measure for the present study was a drinking composite score computed as the mean of the z-scores for usual drinking days per week, usual drinks per drinking day, and frequency of binge drinking (used as the continuous scale from 0 to 4, see above).

2.4. Observational coding methods

Participant change language within the BMI sessions was coded using the Motivational Interviewing Skill Code (MISC), version 2.1 (Miller, Moyers, Ernst, & Amrhein, 2008). Coding procedure involved two passes through each session. The first, uninterrupted pass, assessed the session's global ratings, however they were not used in the present study. During the second

pass, the coder categorized each counselor and patient utterance using one of the 34 proposed codes. Audio-recordings were exported to Dartfish Team Pro 4.0 video analysis software (Dartfish, 2006) and recorded utterances were parsed by one coder and coded by another.

Four master-level students were trained in the MISC 2.1 by the first author. Training consisted of: 1) a short presentation on MI and the MISC, 2) detailed reading of the MISC, 3) independent, then group coding of expert-rated MI sessions, 4) quizzes and 5) independent, then group coding of BMIs from another study. Independent coding was later accompanied by group coding to increase inter-rater reliability. Throughout this training an incremental learning approach was used, starting with simple codes and building up to more complex codes. Each coder received about 60 total hours of training. Further, discrepancies and challenges were addressed weekly in joint trainer-coder meetings, which lasted throughout the entire data collection period.

For the current study, possible variation in the predictive validity of patient utterance measures by strength rating was of primary interest (see Figure 1). The CT codes were the following: Reasons to change or reasons not to change; Ability or inability to change; Desire to change or not to change; Need to change or need not to change; Commitment to change or not to change; and Taking steps toward or away from change. Each client language utterance was assigned a strength value ranging from +3 (strong inclination toward change) to −3 (strong inclination away from change/toward the status quo). The starting point for a straightforward statement of inclination was +2 or -2 (medium strength), and the intensity of inclination was increased or decreased in strength by signals in the content, vocal tone, or context of the utterance. For example, emphasis modifiers (e.g. “definitely”, “surely”, “absolutely”, “really”, “very”), definitive vocal tone of expression, or verbs such as “I swear”, “I promise”, “I *will*”

(with vocal emphasis) would indicate a strength of 3. In contrast, a somewhat qualified statement ("mostly", "pretty much", "probably", "not really") or a highly diminished statement ("I guess", "kind of", "a little") would indicate a strength of 1. Finally, the Follow/Neutral category allowed coders to characterize utterances with no inclination or link to the target behavior change.

- Insert Figure 1 here

We considered 4 ways of measuring change language in the present paper. First, we computed the frequently utilized measure of overall frequency of positive CT (or toward change talk, i.e. all CT codes given a strength from +1 to +3) and negative CT (or against change talk, or sustain talk, i.e. all CT codes given a strength from −1 to −3). Second, we computed a rate between the two latter measures: the percentage of overall positive CT (overall positive CT/(overall positive CT + overall negative CT)). Third, Average CT Strength was computed on the −3 to +3 scale (i.e. mean of all strength ratings). Finally, we computed the frequency of all CT codes given a strength of +1 (CT+1), +2 (CT+2), +3 (CT+3), -3 (CT-3), -2 (CT-2), and-1 (CT-1). Because CT+3 and CT-3 occurred rarely in sessions (less than once on average), we also created dichotomized variables for these two indicators by encoding those sessions/cases having no such utterance vs. those having one or more.

To establish inter-rater reliability, a random subsample of 42 BMI sessions (20.2%) was double-coded. Coders were blinded to whether the session they rated was single- or double-coded. Coding reserved for analyses was randomly selected from each available double-coded BMI. Since independent raters did parsing a priori, double coding could be done on the same utterances. Cohen's kappa was used to address inter-rater reliability at the utterance level (i.e.

pooling all sessions together) and was interpreted according to the categorization by Landis and Koch (1977); intra-class correlations (ICCs) were used for each individual code and was interpreted according to the categorization by Cicchetti (1994). Overall inter-rater reliability in this study was excellent. At the utterance level (i.e. pooling all sessions together), Cohen's kappa was 0.87, indicating almost perfect agreement according to Landis and Koch (1977). We used weighted Kappa to measure inter-rater reliability on the strength rating scale (i.e. from -3 to +3), and this measure was also quite high (0.87). Inter-rater reliability for each individual code (Table 1, right hand column) was assessed using intra-class correlations (ICCs) and interpreted according to the categorization by Cicchetti (1994). Agreement was excellent for all individual codes (ICCs ranging from .81 to .99) except for the frequency of CT+3 which was rarely observed and for which agreement was poor (ICC=0.06). We used Cohen's Kappa for the two dichotomized variables (occurrence of CT-3 and CT+3, see above) and found almost perfect agreement for CT-3, and fair agreement for CT+3.

