



Published in final edited form as:

*Addiction*. 2016 December ; 111(12): 2145–2154. doi:10.1111/add.13518.

## Finding Success in Failure: Using Latent Profile Analysis to Examine Heterogeneity in Psychosocial Functioning among Heavy Drinkers Following Treatment

**Adam D. Wilson,**

Center on Alcoholism, Substance Abuse, & Addictions, University of New Mexico, 2650 Yale Blvd SE, Albuquerque, NM, 87106 USA

**Adrian J. Bravo,**

Department of Psychology, Old Dominion University, 250 Mills Godwin Life Sciences Bldg, Norfolk, VA 23529 USA

**Matthew R. Pearson,** and

Center on Alcoholism, Substance Abuse, & Addictions, University of New Mexico, 2650 Yale Blvd SE, Albuquerque, NM 87106 USA

**Katie Witkiewitz**

Department of Psychology, Center on Alcoholism, Substance Abuse, & Addictions, University of New Mexico, 2650 Yale Blvd SE, Albuquerque, NM 87106 USA

### Abstract

**Aims**—To estimate differences in post-treatment psychosocial functioning among treatment “failures” (i.e., heavy drinkers, defined as 4+/5+ drinks for women/men) from two large multi-site clinical trials, and to compare these levels of functioning to those of the purported treatment “successes” (i.e., non-heavy drinkers).

**Design**—Separate latent profile analyses of data from COMBINE and Project MATCH, comparing psychosocial outcomes across derived classes of heterogeneous treatment responders.

**Setting**—Eleven U.S. academic sites in COMBINE, 27 U.S. treatment sites local to nine research sites in Project MATCH.

**Participants**—962 individuals in COMBINE (69% male, 77% white, mean age: 44 years) treated January 2001 to January 2004 and 1,528 individuals in Project MATCH (75% male, 80% white, mean age: 40 years) treated April 1991 to September 1994.

**Measurements**—In COMBINE, we analyzed health, quality of life, mental health symptoms, and alcohol consequences 12-months post-baseline. In Project MATCH, we examined social functioning, mental health symptoms, and alcohol consequences 15-months post-baseline.

**Findings**—Latent profile analysis of measures of functioning in both samples supported a three-profile solution for the group of treatment “failures,” characterized by high-functioning, average-

---

Correspondence to: Matthew R. Pearson.

Declarations of Interest: None

functioning, and low-functioning individuals. The high-functioning treatment “failures” were generally performing better across measures of psychosocial functioning at follow-up than participants designated treatment “successes” by virtue of being abstainers or light drinkers.

**Conclusions**—Current Food and Drug Administration guidance to use heavy drinking as indicative of treatment “failure” fails to take into account substantial psychosocial improvements made by individuals who continue to occasionally drink heavily post-treatment.

## Introduction

Historically, success in alcohol use disorder (AUD) treatment has been defined by abstinence from alcohol. In more recent decades, researchers have advocated for the use of “low-risk drinking” as a successful endpoint in alcohol treatment (1). Both the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have adopted low-risk drinking as endpoints for pharmaceutical trials (2–4). Currently, the FDA low risk drinking endpoint is defined as no heavy drinking days, with a heavy drinking day defined as consuming 4 or more drinks for a woman, and 5 or more drinks for a man. The 4+/5+ cutoff to distinguish between treatment success and treatment failure is widely used in the field (Maisto et al., in press).

This 4+/5+ heavy drinking definition originated from research on “binge drinking” among college students (5) and critiques of the 4+/5+ “binge” or “heavy” drinking definition can be levied against the very data that led to its widespread adoption (6). First, this same research found a linear relationship between consumption level and the experience of consequences, suggesting the 4+/5+ cutoff was arbitrary. Second, the research focused on college students, thus researchers should correctly be cautious toward applying these principles to clinical populations where “in the context of full-blown alcoholic drinking... five drinks seem comparatively small” (7) (p. 287). Third, reporting that individuals above the cutoff experience more consequences than individuals below the cutoff (5,8,9) is a very weak validity test.

