Published in final edited form as:

Alcohol Clin Exp Res. 2017 May ; 41(5): 1054–1062. doi:10.1111/acer.13371.

Temporal Stability of Heavy Drinking Days and Drinking Reductions among Heavy Drinkers in the COMBINE Study

Katie Witkiewitz, PhD,

University of New Mexico

Adam D. Wilson, MS,

University of New Mexico

Matthew R. Pearson, PhD,

University of New Mexico

Kevin A. Hallgren, PhD,

University of Washington

Daniel E. Falk, PhD,

National Institute on Alcohol Abuse and Alcoholism

Raye Z. Litten, PhD,

National Institute on Alcohol Abuse and Alcoholism

Henry R. Kranzler, MD,

University of Pennsylvania

Karl F. Mann, MD,

Central Institute of Mental Health

Deborah S. Hasin, PhD,

Columbia University

Stephanie S. O'Malley, PhD, and

Yale University

Raymond F. Anton, MD

Medical University of South Carolina

Corresponding Author: Katie Witkiewitz, PhD, Department of Psychology, University of New Mexico, MSC 03-2220, Albuquerque NM 87131, katie@unm.edu, (505) 925-2334.

Disclosures. Dr. Kranzler has been a consultant, advisory board member, or CME lecturer for Lundbeck, and Indivior. Dr. Mann received honoraria for consultancies from Pfizer, Novartis, Abbvie and speaker fees from Lundbeck. Dr. Hasin is Principal Investigator of a study funded by inVentiv Health Consulting that combines support from Actavis, Inc.; Endo Pharmaceuticals; Janssen Pharmaceuticals, Inc.; Mallinckrodt, LLC; Pfizer, Inc.; Purdue Pharma, L.P.; Rhodes Pharmaceuticals, L.P.; Roxane Laboratories, Inc.; and Zogenix, Inc. Dr. O'Malley reports personal fees from Alkermes, grants from Lilly, personal fees from Hazelden Betty Ford Foundation, other (study drug) from Pfizer, Astra Zeneca, other (consulting agreement) from Cerecor, Amygdala, outside the submitted work. Dr. Anton has been a consultant in the recent past for Lilly, Lundbeck, Novartis, Indivior, Laboratorio Farmaceutico CT, and Alkermes, and on Advisory Boards for Alkermes, Indivior, and Lundbeck. Drs. Witkiewitz, Kranzler, Mann, Hasin, Falk, Litten, O'Malley, and Anton are members of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative (ACTIVE Group), which over the time that this paper was developed was supported by Abbvie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, and Xenoport. Drs. Wilson, Pearson, Hallgren, Falk, and Litten have no disclosures.

Abstract

Background—Recently, the Food and Drug Administration (FDA) proposed to expand the options for primary endpoints in the development of medications for alcohol use disorder (AUD) to include either abstinence from alcohol or a non-abstinent outcome: no heavy drinking days (with a heavy drinking day defined as more than 3 drinks per day for women and more than 4 drinks per day for men [$>3/>4$ cutoff]). The FDA also suggested that 6 months would be the most appropriate length for a clinical trial to demonstrate the stability of this non-abstinent drinking outcome. However, few alcohol clinical trials have examined the stability of non-heavy drinking during and after treatment.

Methods—In a secondary analysis of the COMBINE study data ($n=1383$), we examined transitions in heavy drinking days during the course of treatment (months 1 through 4), during the transition out of treatment (months 4 through 7), and up to 12 months afterwards (months 13 through 16) using latent variable mixture models.

Results—Heavy drinking and non-heavy drinking were relatively stable in consecutive months (minimum agreement (κ) = .64 for months 1 to 2). Most individuals were stable low-risk drinkers/abstainers or heavy drinkers by the end of treatment, as characterized by a 10% probability (or less) of transitioning out of either a no heavy drinking state or a heavy drinking state. More than two-thirds of the heavy drinkers who exceeded the heavy drinking threshold during treatment reported, on average, a 64% reduction in drinking frequency and a 38% reduction in drinking intensity from pretreatment drinking levels.

Conclusions—The results show stability of no heavy drinking as an outcome within the first four months of treatment and that the $>3/>4$ drink cutoff may mask substantial reductions in alcohol consumption among some patients. Future studies should explore the clinical utility of reduction endpoints.

Keywords

heavy drinking days; alcohol use disorder; alcohol clinical trials; alcohol treatment; primary endpoints for alcohol medications

Introduction

Heavy drinking, defined by the National Institute on Alcohol Abuse and Alcoholism as more than three standard drinks (containing 14 g of pure alcohol) for women and more than four standard drinks for men, is common during and following treatment for alcohol use disorders (AUDs). For example, in the COMBINE study (Anton et al., 2006), 70.9% of individuals across treatment groups experienced at least one heavy drinking day and 29.1% were stable non-heavy drinkers or abstainers during the four months of treatment, while during the 12 months following treatment, more than 80% experienced at least one heavy drinking day and fewer than 20% were stable non-heavy drinkers or abstainers. A marker of AUD treatment success, “percent subjects with no heavy drinking days”, has been recommended by the Food and Drug Administration (FDA) as an acceptable alternative to complete abstinence as a primary endpoint for Phase III alcohol clinical trials (FDA, 2015; Falk et al., 2010). An emerging body of literature supports the validity of no heavy drinking

as a useful surrogate for clinically meaningful benefit. For example, studies have examined no heavy drinking days as an efficacy endpoint (Falk et al., 2010) and demonstrated an association between no heavy drinking and long-term outcomes (Kline-Simon et al., 2013).