2.5. Statistical analysis

As a first step, we used classical descriptive statistics to characterize each CT measure in the present sample. In a second step, we tested each CT measure separately as the independent variable in separate ordinary least squares regression models with the drinking composite score at 3-month follow-up as the dependent variable, controlling for the drinking composite score at baseline. To test their unique predictive validity, we also tested frequency of all individual strength measures (i.e. CT+1, CT+2, dichotomized CT+3, CT-1, CT-2, and dichotomized CT-3) entered together in a single, multivariate model with the drinking composite score at 3-month follow-up as the dependent variable, controlling for the drinking composite score at baseline. We

computed variance inflation factors to detect potential multi-collinearity among these 6 independent variables, but found no evidence of multi-collinearity.

As a final step, we tested if baseline characteristics (i.e. readiness to change using the URICA, and alcohol problems severity using the AUDIT) moderated the effects measured above to investigate whether predictive validity varies by clinically informative patient factors. Specifically, one frequency of individual strength measure (i.e. CT+1, CT+2, dichotomized CT+3, CT-1, CT-2, or dichotomized CT-3), one baseline characteristic, and their interaction term were entered to predict the drinking composite score at 3-month follow-up while controlling for the drinking composite score at baseline. In the presence of significant interaction effects, we used the "/plot" subcommand for Model 1 implemented in the PROCESS macro for SPSS developed by Hayes (2013). This subcommand generates a table of predicted values for the dependent variable at specified values of the independent variable and the moderator (set here as the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles). The covariate in the model (drinking at baseline) was set to its sample mean when deriving the predicted values. As AUDIT score was correlated with baseline drinking (r=.68), we repeated this analysis first without baseline drinking as a control variable, and second without AUDIT score but with an interaction between the independent variables and the baseline drinking composite. All analyses resulted in similar patterns of effect, and we thus present our initial model in the present paper.

3. Results

3.1. Baseline descriptive statistics

At baseline, the 174 young men drank, on average, slightly more than 2 days per week (mean[M]=2.3, standard deviation [SD]=1.6) and about 5 drinks per drinking day (M=4.9,

SD=4.6). About one third had binge drinking episodes monthly and one third weekly. According to the AUDIT, severity of alcohol use patterns was in the mid-range on average (M=10.0, SD=4.1), between hazardous drinking (score of 8, Babor et al., 2001) and probable dependence (score of 12, Gache et al., 2005). Readiness to change was low, with mean of 1.6 (SD=2.6) on a possible range from -2 to 14, indicating that participants were mostly in the pre-contemplation stage of change (see HABITS Lab, 2015).

3.2. Change language descriptive statistics

There were about 7 utterances of CT+ and 12 of CT-, on average, in these 20 to 30 minute BMI sessions (Table 1). Consistently, the percent of positive CT (M=33.3) and the averaged strength of change language (M= -0.3) indicated a stronger inclination to the status quo. CT+ utterances rated with a strength rating of 2 were the most frequent. Strength ratings of 1 (i.e. indicating decreased intensity of inclination) were more sporadic (about 2 CT+1 and 1 CT-1 per session on average). The strength ratings of 3 (i.e. increased intensity of inclination) were infrequent, CT+3 and CT-3 occurring less than once per session on average. Dichotomization of these variables showed that about 9% of the sample had 1 utterance or more rated +3 in their session, and about 25% had 1 utterance or more rated -3.

- Insert Table 1 here

3.3. Predicting drinking at 3-month follow-up with change talk measures

We first tested each CT measure as independent variables in separate, ordinary least squares regression models with the drinking composite score at 3-month follow-up as the dependent

variable, controlling for the drinking composite score at baseline. Overall CT- frequency and overall Average CT Strength were significantly related to outcome, both in the expected direction (Table 2). However, Overall CT+ frequency and Percent CT+ were not significantly related to outcome. Regarding the frequency of each strength rating, the frequency of CT-2 was significantly related to poorer outcome while the frequency of CT-3 had a trend (p=0.07) in the same direction. Consistent with overall CT+, the frequency of CT+1 and CT+2 were not significant predictors of outcome. However, the occurrence of one or more CT+3 utterances was associated with better drinking outcome at follow-up.

- Insert Table 2 here

To test unique predictive validity, we then tested frequency of change language strength measures (i.e., CT-3 dichotomized, CT-2, CT-1, CT+1, CT+2, and CT+3 dichotomized) entered together in a single, multivariate model with the drinking composite score at 3-month follow-up as the dependent variable, controlling for the drinking composite score at baseline (Table 3). Results showed that higher frequency of CT-2 was significantly related to more drinking while having one or more utterances of CT+3 was significantly related to less drinking. Frequency of CT+2 showed an unexpected *positive* relationship to drinking and was marginally significant (p=0.06).