Given the weak validity of any current approach that uses a binary cutoff for consumption as an indicator of success or failure, it makes sense to look directly at the outcome for which consumption currently stands as a proxy: psychosocial functioning. Though areas of life-functioning post-treatment have received some attention as potential outcomes in the literature (10–12); the complexity surrounding what “functioning” is, compared to the simplicity of tallying the number of drinks per day or percentage of days that an individual drinks, has inhibited investigation of treatment effects along these admittedly blurrier lines. However, given that the current edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (13), and all preceding versions of the DSM include no mention of consumption level in the criteria that lead to a diagnosis of AUD, an investigation of outcomes based on psychosocial variables deserves merit.

## Purpose

As currently supported by the FDA, individuals who report any heavy drinking, defined as 4+/5+ drinks per occasion for women/men during a follow-up period, are considered

treatment failures. The use of the “any heavy drinking” endpoint as a surrogate measure of clinical benefit, per recommendations by the FDA (3), is based on unpublished data which showed that drinking above these limits was associated with significantly greater consequences and worse psychosocial functioning. The FDA guidance assumes that all individuals engaged in any heavy drinking will show significant consequences and impairment in functioning. However, to date, no research has examined this assumption. Using data from two of the largest alcohol clinical trials conducted in the United States (U.S.), COMBINE (14) and Project MATCH (15), we used latent profile analysis (LPA) to examine potential heterogeneity among heavy drinkers in terms of psychosocial functioning, and compare their post-treatment functioning to those deemed treatment “successes” (i.e., light drinkers and abstainers). LPA is a person-centered statistical approach that can be used to identify distinct subpopulations based on various indicator variables. In our secondary data analysis, we use several measures of psychosocial functioning as indicators in an attempt to identify distinct groups of individuals who differ in psychosocial functioning at follow-up (i.e., 12- and 15-months post-baseline in COMBINE and Project MATCH, respectively). Given previous studies demonstrating heterogeneity of treatment response (16–18), we hypothesized those individuals who exceeded heavy drinking limits (i.e., treatment “failures”) would be a heterogeneous group defined by subsets of individuals with discrete levels of psychosocial functioning.

## Method

### Participants and Procedure

Data for the present study came from two multisite randomized clinical trials, the COMBINE study (19), and Project MATCH (15). The COMBINE study recruited 1,383 participants from inpatient and outpatient referrals at 11 study sites and in the surrounding communities. Participants were randomly assigned to conditions consisting of some combination of pharmacotherapy, medication management and/or a combined behavioral intervention (CBI). Project MATCH recruited 1,726 participants from 27 treatment sites local to nine U.S. research sites. The participants were randomly assigned to cognitive behavioral therapy, motivational enhancement therapy, or 12-step facilitation.

### Measures

**Alcohol consumption**—In both studies, alcohol consumption was assessed using the Form 90 interview (20), which is a calendar-based measure in which participants report the number of standard drinks they consumed on each day during the response period (e.g., past 90 days). We selected individuals into the “failure” subsample if they reported > 0% heavy drinking days (defined as consuming 4+/5+ standard drinks per day for women/men) during the previous 90 days of the follow-up period (10–12 months post-baseline in COMBINE; 13–15 months post-baseline in Project MATCH).

**Alcohol Consequences**—In both studies, alcohol consequences were assessed using the Drinker Inventory of Consequences (DrInC) (21). In both studies, we used a total score as a global indicator of negative consequences. Reliability and means were similar for

COMBINE ( $M=31.14$ ;  $SD=21.55$ ;  $\alpha=.95$ ) and Project MATCH ( $M=36.77$ ;  $SD=25.02$ ;  $\alpha=.96$ ).

**Health**—In COMBINE, health was assessed using the 12-item SF-12 (22) measured on a 5-response scale. We used the mental health ( $M=-0.47$ ;  $SD=1.05$ ;  $\alpha=.86$ ) and physical health ( $M=0.13$ ;  $SD=0.89$ ;  $\alpha=.86$ ) scores as indicators of overall health.