The FDA currently conceptualizes “no heavy drinking” (i.e., either abstinence or low-risk drinking [i.e., drinking below the >3/>4 definition]) throughout a clinical trial as indicative of treatment success. However, recent research has shown that heavy drinkers are heterogeneous in their post-treatment psychosocial functioning and that some treatment “failures” who engage in limited heavy drinking report psychosocial functioning that is similar to individuals with abstinent or non-heavy drinking outcomes during treatment (Wilson et al., 2016; Witkiewitz et al., 2017). The FDA guidance also recommends a trial length of at least 6 months (FDA, 2015) to ensure the stability of the observed outcomes. In the COMBINE study, which was a 4-month treatment period, responder analyses suggested that individuals with some heavy drinking days may still have had a positive treatment response to naltrexone, i.e., that they had fewer heavy drinking days than those who received placebo medication (Falk et al., 2014). Thus, an analysis of the patterns of heavy drinking in alcohol clinical trial participants is critical to determine whether, when, and to what degree heavy drinking stabilizes as persistent heavy drinking or as stable non-heavy drinking.

The aims of the current study were three fold: (1) to gain a better understanding of the rates and stability of heavy drinking among individuals with AUD enrolled in the COMBINE study (Anton et al., 2006) across three different clinically relevant time frames: during treatment (months 1 to 4), the transition from receiving treatment to not receiving treatment (months 4 to 7), and the last four months of the 1-year follow-up (months 13 to 16); (2) to determine whether individuals with heavy drinking days have similar drinking outcomes (e.g., percent drinking days, drinks per drinking day) during treatment; and (3) to assess changes in drinking from baseline to the end of treatment among individuals who are considered treatment non-responders, i.e., they had at least one heavy drinking day.

Method

Participants and Procedures

As described elsewhere (Anton et al., 2006), the COMBINE study (“Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence;” COMBINE Study Group, 2003) randomized 1,383 subjects from 11 research sites into 9 treatment groups, consisting of a combination of medical management (MM) or combined behavioral intervention (CBI) and medications (acamprosate, naltrexone, or placebo versions of each drug). Subjects received treatment for a total of 16 weeks; participants were offered 9 MM visits and a maximum of 20 CBI sessions. They completed assessments at 10, 36, and 52 weeks following treatment. In the present study, patients from all treatment modalities were included in the analyses because we were interested in patterns and stability of drinking changes during treatment, regardless of treatment assignment. Sensitivity analyses were conducted in which the naltrexone ($n = 614$) and the CBI ($n = 619$) groups, which received treatments that were effective in the primary analyses (Anton et al., 2006), were excluded. Exclusion of the naltrexone or CBI groups did not greatly alter the conclusions, thus the

results from all treatment groups are described below in detail; the results with naltrexone or CBI excluded are available from the first author.

The sample was recruited from treatment referrals at the study sites and advertising in the community. Key inclusion criteria included a diagnosis of alcohol dependence based on the Diagnostic and Statistical Manual for Mental Disorders, fourth edition (American Psychiatric Association, 1994), age 18 years or older, abstinence from alcohol at least 4 days prior to randomization, a minimum weekly average of 14 drinks (females) or 21 drinks (males) over a successive 30 days in the 90-day period prior to initiating abstinence, and at least two days of heavy drinking (defined as 4+ drinks for females and 5+ drinks for males) in the 90-day period, with the last drink within 21 days of enrollment. Individuals were excluded if they (1) were dependent on a drug other than alcohol, nicotine, or cannabis, (2) recently used opioids, (3) had a serious mental illness, (4) had any other medical condition that could disrupt study participation, (5) had taken one of the study medications during the 30 days prior to baseline, or (6) took medication that could increase the potential risks of the study

Of the 1,383 participants in COMBINE, 31% were female and 69% were male and nearly a quarter were racial/ethnic minorities (76.3% Non-Hispanic White, 11.6% Hispanic American, 7.8% African American, and 4.1% “other”). The participants' mean age was 44.4 years (SD = 10.2), 71% had at least 12 years of education, and 42% were married. Complete drinking data were available from 94% of all participants during treatment (see missing data handling below). At baseline, participants in COMBINE were drinking on 78.6% of days (SD=22.5%), with most of these days considered heavy drinking days (Mean = 67.8%; SD=28.0%), and an average of 11.0 drinks per drinking day (SD=6.8).

Measures

The occurrence of one or more heavy drinking days versus no heavy drinking days by month was our primary binary outcome variable. Alcohol consumption was assessed using a calendar method with the Form-90 Interview (Miller, 1996). Daily alcohol consumption was assessed monthly during treatment and at assessments conducted at 10 weeks, 36 weeks, and 52 weeks following treatment. Heavy drinking days were defined as days on which participants drank more than 4 standard drinks (“4+”) for women and more than 5 standard drinks (“5+”) for men, using the NIAAA definition of a standard drink (one that contains 14 grams of alcohol). Although the FDA and NIAAA define heavy drinking as more than 3 drinks for women (“>3”) and more than 4 drinks per men (“>4”), we defined heavy drinking by the whole drink equivalents of 4 or more for women and 5 or more for men, which is consistent with many clinical trials (Maisto et al., 2016; National Institute on Alcohol Abuse and Alcoholism, 2005). An inspection of the standard drinks consumed on a random sample of days during each month indicated that less than 3% of the COMBINE sample consumed between 3 and 4 standard drinks among women or between 4 and 5 standard drinks among men. Thus, the results using the 4+/5+ definition in this study would be expected to be similar to those obtained had the >3/>4 cut-offs been used.

Each month in which a heavy drinking day occurred was coded as positive (1); each month with no heavy drinking days was coded as negative (0). We examined three four-month time

periods: during treatment (months 1 to 4), the transition from in-treatment (month 4) to post-treatment (months 5 to 7), and during the last four months of the 1-year post-treatment follow-up (months 13 to 16). These four-month periods were examined because the treatment duration was four months, and we wished to use periods of equivalent duration during the transition out of treatment and at follow-up. Also, the models to be estimated, described below, could not accommodate the full duration of the period from months 1 to 16 and thus we estimated the models in four-month periods.

We then computed secondary drinking outcome measures defined by frequency (percent drinking days, PDD) and intensity of drinking (drinks per drinking day, DDD), among individuals who had at least one heavy drinking day. These enabled us to characterize the frequency and intensity of drinking being reported among individuals who would be considered treatment failures because they had at least one heavy drinking day in a given month.