- Insert Table 3 here

3.4. Effect of change language strength measures on drinking at 3-month by baseline characteristics

Finally, we tested 12 models, namely if 1) baseline readiness to change and 2) alcohol problem severity moderated the effects of the 6 change language strength measures (i.e., CT-3 dichotomized, CT-2, CT-1, CT+1, CT+2, CT+3 dichotomized) on drinking at 3-month (controlling for baseline drinking). Three models out these 12 models showed significant moderations, i.e. interaction effects: CT+1 X Readiness (p=0.04), CT+2 X Readiness (p=0.01), and CT+2 X Severity (p=0.02). To help visualize and interpret the nature of the moderation, we used the "/plot" subcommand in Model 1 of the PROCESS macro for SPSS (Hayes, 2013). The complete output from PROCESS is available as online Supplementary Material for interpretation of the models to be properly made. The graphs for the 3 significant interactions are presented in Figure 2. The three graphs showed similar patterns. As predicted by MI change process theory, young men with higher baseline readiness to change had lower predicted drinking when they expressed more CT+1 and CT+2; those with higher alcohol problem severity did so when they expressed more CT+2. However, young men having lower baseline readiness to change had unexpectedly *higher* predicted drinking when they expressed more CT+1 or CT+2; those with lower alcohol problem severity also had unexpectedly *higher* predicted drinking when they expressed more CT+2.

- Insert Figure 2 here

4. Discussion

This study examined the relative predictive validity of client change language at each level of strength among a sample of non-treatment seeking, young male hazardous drinkers. We further considered whether client alcohol severity and readiness for change moderated these effects. Our results were not entirely anticipated and speak to the impact of contextual factors on the performance of theoretical models, and more importantly, behavioral intervention process. There is support for client positive change talk and negative change talk (or sustain talk) as mechanisms of MI efficacy, but recent studies suggest there may be key differences in the validity of these indicators among younger populations with lower motivation for change (e.g. mandated college students; Apodaca et al., 2014). Our findings contribute to this emerging story.

Consistent with findings in a meta-analysis examining the role of CT on drinking outcomes (Magill et al., 2014), negative CT (or sustain talk) robustly predicted maintenance or an increase of current drinking behaviors. As past studies have also found these results, specifically with young adults (Apodaca et al., 2014; Baer et al., 2008; Borsari et al., 2014; Gaume et al., 2013; Vader et al., 2010), there may be similar mechanisms at work. Altogether, these results suggest the need for clinicians to pay particular attention to this type of client language during BMI, and the importance of strategic techniques to soften sustain talk and ideally reframe it to change talk (Barnett et al., 2014b).

Somewhat contrary to MI theory, overall positive CT (i.e. language toward change) did not predict lower drinking. Our multivariate model showed that having one or more utterances of CT+3 was significantly related to less drinking while the frequency of CT+2 was related to more drinking at a marginally significant level ($p=0.06$). These inconsistencies among positive

strength ratings might explain previous findings as to the varying predictive value of positive client language in MI (Magill et al., 2014).

One finding consistent with MI hypotheses was that high strength ratings of positive CT, specifically CT+3, was a predictor of lower drinking. Although this type of utterance was infrequent and coded with fair reliability, the fact that one or more of these CT+3 ratings predicted lower drinking over and above negative CT is clinically important to consider and warrants further examination in process research. As counselors often strive to gain strong statements towards change, the current findings lend support for the importance of noticing these utterances within sessions, even if infrequently expressed. Although beyond the score of this study, future research should consider exploring the sequential order to clients' and counselors' statements that lead to CT+3, and therefore illuminate ways to draw these statements from clients within MI sessions. Finding specific counselor behaviors that have this potential may help in new treatment development, as well as training and supervision approaches for brief alcohol interventions.

Interestingly, CT+2 was close to predicting poorer outcomes, which leads to questions about how different CT+2 utterance is from a CT+3 and what differentiates between the two. Understanding this nuanced finding may help counselors to move the client more towards change and to amplify this change perspective. An important finding is that our moderation analyses identified that one theory does not fit all models. Specifically, moderation effects showed that CT+1 and CT+2 when combined with high readiness to change and CT+2 when combined with high alcohol problem severity did predict change in alcohol consistent with MI theory. However, findings inconsistent with the theory were observed at low scores on these scales. One possible explanation is that young, non-treatment seeking men with hazardous

drinking but medium alcohol problems severity and low readiness to change may offer low positive change talk throughout the intervention to simply “follow the counselor” without giving much thought to their statements. Another explanation may be related to social desirability, in that sense that these young men would ‘give’ change talk as they feel that the counselor is expecting it. On the other hand, CT+3 was not moderated by similar paths and might thus capture a marked intensity toward change or client reflection of true salience. This may reflect authenticity on the part of the client that leads to expected drinking outcomes (Miller & Rose, 2009). Similar hypotheses have been proposed when examining identical constructs with mandated college students (Apodaca et al., 2014; Borsari et al., 2014). In point of fact, these findings may be a partial explanation for the findings of the larger study that clinician experience with MI was critical (Gaume et al .2014). Clinicians with more MI experience may be better at noticing "shallow" positive change talk and may be more willing to dig deeper. It might also be that more seasoned MI clinicians convey a greater sense of autonomy to the clients, resulting in more genuine CT. Further studies should test these hypotheses, for example by administering a social desirability scale, by asking participants to report on how honest they were during the session, or by seeing if CT scores relate to therapeutic bond, working alliance, or support of client autonomy.