**Quality of life**—In COMBINE, quality of life was assessed using the 25-item World Health Organization Quality of Life Scale- Brief Version (WHOQOL-BREF) (23) measured on a 5-response scale. We used the physical health ( $M=28.07$ ;  $SD=4.39$ ; 7 items;  $\alpha=.60$ ), psychological ( $M=21.85$ ;  $SD=3.98$ ; 6 items;  $\alpha=.76$ ), social relationships ( $M=10.29$ ;  $SD=2.49$ ; 3 items;  $\alpha=.71$ ), and environment ( $M=30.08$ ;  $SD=5.38$ ; 8 items;  $\alpha=.82$ ) subscales as indicators of overall biopsychosocial functioning.

**Mental health symptoms**—In COMBINE, mental health symptoms were assessed using the 53-item Brief Symptom Inventory (24) measured on a 5-response scale. We used the depression ( $M=58.29$ ;  $SD=11.13$ ; 7 items;  $\alpha=.90$ ), anxiety ( $M=53.61$ ;  $SD=11.19$ ; 6 items;  $\alpha=.85$ ), hostility ( $M=52.13$ ;  $SD=9.90$ ; 5 items;  $\alpha=.80$ ) and interpersonal sensitivity ( $M=54.51$ ;  $SD=10.85$ ; 4 items;  $\alpha=.86$ ) subscales as indicators of psychological functioning. In Project MATCH, mental health symptoms were assessed using the psychiatric severity subscale of the 57-item Addiction Severity Index (ASI) (25) measured on a binary yes/no response scale ( $M=0.16$ ;  $SD=0.21$ ;  $\alpha=.67$ ), and the 21-item Beck Depression Inventory (BDI) measured on a 4-point response scale ( $M=10.44$ ;  $SD=9.26$ ;  $\alpha=.92$ ).

**Social functioning**—In Project MATCH, social functioning was assessed using the 81-item Psychosocial Functioning Inventory (26) measured on a 4-response scale. We used the social behavior role ( $M=3.26$ ;  $SD=0.53$ ; 14 items;  $\alpha=.85$ ) and overall social performance ( $M=3.70$ ;  $SD=0.78$ ; 3 items  $\alpha=.85$ ) subscales as indicators of psychosocial functioning.

## Statistical Analysis

Within two independent samples of treatment “failures,” we conducted latent profile analyses (LPAs) using *Mplus* 7.11 (27) to determine the number of distinct heavy drinking subpopulations based on their psychosocial functioning 12–15 months post-baseline. As recommended by previous research (28,29), we relied on goodness-of-fit indexes, such as the Akaike Information Criterion (AIC) (30) and Bayesian Information Criterion (BIC) (31), as well as tests of statistical significance to settle upon the number of latent classes. Specifically, to determine the number of latent classes across our two analytic samples (i.e., COMBINE and Project MATCH), we used the Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (32,33), which compares whether a  $k$  class solution fits better than a  $k - 1$  class solution.

Further, upon settling on a class solution, we then compared the latent classes to each other and to a group of abstainers/light drinkers using analysis of variance (ANOVA) with Tukey post hoc comparisons. We also tested the equality of means across latent classes on each psychosocial variable using pseudo-class-based multiple imputations (34), and found the same pattern of results (see Tables S1 and S2).

## Results

### COMBINE Sample

Of 1,383 total participants in COMBINE, 962 provided necessary data during the 12-month follow-up period (i.e., approximately 10-, 11-, and 12-months post-baseline) for analysis. This subsample was not significantly different than the full sample on most baseline psychosocial functioning indicators except that the full sample had a slightly higher baseline DrInC score [ $t(1379)=2.50, p=.013, d=.146$ ] and slightly lower baseline depression score [ $t(1379)=-2.01, p=.045, d=.118$ ]. In this subsample, 416 (43.24%) either abstained or drank below the heavy drinking threshold and these individuals would be considered treatment “successes” according to current FDA guidance. However, 546 (56.76%) would be considered treatment “failures” based on reporting heavy drinking (defined as 4+/5+ drinks on any occasion for women/men) during the follow-up period. Latent profile analysis was conducted on these 546 subjects who were categorized as heavy drinkers and provided data on the psychosocial outcome indicators