Statistical Analyses

A combination of descriptive and inferential analyses was used to examine stability and change in the occurrence of heavy drinking across time in COMBINE. Descriptive analyses using frequencies and crosstabs were conducted to determine the observed monthly prevalence of heavy drinking and no heavy drinking during each period. A Cohen's kappa statistic (Cohen, 1960) was calculated to assess the degree of stability (i.e., chance-corrected agreement) in observed heavy drinking and no heavy drinking across time, where kappa = 0 would indicate no stability in the observed drinking status across time and kappa = 1 would indicate perfect stability in the observed drinking status from one month to the next month. Kappa represents chance-corrected agreement, with a value greater than .61 defined as substantial agreement (i.e., stability in this case) and kappa greater than .81 defined as nearly perfect agreement (Landis and Koch, 1977). At this descriptive stage, we used observed data without imputing missing values.

The aims of inferential analyses were (1) to gain a better understanding of the patterns of heavy drinking among individuals with AUD; (2) to determine whether those with heavy drinking days are homogeneous (i.e., come from the same population) or heterogeneous in terms of their drinking outcomes (e.g., percent drinking days, drinks per drinking day) during treatment; and (3) to assess the degree of drinking changes from baseline to the end of treatment among individuals with one or more heavy drinking days.

Latent variable mixture modeling was the primary analysis technique for all models. Mixture modeling is a useful tool for accommodating non-normality of drinking outcomes at a given time point and discontinuous changes in drinking over time (Lanza and Collins, 2006; Muthén and Muthén, 2000; Witkiewitz et al., 2010) by estimating a non-normal distribution with a certain number of separate distributions (i.e., latent classes), which are hypothesized to be meaningful sub-populations of individuals with similar outcomes. Model building and selection were based on prior research, theoretical considerations, and empirical indices of model fit. All models were estimated using *Mplus* version 7.3 (Muthén and Muthén, 2012) using maximum likelihood estimation with robust estimation of standard errors (Yuan and Bentler, 2010) and a logit link function. Missing data were accommodated

via maximum likelihood estimation procedures, which provide an estimate of the variance-covariance matrix given all available data. Maximum likelihood estimation is an effective approach for handling data that are missing at random (Enders, 2010; Graham, 2009; Hallgren and Witkiewitz, 2013; Witkiewitz et al., 2014). Models with differing numbers of classes across models were compared using the Bayesian Information Criterion (BIC; Schwarz, 1978), a preferred index for determining the number of classes in longitudinal mixture models (Nylund et al., 2007). The model with the lowest BIC within a group of comparison models was selected. An entropy statistic, which reflects the probability of accurate classification within latent classes from 0.0 (0% accurate) to 1.00 (100% accurate), was examined to evaluate classification precision.

Transitions in Heavy Drinking: Latent Markov Models of Heavy Drinking Days (HDD)

To gain a better understanding of the rates and patterns of heavy drinking among individuals with AUD enrolled in the COMBINE study across three different clinically relevant time frames we used latent Markov models (LMMs) to examine the probability of transitioning between latent classes across time (months during each period). LMMs can be used to examine a finite number of unobserved (i.e., latent) states and the probabilities of remaining in the same latent state or shifting to a different latent state across time (Böckenholt, 2005). LMMs were used to estimate the probability of heavy drinking during the next month, given heavy drinking or non-heavy drinking during the prior month, as well as the probability of stability and change in heavy drinking across months in each time period. Ultimately, we were most interested in the rates of transitioning (indexed by the transitional probabilities [τ parameters])) between latent heavy drinking and non-heavy drinking classes.

We estimated three sets of models for each dataset: *during treatment* (occurrence of heavy drinking days across the four months of treatment in COMBINE), *transition out of treatment* (occurrence of heavy drinking days across the last month of treatment [month 4] and three months following treatment [months 5-7] in COMBINE), and *following treatment* (occurrence of heavy drinking days over the past four months at the 12-month follow-up in COMBINE). For each set of models, we found that the models with two latent classes per time point, described in detail for each time period below, provided a better fit based on lower BIC than models with three, four, or five latent classes, and the classification precision (entropy) was also highest for the 2-class model. For example, during treatment, the BIC of the 2-class model was 5668.4 and BIC increased with each additional class (3-class BIC=5759.3; 4-class BIC=5902.0; 5-class BIC=6089). Entropy was 0.83 for the 2-class model (indicating acceptable classification precision; Nylund et al., 2007) and decreased with each additional class (3-class entropy=.781; 4-class entropy=.712; 5-class entropy=.647).

Model difference testing across alternative nested models was conducted using the Satorra-Bentler scaled χ^2 difference test. Specifically, we conducted tests of invariance across time for two parameters of the latent Markov model: the measurement parameters (ρ), which index the probability of heavy drinking conditional on latent class membership at each time point; and the transition probabilities (τ), which index the probability of expected class membership at time t+1 given expected class membership at time t. For all models,

described below, the scaled χ^2 difference test indicated that the measurement parameters (ρ) could be constrained to be invariant across time without sacrificing model fit (all $ps > 0.06$), such that latent states at each time point could be constrained to be equivalent. The models did differ on the invariance of the τ parameters, as reported in detail below. Based on prior work (Witkiewitz, 2008; Witkiewitz et al., 2010), we hypothesized that the probability of transitioning would be greatest during treatment and in the months following treatment, whereas the rates of transitioning would decrease at the 12-month follow-up when heavy drinking (or non-heavy drinking) states may become more stable. Covariates and treatment group were not considered in the analyses because we were most interested in studying the patterns of stability and change in heavy drinking across time, irrespective of treatment and other covariates. Sensitivity analyses conducted with demographic variables (e.g., sex, race, education) and treatment group included as covariates showed that none of the results changed substantively when covariates were included.