Regarding tested summary scores, the Averaged strength score was a significant predictor of better outcomes, while the Percent of positive CT was not. This might be explained first by the different directionality within the overall positive CT frequency measure. Second, data suggest stronger effects of CT at both ends of strength continuum (stronger detrimental effect at -3 and stronger positive effect at +3). The averaged strength measure thus appears as a potent summary score for CT, as it was already observed in a similar setting (Gaume et al., 2013).

4.2. Limitations

The present results are to be considered within the specific population (i.e. young, French-speaking, hazardous drinkers males from the general population not seeking treatment) and the particular type of MI intervention (i.e. a single BMI) they emerged from. The resulting low readiness to change observed within this sample should also be taken into account. Psycholinguistic coding was conducted with overall high inter-rater reliability, except for the measure of highest strength of positive change talk (CT+3). This was certainly due to the low frequency of such a code within our particular population. However, even if infrequent and coded with only fair reliability, the presence of such a language behavior was a strong predictor of actual change. This definitely calls for replication and further research on the predictive validity of high strength change talk.

4.3. Conclusions

Overall, sustain talk strongly predicted poorer outcomes but, inconsistent with what was expected, overall positive CT was not a significant predictor of lower drinking. Regarding tested summary scores, the Averaged strength score was a significant predictor of better outcomes, highlighting the utility and analytical parsimony of this score as a measure of client in-session talk. Multivariate models showed that the presence of CT+3 significantly predicted lower drinking while CT+2 was a trend-level predictor of *higher* follow-up drinking ($p=0.06$). Moderation analyses showed that this effect was explained by young men having lower baseline readiness to change and alcohol problem severity for whom more CT+1 or CT+2 was related to unexpected higher drinking at follow-up. On the other hand, those with higher baseline readiness

to change or alcohol problem severity had lower drinking when they expressed more CT+1 or CT+2, as predicted by MI change process theory.

MI is a 'talk therapy' that recognizes, explores, and attempts to resolve client ambivalence (Miller, 2000). Within the therapeutic context, verbalized pro- and anti-change statements are a manifestation of ambivalence, and according to MI theory, resolved ambivalence will result in an interpersonal 'contract' between therapist and client (Miller & Rollnick, 2013). This study raises concerns about the value of that contract under certain conditions.

These results should be considered preliminary, but highlight the importance of matching the complexity of treatment-facilitated change with more nuanced examination of therapeutic process (Magill & Longabaugh, 2013). The present findings also have broader implications for efficacy research on MI, and other behavioral treatments, showing small to moderate and variable effects. These treatments might only be working for certain people and under certain conditions. Further research is thus needed and efficacy studies should more often embed tests of underlying mechanisms and moderation of effects.

Acknowledgments

This research was supported by the Swiss National Science Foundation, Grants FN 33CSC0-122679 and PBLAP3-134347; and by the Swiss Alcohol Research Foundation. The funding sources had no influence in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

We gratefully thank Aurélia Brugger, Claudia Campistol, Marie Desponds, Maïlys Korber, and Julien Flückiger for the data collection; Sylvie Bourban for the parsing of interventions;

Odile Cantero, Ariane Rosselet, Stéphanie Rupp, and Sébastien Froidevaux for the parsing and coding of interventions; and the 18 counselors who provided the BMI.

References

Amrhein, P. C. (2004). How Does Motivational Interviewing Work? What Client Talk Reveals. *Journal of Cognitive Psychotherapy, 18*(4), 323-336.

Amrhein, P. C., Miller, W. R., Yahne, C. E., Palmer, M., & Fulcher, L. (2003). Client commitment language during motivational interviewing predicts drug use outcomes. *J Consult Clin Psychol, 71*(5), 862-878. doi:10.1037/0022-006X.71.5.862

Apodaca, T. R., Borsari, B., Jackson, K. M., Magill, M., Longabaugh, R., Mastroleo, N. R., & Barnett, N. P. (2014). Sustain talk predicts poorer outcomes among mandated college student drinkers receiving a brief motivational intervention. *Psychol Addict Behav, 28*(3), 631-638. doi:10.1037/a0037296

Arkowitz, H., Miller, W. R., Westra, H. A., & Rollnick, S. (2008). Motivational Interviewing in the treatment of psychological problems: Conclusions and future directions. In H. Arkowitz, W. R. Miller, H. A. Westra & S. Rollnick (Eds.), *Motivational Interviewing in the Treatment of Psychological Problems* (pp. 324-341). New York: Guilford.

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *AUDIT - The Alcohol Use Disorders Identification Test* (2nd ed.). Geneva: World Health Organization.