### COMBINE Latent Profile Analysis

Latent classes were based on the pattern of means of eleven separate psychosocial functioning variables from four separate measures (indicators are in parentheses) at 12 months post-baseline: the SF-12 (physical and mental health), the WHOQOL-BREF (physical health, psychological domain, social relationships domain, and environment domain), the BSI (depression, anxiety, hostility, and interpersonal sensitivity), and the DrInC (total consequences). Higher scores on the SF-12 and WHOQOL-BREF variables indicate healthier psychosocial functioning; while lower scores on the BSI and DrInC variables indicate healthier psychosocial functioning. As can be expected, all SF-12 and WHOQOL-BREF psychosocial variables were significantly positively correlated with each other (except SF-12 mental and SF-12 physical health), and significantly negatively correlated with all BSI and DrInC variables. Further, all BSI and DrInC variables were significantly positively correlated with each other. Within the LPA analytic sample ( $n=546$ ), the Likelihood Ratio Test (LRT) suggested that a 3-class solution fit better than a 2-class solution ( $p=.002$ ) and a 4-class solution did not fit significantly better than a 3-class solution ( $p=.208$ ). Although the AIC and BIC continue to improve (i.e., decrease) for all class solutions (see Table 1), given the LRT, we selected the 3-class solution.

The relative entropy value of .892 indicates that it is estimated that about nine-tenths of subjects were correctly classified in the appropriate latent class, which is considered high classification quality (i.e.,  $>.80$ ) (35).

### Class Comparisons: Comparing “Successes” and “Failures”

Figure 1 depicts the pattern of means across the latent classes, Table 2 summarizes the statistical tests of these differences, and Table S3 summarizes the effect sizes of changes from baseline to follow-up. Scores have been standardized based on the entire COMBINE sample distribution to have a mean of 0 and standard deviation of 1.0, so that positive values are above the mean and negative values are below the mean (a value of  $+1.0=1$  standard deviation above the mean). Class 1 comprised 18.30% of the sample ( $N=99.90$ ), and we

label this class the “Low Functioning” group as they were extremely low on quality of life indicators (i.e., SF-12 and WHOQOL-BREF;  $-1.53 < z_s < -.52$ ), while extremely high on mental health/negative consequences indicators (i.e., BSI and DrInC;  $1.43 < z_s < 1.64$ ). On average, their psychosocial functioning worsened from baseline to follow-up (mean  $d = -.136$ ). Class 2 comprised 37.56% of the sample ( $N=205.06$ ), and we label this class the “Average Functioning” group as they were close to average on quality of life indicators ( $-.54 < z_s < -.01$ ), and close to average on mental health/negative consequences indicators ( $.45 < z_s < .62$ ). On average, their psychosocial functioning improved moderately from baseline to follow-up (mean  $d = .271$ ). Finally, the largest group, Class 3, comprised 44.15% of the sample ( $N=241.03$ ), and we label this class the “High Functioning” group as they were relatively high on quality of life indicators ( $.08 < z_s < .52$ ), and relatively low on mental health/negative consequences indicators ( $-.60 < z_s < -.20$ ). On average, their psychosocial functioning improved substantially from baseline to follow-up (mean  $d = .600$ ). Importantly, with the exception being higher on DrInC scores, the “High Functioning Failures” had significantly better psychosocial functioning on many indicators (12 Mental Health, WHOQUOL-BREF Psychological Domain, BSI Depression, BSI Anxiety, and BSI Interpersonal Sensitivity) and were not significantly different on the other indicators (SF-12 Physical Health, WHOQUOL-BREF Physical Health, Social Relationship Domain, Environmental Domain, and BSI Hostility) compared to “Successes”.