Homogeneity of Heavy Drinking Population: Latent Profile Models of Drinking Frequency and Intensity (PDD and DDD) Among Heavy Drinkers During Treatment

Latent profile models were used to examine contemporaneous drinking outcomes (percent drinking days [PDD] and drinks per drinking day [DDD]) among individuals who engaged in heavy drinking during each month of treatment. Specifically, the PDD and DDD variables were used as indicators in latent profile models (analogous to a latent class analysis) during treatment. The same process of model testing was conducted with the BIC, entropy, the Lo Mendell Rubin Likelihood ratio test (LMR), and the substantive interpretation of latent classes as the primary methods to select the final model within each month. The LMR, which provides a statistical test of the improvement in model fit by each additional class, was used to test the null hypothesis that individuals who engaged in at least some heavy drinking come from a single population (i.e., a one-class model provides the best fit to the data) and were homogeneous with respect to their drinking outcomes. Specifically, we used the LMR test to compare models with an increasing number of latent classes, with each latent class theoretically characterizing different subpopulations of average PDD and DDD. If the LMR rejected the 1-class solution, then alternative class solutions were examined using the BIC and LMR to identify the number of latent classes within each treatment month. The LMR test is available for mixture models with only one latent class variable (e.g., the latent profile model) and is not available when there is more than one latent class variable (e.g., the latent Markov model), thus the LMR test was used only to compare latent profile models and not to estimate the latent Markov models.

Drinking Reductions from Baseline to the End of Treatment among Heavy Drinkers: Latent Profile Models of Drinking Frequency and Intensity (PDD and DDD) Among Heavy Drinkers During the Last Month of Treatment

Assuming heterogeneity of the latent classes, we compared the baseline PDD and DDD among individuals who were most likely classified in each of the latent classes using the latent profile model from the end of treatment (month 4). The goal of this analysis was to evaluate how individuals changed their drinking from baseline to the end of the treatment period based on latent drinking class membership during the last month of treatment. Mean differences in baseline PDD and DDD were estimated using a model-based approach (the

BCH method; Asparouhov & Muthén, 2014) in which the average class-specific means of the baseline PDD and DDD variables were estimated with classification error of the model taken into account in estimating the weighted means by latent class membership in the end of treatment latent profile model.

Results

Descriptive Statistics

The observed (non-imputed) prevalence of heavy drinking in a given month in the sample was consistent across the 16 months, averaging 51.7% (SE = 1.51%; Table 1). The observed cumulative prevalence of any heavy drinking, also shown in Table 1, increased over time with the largest increases in cumulative number of any heavy drinking occurring during the first 3 months of treatment. Specifically, as shown in the third column of Table 1, across the entire 16-month period 1113 (85.3%) individuals in the COMBINE sample engaged in any heavy drinking and 897 (66.7%) engaged in heavy drinking by month 3. In other words, 80.6% of those who ever engaged in heavy drinking (i.e., 897/1113) across the entire 16-month period, engaged in heavy drinking by month 3.

The observed associations of non-imputed heavy drinking between consecutive months across the 16-month follow-up period are shown in Table 2. For example, 84.1% of subjects who reported heavy drinking during the first month of treatment also reported heavy drinking during the second month, whereas 15.9% of the sample reported initial heavy drinking followed by no heavy drinking. Similarly, 76.7% of subjects reporting no heavy drinking in Month 1 remained in that category during Month 2, whereas 23.3% were engaging in heavy drinking in Month 2. In general, as indicated by the estimates of kappa reported in Table 2, the agreement for heavy drinking was substantial (greater than .64) in consecutive months and increased over time to a maximum of kappa = .78 from months 11 to 12, suggesting an increasing stability of the no heavy drinking outcome.

Transitions in Heavy Drinking: Latent Markov Models of Heavy Drinking Days (HDD)

During treatment (months 1-4)—As noted above, the model with two latent classes per time point provided a better fit than models with three, four, or five latent classes, and the classification precision (entropy) was also highest for the 2-class model. The two latent classes could best be described as one class with a high probability of no heavy drinking (probability of no heavy drinking = 0.935) and a second latent class characterized by a high probability of at least some heavy drinking (probability of heavy drinking = 0.940). The class prevalence changed over time with expected classification (based on estimated posterior probabilities) of membership in the no heavy drinking class of 52.3% at month 1, 47.8% at month 2, 49.1% at month 3, and 51.7% at month 4. Note that these results are consistent with the observed categories shown in Table 1, with slight differences due to modeling of probabilities of heavy drinking and no heavy drinking.

Second, we examined whether we could assume measurement invariance of the transition probabilities (τ) across months 1 to 4. The model with τ parameters constrained to equality across months 1 to 4 fit significantly worse (Scaled χ^2 difference (df) = 15.00 (4), $p =$

0.005) than the less restricted model with τ parameters allowed to vary across months 1 to 4, supporting changing transitional probabilities during treatment. Consistent with the descriptive results, the probability of transitioning from non-heavy drinking to heavy drinking decreased over time, with the greatest probability (P) = 0.143 in month 1 to month 2, followed by P = 0.083 in month 2 to month 3, and P = 0.066 in month 3 to month 4. In contrast, the probability of transitioning from heavy drinking to non-heavy drinking increased over time during treatment, with P = 0.063 from months 1 to 2, P = 0.100 from months 2 to 3, and P = 0.114 from months 3 to 4.

Transition out of treatment (months 4-7): Again, the model with two latent classes per time point provided a better fit (i.e., had a lower BIC) than 3-, 4-, or 5-class models. The two latent classes could be characterized by one latent class with a high probability of no heavy drinking (probability of no heavy drinking = 0.944) and a second latent class characterized by a high probability of at least some heavy drinking (probability of heavy drinking = 0.966).

The model τ parameters constrained to equality across months 4 to 7 did not fit significantly worse than the less restricted model with τ parameters allowed to vary across months 4 to 7 (Scaled χ^2 difference (df) = 2.84 (4), p = 0.58), supporting measurement invariance of τ parameters across months 4 to 7. This indicates that the transitions between non-heavy and heavy drinking over time could be constrained to be equal over time without a significant decrement in overall model fit. The latent transition probabilities (which were constrained to be equal over time) showed considerable stability, with a 0.917 probability of remaining classified in the non-heavy drinking class across consecutive months and a 0.941 probability of remaining classified in the heavy drinking class across consecutive months.