Baer, J. S., Beadnell, B., Garrett, S. B., Hartzler, B., Wells, E. A., & Peterson, P. L. (2008). Adolescent Change Language Within a Brief Motivational Intervention and Substance Use Outcomes. *Psychology of Addictive Behaviors, 22*(4), 570-575. doi:Doi 10.1037/A0013022

Baer, J. S., Rosengren, D. B., Dunn, C. W., Wells, E. A., Ogle, R. L., & Hartzler, B. (2004). An evaluation of workshop training in motivational interviewing for addiction and mental health clinicians. *Drug Alcohol Depend, 73*(1), 99-106.

Barnett, E., Moyers, T. B., Sussman, S., Smith, C., Rohrbach, L. A., Sun, P., & Spruijt-Metz, D. (2014a). From counselor skill to decreased marijuana use: does change talk matter? *J Subst Abuse Treat, 46*(4), 498-505. doi:10.1016/j.jsat.2013.11.004

Barnett, E., Spruijt-Metz, D., Moyers, T. B., Smith, C., Rohrbach, L. A., Sun, P., & Sussman, S. (2014b). Bidirectional Relationships Between Client and Counselor Speech: The Importance of Reframing. *Psychol Addict Behav*. doi:10.1037/a0036227

Bertholet, N., Faouzi, M., Gmel, G., Gaume, J., & Daeppen, J. B. (2010). Change talk sequence during brief motivational intervention, towards or away from drinking. *Addiction, 105*(12), 2106-2112. doi:10.1111/j.1360-0443.2010.03081.x

Blow, F. C., Ilgen, M. A., Walton, M. A., Czyz, E. K., McCammon, R., Chermack, S. T., . . . Barry, K. L. (2009). Severity of baseline alcohol use as a moderator of brief interventions in the emergency department. *Alcohol Alcohol, 44*(5), 486-490. doi:10.1093/alcalc/agp031

Borsari, B., Apodaca, T. R., Jackson, K. M., Mastroleo, N. R., Magill, M., Barnett, N. P., & Carey, K. B. (2014). In-Session Processes of Brief Motivational Interventions in Two Trials With Mandated College Students. *J Consult Clin Psychol*. doi:10.1037/a0037635

Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. *J Consult Clin Psychol, 71*(5), 843-861. doi:10.1037/0022-006X.71.5.843

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. *Arch Intern Med, 158*(16), 1789-1795.

Cicchetti, D. V. (1994). Guidelines, criteria and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Daeppen, J. B., Bertholet, N., Gaume, J., Fortini, C., Faouzi, M., & Gmel, G. (2011). Efficacy of brief motivational intervention in reducing binge drinking in young men: A randomized controlled trial. *Drug Alcohol Depend, 113*(1), 69-75. doi:10.1016/j.drugalcdep.2010.07.009

Dartfish. (2006). *Dartfish Software TeamPro 4.0*. Fribourg, Switzerland: Dartfish Ltd.

Dunn, C., Deroo, L., & Rivara, F. P. (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review. *Addiction, 96*(12), 1725-1742. doi:10.1080/09652140120089481

Gache, P., Michaud, P., Landry, U., Accietto, C., Arfaoui, S., Wenger, O., & Daeppen, J. B. (2005). The Alcohol Use Disorders Identification Test (AUDIT) as a screening tool for excessive drinking in primary care: reliability and validity of a French version. *Alcohol Clin Exp Res, 29*(11), 2001-2007.

Gaume, J., Bertholet, N., Faouzi, M., Gmel, G., & Daeppen, J. B. (2013). Does change talk during brief motivational interventions with young men predict change in alcohol use? *J Subst Abuse Treat, 44*(2), 177-185. doi:10.1016/j.jsat.2012.04.005

Gaume, J., Gmel, G., & Daeppen, J. B. (2008). Brief alcohol interventions: do counsellors' and patients' communication characteristics predict change? *Alcohol Alcohol, 43*(1), 62-69. doi:10.1093/alcalc/agm141

Gaume, J., Magill, M., Longabaugh, R., Bertholet, N., Gmel, G., & Daeppen, J. B. (2014). Influence of Counselor Characteristics and Behaviors on the Efficacy of a Brief Motivational Intervention for Heavy Drinking in Young Men-A Randomized Controlled Trial. *Alcohol Clin Exp Res, 38*(7), 2138-2147. doi:10.1111/acer.12469

Gaume, J., Magill, M., Longabaugh, R., Bertholet, N., Gmel, G., & Daeppen, J. B. (*in press*). Under what conditions? The role of change talk in brief motivational intervention with non-treatment seeking young adults. *J Consult Clin Psychol*.

HABITS Lab. (2015). URICA – Readiness Score. Calculating And Understanding The Readiness Score. Baltimore, MD: University of Maryland, Baltimore County. Retrieved from http://habitslab.umbc.edu/urica-readiness-score. Last accessed July 23, 2015.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: The Guilford Press.