12-months following Treatment (months 13-16): The BIC was lowest and the entropy statistic was highest for the 2-class model. The two latent classes could best be described as one class with a high probability of no heavy drinking (probability of no heavy drinking = 0.956) and a second latent class characterized by a high probability of at least some heavy drinking (probability of heavy drinking = 0.979).

The model with τ parameters constrained to equality across months 13 to 16 did not fit significantly worse (Scaled χ^2 difference (df) = 5.03 (4), p = 0.28) than the less restricted model with τ parameters allowed to vary across months 13 to 16, again supporting measurement invariance of τ parameters across time. The latent transition probabilities indicated considerable stability, with a 0.971 probability of remaining classified in the non-heavy drinking class across consecutive months and a 0.947 probability of remaining classified in the heavy drinking class across consecutive months.

Homogeneity of Heavy Drinking Population: Latent Profile Models of Drinking Frequency and Intensity (PDD and DDD) Among Heavy Drinkers During Treatment

The second set of analyses examined drinking frequency (PDD) and intensity (DDD) among subjects with a heavy drinking day in a given month. The results of the LMR likelihood ratio test for the latent profile models for each month showed that two to three latent classes generally provided the best fit during treatment. Importantly, the LMR rejected the one-class

model for all months of treatment (all $p < 0.001$; showing that individuals who engage in heavy drinking come from more than a single population and can thus be described by at least two classes). For all models, the BIC tended to decrease with each additional latent class and the entropy statistic indicated good classification precision across all models. Given substantive interpretation of the resulting class solutions and prior research (Witkiewitz, 2008; Witkiewitz et al., 2010), we selected the models with three latent classes as the final model within each month of treatment.

The probabilities of class membership (expressed as the percentage of heavy drinkers within each class) and average PDD and DDD within each class by month are shown in Table 3. The largest class, Class 1 (which we labeled “moderate intensity, occasional drinking”), had an estimated probability of class membership ranging from .691 (month 1) to .617 (month 4), thus representing more than two-thirds (61.7% to 69.1%) of the sample. Individuals with expected classification in Class 1 drank, on average, six DDD on approximately 24% of days in a month (i.e., seven days per month). The second largest class, Class 2 (which we labeled as “moderate intensity, frequent drinking”), had an estimated probability of class membership ranging from .268 (month 1) to .365 (month 4) and drank, on average, six to seven DDD on approximately 79% to 84% of days in a month (24 to 25 days per month). The smallest class, Class 3 (which we labeled as “high intensity, variable frequency”), had an estimated probability of class membership ranging from .018 (month 4) to .047 (month 2), representing a minority (1.8% to 4.7%) of the sample. This class had very high drinking intensity, with average DDD ranging from 21.1 (month 1) to 28.2 (month 4) and variable drinking frequency, with PDD ranging from 24.6% (month 1) to 77.4% (month 3).

Drinking Reductions from Baseline to the End of Treatment (month 4) among Heavy Drinkers

The average PDD and DDD at baseline by the month 4 latent classes, shown in Figure 1, indicated that Class 2 (moderate intensity, frequent drinking) heavy drinkers, had significantly higher PDD at baseline than either Class 1 heavy drinkers (moderate intensity, occasional drinking; Wald $\chi^2(1) = 148.23, p < 0.001$) or Class 3 heavy drinkers (high intensity, variable frequency; Wald $\chi^2(1) = 9.01, p = 0.003$), with average PDD of 85.6% in Class 2 versus 68.0% in Class 1 and 63.4% in Class 3. Class 3 (high intensity, variable frequency) had significantly higher DDD at baseline than Class 1 (Wald $\chi^2(1) = 21.29, p < 0.001$) and Class 2 (Wald $\chi^2(1) = 23.96, p < 0.001$), with average DDD at baseline of 25.4 DDD in Class 3 versus 11.0 DDD in Class 1 and 10.1 DDD in Class 2.

Class 1 (moderate intensity, occasional drinking) showed the most improvement of the three classes with a 64% reduction in the frequency of drinking (from 68.0% at baseline to 24.3% at Month 4) and a 38% reduction in DDD (from 11.0 drinks at baseline to 6.8 drinks at Month 4) from baseline to the end of treatment. Class 2 (moderate intensity, frequent drinking) had only a 1.1% reduction in the frequency of drinking, but reduced DDD by 25% from baseline to the end of treatment. Class 3 (high intensity, variable frequency) showed some improvement in drinking frequency with a 33% reduction in PDD but a 10% increase in DDD from baseline to the end of treatment.

Discussion

We used data from a large multi-site clinical trial for AUD to investigate the rates and transitions in heavy drinking among AUD patients across three clinically relevant timeframes: during treatment, in the transition period from treatment to post-treatment, and in the last four months of the year following treatment. The Markov modeling approach used in the present study allowed us to examine individual transitions from month to month. As hypothesized, rates of heavy drinking days showed increasing stability over time, with most transitions into or out of heavy drinking occurring during treatment. After the four-month treatment period of COMBINE, the probability of changes in heavy drinking status (reported in Table 2) across consecutive months decreased by almost half, from 16.3% in months 4 to 5 to 9.4% in months 15 to 16.

Further, we sought to determine whether individuals with heavy drinking days could be considered a homogeneous group in terms of their drinking outcomes. Findings from the latent profile models indicated that a one class solution (indicating a homogeneous population) was rejected during each month of treatment, indicating heterogeneity in drinking among the heavy drinking patients and suggesting heavy drinkers are not necessarily a uniform group of treatment non-responders. The vast majority of heavy drinkers at the end of treatment had large reductions in drinking intensity or frequency (or both) from baseline levels. Over 60% of these heavy drinkers were classified as moderate intensity, occasional drinkers (Class 1, Figure 1) reporting, on average, a 64% reduction in drinking frequency (PDD) and a 38% reduction in drinking intensity (DDD) from baseline. Over one-third of these heavy drinkers were classified as moderate intensity, frequent drinkers (Class 2, Figure 1) reporting, on average, a 25% reduction in drinking intensity with no change in drinking frequency from baseline. Finally, a small percentage of heavy drinkers (approximately 1.8%) were classified as high intensity, variable frequency drinkers (Class 3, Figure 1) reporting, on average, a 33% reduction in drinking frequency with a small increase in drinking intensity. Overall, the notion that some heavy drinkers might have clinically meaningful reductions in heavy drinking, as compared to baseline levels, is consistent with recent work demonstrating heterogeneity in psychosocial functioning among heavy drinkers, such that the majority of heavy drinkers with AUD who receive treatment in alcohol clinical trials exhibit psychosocial functioning similar to abstainers and light drinkers following treatment (Wilson et al., 2016; Witkiewitz et al., 2016). This finding has important clinical implications for a harm reduction approach to AUD.

The present study can be viewed as providing a high-definition view of the temporal stability of drinking behavior from month-to-month during treatment, which has previously been assessed in different epidemiological and clinical samples over seven years (Delucchi and Weisner, 2010) and three years (Dawson et al., 2007). Consistent with these prior findings, our analyses showed increasing stability in drinking patterns, but are the first to highlight the relative stability of drinking behavior across four months of treatment, which has important research and clinical implications. Most prior longitudinal studies of alcohol treatment outcomes have examined drinking behavior defined by continuous change in alcohol use (e.g., see Maisto et al., 2014 for a review); however, this approach fails to acknowledge discontinuous change in alcohol use from month-to-month, including the important

transitions from no heavy drinking to heavy drinking and *vice versa*. The FDA has suggested that studies supporting the use of a medication for alcohol treatment be at least six months in duration (FDA, 2015), based partially on the need for the stability of observed drinking reports. Our results provide evidence of marked stability from month-to-month in drinking outcome using the “any heavy drinking” definition during four months of treatment, the length of the COMBINE Study. Clinically, the results suggest that individuals who are doing well during the first three months of treatment have a lower probability of shifting to a heavy drinking outcome. Consistent with other recent analyses of alcohol clinical trial data (Witkiewitz et al., 2017), very high intensity heavy drinking during treatment (as observed in Class 3) might indicate a need to shift strategies or refer the patient to a higher level of care, clearly an indication of treatment failure.

The present study benefitted from a large number of participants across 11 sites nationwide and analyses using both basic descriptive and advanced modeling approaches. One of the primary limitations of the present study was the inability to evaluate whether drinking remains stable over a longer trial (i.e., a 6-month trial, which is currently the FDA recommended length for alcohol clinical trials). An additional limitation was the decision to use 4+/5+ definition of heavy drinking that is commonly employed in clinical trials, rather than the >3/>4 definition provided by the NIAAA and the FDA or additional alternative endpoints. There were few subjects (less than 3%) who were misclassified as non-heavy drinkers using this approach and including these individuals (who were drinking below the 4/5 cutoff, but above the >3/>4 cutoff) as heavy drinkers is unlikely to have a meaningful impact on the results. It is also important to note that most specialty treatment programs in the United States remain abstinence-based. In our analysis of treatment success, abstainers and non-heavy drinkers were considered as one group. Thus, it is unclear how the pattern of results might be different if we considered abstinence as the only marker for success. Because of the complexity of the analyses, it was not feasible to include an examination of an abstinence outcome in this article. It is also unclear whether the findings from the current study would generalize to a treatment that explicitly supports non-abstinence goals and provides training in controlled drinking. For example, selection of a non-abstinence goal in treatment has been associated with greater likelihood of non-heavy drinking, whereas those with abstinence goals in the same treatment were more likely to engage in heavy drinking when drinking occurred (Adamson et al., 2010). Finally, the COMBINE study excluded individuals with severe psychiatric disorders and those with comorbid drug dependence (aside from nicotine), thus the current results may not generalize to individuals with comorbid psychiatric or substance use disorder diagnoses.

Overall, the findings from this study suggest that clinicians and researchers may be able to assess the efficacy of a given treatment after an initial two-to-three month period using no heavy drinking days as a primary outcome, given that most change in heavy drinking occurs within the first three months. Additionally, these findings indicate that there are substantial reductions in drinking from baseline among individuals that we have traditionally labeled as treatment non-responders, based on the no heavy drinking responder definition. The notion that individuals can move in and out of heavy drinking and be functioning quite well is consistent with epidemiological data (Dawson et al., 2007; Delucchi and Weisner, 2010) and recent studies in AUD samples (Wilson et al., 2016; Witkiewitz et al., 2017). On average, the

majority of patients (those in Class 1) show considerable improvements in the frequency and quantity of alcohol consumed during and after treatment, despite being categorized as non-responders by virtue of having had at least one heavy drinking day in a specific period during the clinical trial. This finding suggests that other endpoints need to be considered to evaluate the impact of alcohol treatment. These endpoints could include other conceptualizations of drinking reduction (Hasin et al., in press; Witkiewitz et al., 2017), as well as reductions in non-consumption outcomes (e.g., alcohol-related consequences, quality of life, physical/mental health). Importantly, future work should particularly seek to validate new drinking endpoints by demonstrating that improvement on these endpoints correlates with improvement on other clinically meaningful outcomes of health and social consequences.

Acknowledgments

The current study was funded by the United States National Institute on Alcohol Abuse and Alcoholism (R01AA022328, K01AA023233, T32AA007455). In addition to the authors, the following individuals are or were members of the Alcohol Clinical Trials Initiative (ACTIVE) Workgroup and provided intellectual input into this manuscript during attendance at Workgroup meetings: Joanne Fertig, Ph.D., National Institute on Alcohol Abuse and Alcoholism; Keith Isenberg, M.D., Anthem Blue Cross Blue Shield; David McCann, Ph.D., National Institute on Drug Abuse; Didier Meulien, M.D., Lundbeck SAS; Roger Meyer, M.D., Best Practice Project Management, Inc.; Charles O'Brien, M.D., Ph.D., University of Pennsylvania; Megan Ryan, Ph.D., National Institute on Alcohol Abuse and Alcoholism; Bernard Silverman, M.D., Alkermes, Inc.; Francoise Trinquet, M.D., Ethypharm; Benjamin Zakine, M.D., Ethypharm.