Hettema, J., Steele, J., & Miller, W. R. (2005). Motivational interviewing. *Annu Rev Clin Psychol, 1*, 91-111. doi:10.1146/annurev.clinpsy.1.102803.143833

Hodgins, D. C., Ching, L. E., & McEwen, J. (2009). Strength of commitment language in motivational interviewing and gambling outcomes. *Psychol Addict Behav, 23*(1), 122-130. doi:10.1037/a0013010

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Magill, M., Gaume, J., Apodaca, T. R., Walthers, J., Mastroleo, N. R., Borsari, B., & Longabaugh, R. (2014). The Technical Hypothesis of Motivational Interviewing: A Meta-Analysis of MI's Key Causal Model. *J Consult Clin Psychol*. doi:10.1037/a0036833

Magill, M., & Longabaugh, R. (2013). Efficacy combined with specified ingredients: a new direction for empirically supported addiction treatment. *Addiction, 108*(5), 874-881. doi:10.1111/add.12013

Miller, W. R. (2000). Motivational interviewing: IV. Some parallels with horse whispering. *Behavioural and Cognitive Psychotherapy, 28*, 285-292.

Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. C. (2008). Manual for the Motivational Interviewing Skill Code (MISC). Albuquerque, NM: University of New Mexico. Retrieved from http://casaa.unm.edu/download/misc.pdf. Last accessed July 23, 2015.

Miller, W. R., & Rollnick, S. (2004). Talking Oneself Into Change: Motivational Interviewing, Stages of Change, and Therapeutic Process. *Journal of Cognitive Psychotherapy, 18*(4), 299-308.

Miller, W. R., & Rollnick, S. (2013). *Motivational Interviewing: Helping People Change* (3rd ed.). New York, NY: Guilford Press.

Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *Am Psychol, 64*(6), 527-537. doi:10.1037/a0016830

Moyers, T. B., Martin, T., Houck, J. M., Christopher, P. J., & Tonigan, J. S. (2009). From in-session behaviors to drinking outcomes: a causal chain for motivational interviewing. *J Consult Clin Psychol, 77*(6), 1113-1124. doi:10.1037/a0017189

Pirlott, A. G., Kisbu-Sakarya, Y., Defrancesco, C. A., Elliot, D. L., & Mackinnon, D. P. (2012). Mechanisms of motivational interviewing in health promotion: a Bayesian mediation analysis. *Int J Behav Nutr Phys Act, 9*(1), 69. doi:10.1186/1479-5868-9-69

Rehm, J. (1998). Measuring quantity, frequency, and volume of drinking. *Alcohol Clin Exp Res, 22*(2 Suppl), 4S-14S.

Reinert, D. F., & Allen, J. P. (2007). The alcohol use disorders identification test: an update of research findings. *Alcohol Clin Exp Res, 31*(2), 185-199. doi:10.1111/j.1530-0277.2006.00295.x

Smedslund, G., Berg, R. C., Hammerstrom, K. T., Steiro, A., Leiknes, K. A., Dahl, H. M., & Karlsen, K. (2011). Motivational interviewing for substance abuse. *Cochrane Database Syst Rev*(5), CD008063. doi:10.1002/14651858.CD008063.pub2

Soderstrom, C. A., DiClemente, C. C., Dischinger, P. C., Hebel, J. R., McDuff, D. R., Auman, K. M., & Kufera, J. A. (2007). A controlled trial of brief intervention versus brief advice for at-risk drinking trauma center patients. *J Trauma, 62*(5), 1102-1111; discussion 1111-1102. doi:10.1097/TA.0b013e31804bdb26

Vader, A. M., Walters, S. T., Prabhu, G. C., Houck, J. M., & Field, C. A. (2010). The language of motivational interviewing and feedback: counselor language, client language, and client drinking outcomes. *Psychol Addict Behav, 24*(2), 190-197. doi:10.1037/a0018749

Vasilaki, E. I., Hosier, S. G., & Cox, W. M. (2006). The efficacy of motivational interviewing as a brief intervention for excessive drinking: a meta-analytic review. *Alcohol Alcohol, 41*(3), 328-335. doi:10.1093/alcalc/agl016

Table 1. Change talk measures descriptive statistics and interrater reliability

	Mean	SD	Min	Max	ICC
Percent CT+	33.3	25.5	0	100	0.80
Average CT Strength	-0.3	0.4	-1.2	1.5	0.92
Frequencies					
Overall CT+	7.0	7.8	0	44	0.81
CT+1	1.8	2.2	0	10	0.84
CT+2	5.0	6.7	0	37	0.83
CT+3	**0.2**	**0.6**	**0**	**5**	**0.06**
Overall CT-	11.7	8.3	0	42	0.99
CT-1	1.1	1.4	0	6	0.84
CT-2	10.1	7.4	0	36	0.96
CT-3	0.5	1.0	0	6	0.84
Dichotomized					
	N (1+)		% (1+)		K
CT+3 [0 vs. 1+]	**16**		**9.2**		**0.22**
CT-3 [0 vs. 1+]	43		24.7		0.82

Descriptive statistics based on N=174 coded sessions with follow-up measures. Interrater reliability statistics based on N=42 double-coded sessions. Bold characters indicate low interrater agreement. CT, change talk; SD, standard deviation; ICC, intra-class correlation, K, Cohen's kappa.

Table 2. Linear regression models predicting drinking at 3-month follow-up and adjusting for drinking at baseline

	B	SE	t	p	[95%	CI]
Percent CT+	-0.0004	0.002	-0.22	0.83	-0.004	0.003
Average CT Strength	-0.25	0.11	-2.21	0.03	-0.47	-0.03
Frequencies						
Overall CT+	0.01	0.01	1.12	0.26	0.00	0.02
CT+1	0.01	0.02	0.58	0.56	-0.03	0.05
CT+2	0.01	0.01	1.25	0.21	0.00	0.02
CT+3	-0.11	0.07	-1.55	0.12	-0.25	0.03
Overall CT-	0.02	0.01	3.56	<0.001	0.01	0.03
CT-1	0.05	0.03	1.49	0.14	-0.02	0.11
CT-2	0.02	0.01	3.49	0.001	0.01	0.03
CT-3	0.08	0.04	1.81	0.07	-0.01	0.17
Dichotomized						
CT+3 > 0 (reference=0)	-0.33	0.15	-2.16	0.03	-0.63	-0.03
CT-3 > 0 (reference=0)	0.14	0.10	1.32	0.19	-0.07	0.34

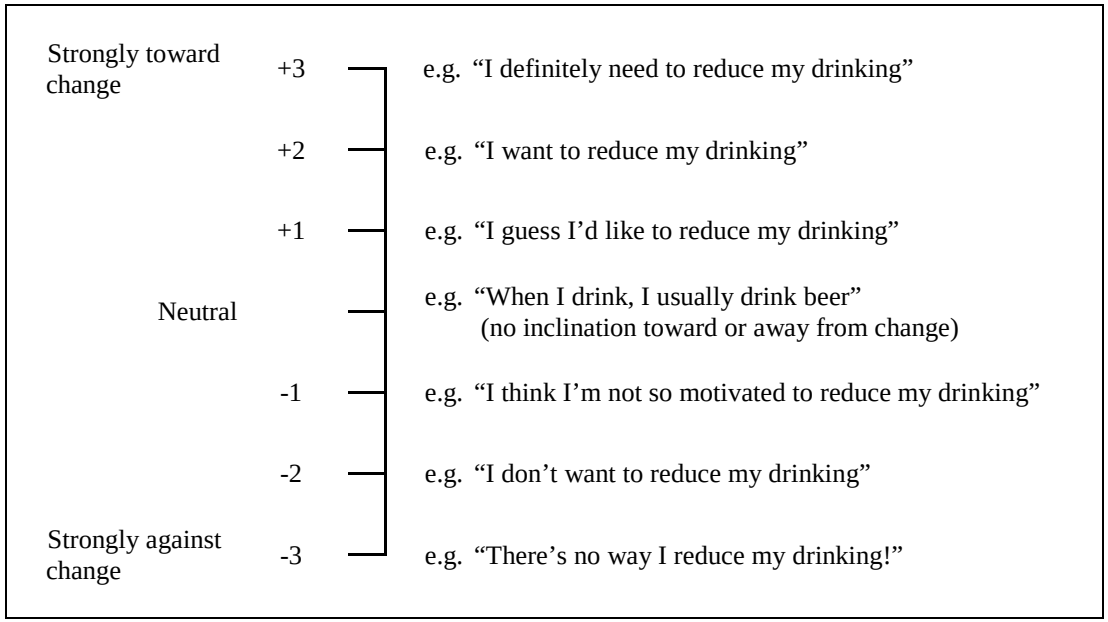
N=174. CT, change talk; SE, standard error; CI, confidence interval. Each variable was tested in a single, separate regression model.

Table 3. Multivariate linear regression model predicting drinking at 3-month follow-up and adjusting for drinking at baseline

	B	SE	t	p	[95%	CI]
CT+1 frequency	0.00	0.02	-0.08	0.94	-0.04	0.04
CT+2 frequency	0.01	0.01	1.93	0.06	0.00	0.03
CT+3 > 0 (reference=0)	-0.49	0.16	-3.02	0.003	-0.80	-0.17
CT-1 frequency	0.02	0.03	0.66	0.51	-0.04	0.08
CT-2 frequency	0.02	0.01	2.77	0.006	0.01	0.03
CT-3 > 0 (reference=0)	0.06	0.10	0.56	0.58	-0.15	0.26