Lindsay Snyder and Sarah Timm (ASCP) provided important administrative support to the ACTIVE workgroup.

References

- Adamson SJ, Heather N, Morton V, Raistrick D. UKATT Research Team. Initial preference for drinking goal in the treatment of alcohol problems: II. Treatment outcomes. *Alcohol Alcohol*. 2010; 45:136–42. [PubMed: 20130150]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th ed; DSM-IV). 4th. Author; Washington, DC: 1994.
- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, Gastfriend DR, Hosking JD, Johnson BA, LoCastro JS, Longabaugh R, Mason BJ, Mattson ME, Miller WR, Pettinati HM, Randall CL, Swift R, Weiss RD, Williams LD, Zweben A. Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. *JAMA*. 2006; 295:2003–17. [PubMed: 16670409]
- Asparouhov T, Muthén BO. Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary second model (No. 21). *Mplus Web Notes*. 2014
- Böckenholt U. A latent Markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions. *Psychol Methods*. 2005; 10:65–83. [PubMed: 15810869]
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960; 20:37–46.
- COMBINE Study Group. Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: rationale and methods. *Alcohol Clin Exp Res*. 2003; 27:1107–22. [PubMed: 12878917]
- Dawson DA, Goldstein RB, Grant BF. Rates and correlates of relapse among individuals in remission from DSM-IV alcohol dependence: a 3-year follow-up. *Alcohol Clin Exp Res*. 2007; 31:2036–45. [PubMed: 18034696]
- Delucchi KL, Weisner C. Transitioning into and out of problem drinking across seven years. *J Stud Alcohol Drugs*. 2010; 71:210–8. [PubMed: 20230718]
- Enders, CK. *Applied missing data analysis*. Guilford Press; New York, NY: 2010.

- Falk DE, Litten RZ, Anton RF, Kranzler HR, Johnson BA, the ACTIVE Workgroup. Cumulative proportion of responders analysis as a tool to assess treatment outcome in alcohol clinical trials. *J Stud Alcohol Drugs*. 2014; 75:335–346. [PubMed: 24650828]
- Falk DE, Wang XQ, Liu L, Fertig J, Mattson M, Ryan M, Johnson B, Stout R, Litten RZ. Percentage of subjects with no heavy drinking days: evaluation as an efficacy endpoint for alcohol clinical trials. *Alcohol Clin Exp Res*. 2010; 34:2022–34. [PubMed: 20659066]
- Food and Drug Administration. Alcoholism: Developing drugs for treatment (No FDA D-0152-001). Silver Spring; MD: 2015.
- Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol*. 2009; 60:549–76. [PubMed: 18652544]
- Hallgren KA, Witkiewitz K. Missing data in alcohol clinical trials: a comparison of methods. *Alcohol Clin Exp Res*. 2013; 37:2152–60. [PubMed: 23889334]
- Hasin DS, Wall M, Witkiewitz K, Kranzler HR, Falk DE, Litten RZ, Mann KF, O'Malley SS, Scodes J, Robinson RK, Anton RF. Change in non-abstinent WHO risk drinking levels and alcohol dependence: A 3-year follow-up study in the United States general population. *Lancet Psychiatry*. in press.
- Kline-Simon AH, Falk DE, Litten RZ, Mertens JR, Fertig J, Ryan M, Weisner CM. Posttreatment low-risk drinking as a predictor of future drinking and problem outcomes among individuals with alcohol use disorders. *Alcohol Clin Exp Res*. 2013; 37(1):E373–80. [PubMed: 22827502]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
- Lanza ST, Collins LM. A mixture model of discontinuous development in heavy drinking from ages 18 to 30: the role of college enrollment. *J Stud Alcohol*. 2006; 67:552–61. [PubMed: 16736075]
- Maisto SA, Kirouac M, Witkiewitz K. Alcohol use disorder clinical course research: How does it inform clinicians' treatment planning? *J Stud Alcohol Drugs*. 2014; 75:799–807. [PubMed: 25208198]
- Maisto SA, Witkiewitz K, Moskal D, Wilson AD. Is the construct of relapse heuristic, and does it advance alcohol use disorder clinical practice? *J Stud Alcohol Drugs*. 2016; 77:849–858. [PubMed: 27797685]
- Miller, WR. Project MA ed. National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 1996. Form 90: A structured assessment interview for drinking and related behaviors.
- Muthén BO, Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcohol Clin Exp Res*. 2000; 24:882–91. [PubMed: 10888079]
- Muthén LK, Muthén BO. Mplus users guide (Version 7). 2012
- National Institute on Alcohol Abuse and Alcoholism. Helping patients who drink too much: A clinician's guide. National Institutes of Health; Bethesda, MD: 2005.
- Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Struct Equ Model*. 2007; 14:535–569.
- Schwarz GE. Estimating the dimension of a model. *Ann Stat*. 1978; 6:461–464.
- Wilson AD, Bravo AJ, Pearson MR, Witkiewitz K. Finding success in failure: using latent profile analysis to examine heterogeneity in psychosocial functioning among heavy drinkers following treatment. *Addiction*. 2016; 111:2145–2154. DOI: 10.1111/add.13518 [PubMed: 27367263]
- Witkiewitz K. Lapses following alcohol treatment: modeling the falls from the wagon. *J Stud Alcohol Drugs*. 2008; 69:594–604. [PubMed: 18612576]
- Witkiewitz K, Falk DE, Kranzler HR, Litten RZ, Hallgren KA, O'Malley SS, Anton RF. Methods to analyze treatment effects in the presence of missing data for a continuous heavy drinking outcome measure when participants drop out from treatment in alcohol clinical trials. *Alcohol Clin Exp Res*. 2014; 38:2826–34. [PubMed: 25421518]
- Witkiewitz K, Hallgren KA, Kranzler HR, Mann KF, Hasin DS, Falk DE, Litten RZ, Anton RF. Clinical validation of reduced alcohol consumption after treatment for alcohol dependence: Using the World Health Organization risk drinking levels. *Alcohol Clin Exp Res*. 2017; 41:179–186. [PubMed: 28019652]