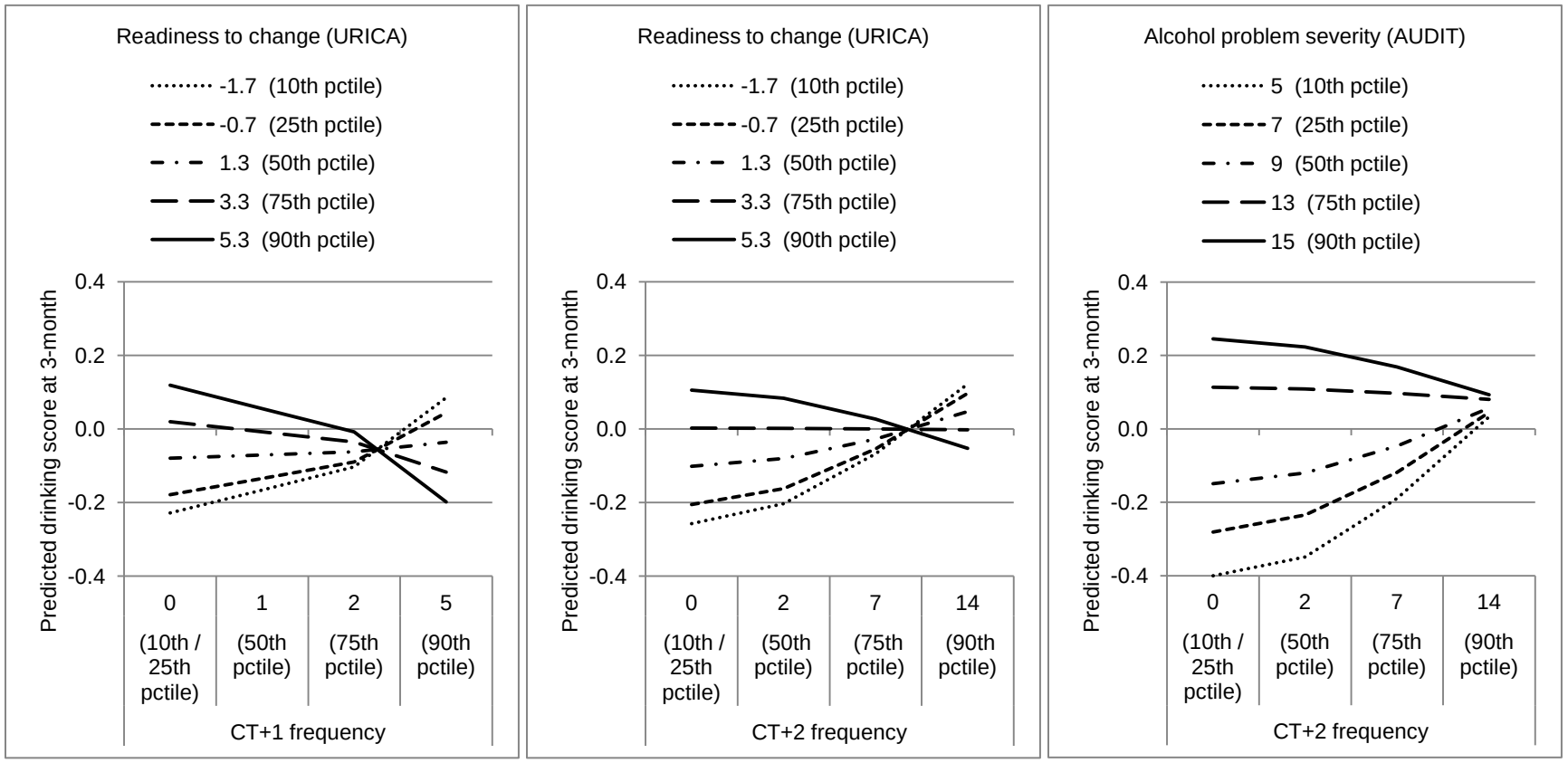
N=174. CT, change talk; SE, standard error; CI, confidence interval. CT+3 and CT-3 are dichotomized variables, one or more utterances (>0) is compared to no utterance (0). Variance inflation factors (VIF) showed no evidence of multicollinearity (CT+1: 1.10; CT+2: 1.25; CT+3>0: 1.18; CT-1: 1.17; CT-2: 1.24; CT-3>0: 1.09; baseline drinking: 1.09).

Figure 1. Change talk strength scale



The sample utterances given therein represent an arbitrary assignment of utterance strengths across possible change talk dimensions. All change talk dimensions (i.e. Reasons; Ability; Desire; Need; Commitment; and Taking steps) were assigned a strength value.

Figure 2. Moderation of the effect of CT strength frequency measures on drinking at 3-month by baseline readiness to change and AUDIT scores (models with significant interactions)



N=174. The covariate in the model (drinking at baseline) was set to its sample mean when deriving the predicted values. Due to right skewed distribution, the 10th and 25th percentiles both are 0. URICA, University of Rhode Island Change Assessment Scale - DELTA Project Reduced Drinking Version; AUDIT, Alcohol Use Disorder Identification Test; pctile, percentile; CT, change talk.