- Witkiewitz K, Maisto SA, Donovan DM. A comparison of methods for estimating change in drinking following alcohol treatment. *Alcohol Clin Exp Res*. 2010; 34:2116–2125. [PubMed: 20860611]
- Witkiewitz K, Roos CR, Pearson MR, Hallgren KA, Maisto SA, Kirouac M, Forcehimes AA, Wilson AD, Robinson CS, McCallion EA, Tonigan JS, Heather N. How much is too much? Patterns of drinking during alcohol treatment and associations with post-treatment outcomes across three alcohol clinical trials. *J Stud Acohol Drugs*. 2017; 78:59–69.
- Yuan KH, Bentler PM. Finite Normal Mixture SEM Analysis by Fitting Multiple Conventional SEM Models. *Sociol Methodol*. 2010; 40:191–245. [PubMed: 21170153]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Baseline and Month 4 Percent Drinking Days (PDD) and Drinks per Drinking Day (DDD) among Classes of Drinking Outcomes in Month 4.

Table 1
Number of Participants With and Without Heavy Drinking Days in each Month

Month	N (%)		N (%)
	Heavy Drinking Days	No Heavy Drinking Days	Cumulative N (%) of Individuals with Any Heavy Drinking Days
1	664 (48.3)	711 (51.7)	664 (48.3)
2	699 (52.5)	632 (47.5)	825 (60.9)
3	660 (50.7)	643 (49.3)	897 (66.7)
4	630 (48.9)	658 (51.1)	951 (70.9)
5	621 (49.4)	636 (50.6)	983 (73.7)
6	646 (52.1)	593 (47.9)	1009 (75.9)
7	651 (52.8)	583 (47.2)	1028 (77.5)
8	607 (51.0)	583 (49.0)	1046 (79.3)
9	607 (51.3)	576 (48.7)	1056 (80.2)
10	624 (52.9)	556 (47.1)	1070 (81.3)
11	620 (52.4)	560 (47.5)	1075 (81.7)
12	633 (54.0)	539 (46.0)	1094 (83.1)
13	626 (53.7)	540 (46.3)	1101 (83.7)
14	576 (51.3)	547 (48.7)	1107 (84.7)
15	568 (51.3)	540 (48.7)	1109 (84.9)
16	558 (50.8)	541 (49.2)	1113 (85.3)

Note. All numbers are observed (percentages are based on valid number of cases) with no imputation for missing data.

Table 2
Observed Associations and Agreement (Kappa) of Heavy Drinking Days (HDD) and No Heavy Drinking Days (NHDD) in Consecutive Months across the 16-Month Follow-up

$\kappa = .64$	Month 2	$\kappa = .66$	Month 3	$\kappa = .65$	Month 4	$\kappa = .65$	Month 5	$\kappa = .69$	Month 6
Month 1	HDD	NHDD	HDD	NHDD	HDD	NHDD	HDD	NHDD	HDD
HDD	84.1%	15.9%	HDD	80.5%	HDD	80.0%	HDD	83.8%	HDD
NHDD	23.3%	76.7%	NHDD	17.4%	NHDD	16.9%	NHDD	16.3%	NHDD
$\kappa = .69$	Month 7	$\kappa = .73$	Month 8	$\kappa = .75$	Month 9	$\kappa = .71$	Month 10	$\kappa = .77$	Month 11
Month 6	HDD	NHDD	HDD	NHDD	HDD	NHDD	HDD	NHDD	HDD
HDD	87.4%	12.6%	HDD	87.9%	HDD	91.5%	HDD	91.4%	HDD
NHDD	15.3%	84.7%	NHDD	9.3%	NHDD	9.9%	NHDD	12.4%	NHDD
$\kappa = .78$	Month 12	$\kappa = .72$	Month 13	$\kappa = .76$	Month 14	$\kappa = .74$	Month 15	$\kappa = .72$	Month 16
Month 11	HDD	NHDD	HDD	NHDD	HDD	NHDD	HDD	NHDD	HDD
HDD	92.8%	7.2%	HDD	89.1%	HDD	87.2%	HDD	90.8%	HDD
NHDD	11.3%	88.7%	NHDD	12.6%	NHDD	8.2%	NHDD	9.8%	NHDD

Note. HDD = Any heavy drinking days; NHDD = No heavy drinking days.

Table 3
Class Specific Percentage of Patients with Expected Classification in Each Class and Means from Latent Profile Models of Drinking Frequency and Intensity (PDD and DDD) among Heavy Drinkers by Treatment Month

	Class 1 Mean “Moderate Intensity, Occasional Drinkers”	Class 2 Mean “Moderate Intensity, Frequent Drinkers”	Class 3 Mean “High Intensity, Variable Frequency”
Month 1	69.1% of heavy drinkers $n \approx 459$	26.8% of heavy drinkers $n \approx 178$	3.9% of heavy drinkers $n \approx 26$
PDD	Mean=24.6%	Mean=79.3%	Mean=24.6%
DDD	Mean=6.438	Mean=6.437	Mean=21.093
Month 2	63.7% of heavy drinkers $n \approx 445$	31.5% of heavy drinkers $n \approx 220$	4.7% of heavy drinkers $n \approx 33$
PDD	Mean=24.0%	Mean=82.3%	Mean=40.8%
DDD	Mean=6.482	Mean=6.785	Mean=23.777
Month 3	64.9% of heavy drinkers $n \approx 428$	32.5% of heavy drinkers $n \approx 215$	2.6% of heavy drinkers $n \approx 17$
PDD	Mean=24.0%	Mean=84.3%	Mean=77.4%
DDD	Mean=6.803	Mean=6.880	Mean=23.905
Month 4	61.7% of heavy drinkers $n \approx 388$	36.4% of heavy drinkers $n \approx 229$	1.8% of heavy drinkers $n \approx 11$
PDD	Mean=24.3%	Mean=84.6%	Mean=42.4%
DDD	Mean=6.824	Mean=7.566	Mean=28.153