### Project MATCH Sample

Of 1,726 total participants in Project MATCH, 1,528 provided necessary data during the 15-month follow-up period (i.e., approximately 13-, 14-, and 15-months post-baseline) for analysis. This subsample was not significantly different than the full sample on all baseline psychosocial functioning indicators except that the full sample had a slightly higher baseline psychiatric severity score [ $t(1712)=1.987$ ,  $p=.047$ ,  $d=.169$ ]. In this subsample, 716 (46.86%) either abstained or drank below the heavy drinking threshold (treatment “successes”) and 812 (53.14%) would be considered treatment “failures” based on reporting heavy drinking during the follow-up period. Latent profile analysis was conducted on these 812 subjects who were categorized as heavy drinkers and provided data on the psychosocial outcome indicators.

### Project MATCH Latent Profile Analysis

Latent classes were based on the pattern of means of five separate psychosocial functioning variables from four separate measures (indicators are in parentheses) at 15 months post-baseline: the DrInC (total consequences), the ASI (psychiatric severity), the BDI (depression), and the PFI (social role behavior and overall social role performance). Lower scores on the DrInC, ASI, and BDI variables indicate healthier psychosocial functioning, while higher scores on the PFI variables indicate healthier psychosocial functioning. As can be expected, the DrInC, ASI, and BDI psychosocial variables were significantly positively correlated with each other, and significantly negatively correlated with both PFI variables. Within the LPA analytic sample ( $n=812$ ), the LRT suggested that a 3-class solution fit better than a 2-class solution ( $p=.048$ ) and a 4-class solution did not fit significantly better than a 3-class solution ( $p=.109$ ). Although the AIC and BIC continue to improve (i.e., decrease) for all class solutions (see Table 1), given the LRT, we selected the 3-class solution. The relative

entropy value of .821 indicates that it is estimated that about five in six subjects were correctly classified in the appropriate latent class.

### **Class Comparisons: Comparing “Successes” and “Failures”**

Figure 2 depicts the pattern of means across the latent classes, Table 3 summarizes the statistical tests of these differences, and Table S4 summarizes the effect sizes of changes from baseline to follow-up. Again, scores were standardized based on the entire Project MATCH sample so that positive values are above the mean and negative values are below the mean. Class 1 comprised 11.50% of the sample ( $N=93.36$ ), and we label this class the “Low Functioning” group as they were extremely high on the mental health/negative consequences indicators (i.e., DrInC, ASI, and BDI;  $1.37 < z < 2.40$ ), and extremely low on the social functioning indicators (i.e., PFI;  $z = -1.77$ ). On average, their psychosocial functioning worsened from baseline to follow-up (mean  $d = -.404$ ). Class 2 comprised 35.28% of the sample ( $N=286.46$ ), and we label this class the “Average Functioning” group as they were closer to average on mental health/negative consequences indicators ( $.46 < z < .75$ ) and social functioning indicators ( $-.79 < z < -.74$ ). On average, their psychosocial functioning improved modestly from baseline to follow-up (mean  $d = .126$ ). Finally, the largest group, Class 3, comprised 53.22% of the sample ( $N=432.18$ ), and we label this class the “High Functioning” group as they were relatively low on all mental health/negative consequences indicators ( $-.43 < z < -.36$ ), and higher than average on the social functioning indicators ( $.34 < z < .40$ ). On average, their psychosocial functioning improved substantially from baseline to follow-up (mean  $d = .604$ ). The “High Functioning Failures” were functioning at levels that were similar to (PFI Overall Social Role Performance) or significantly better than the “Successes” on almost all indicators (ASI Psychiatric Severity, BDI Depression, and PFI Social Behavior Role), with the only exception being the scores on the DrInC.

### **Supplementary Analyses**

We conducted additional LPA analyses in the full sample in COMBINE and Project MATCH and then distinguished the classes using the binary indicator of “success” versus “failure.” Consistent with the results described above (see Tables S5–S7 and Figures S1–S2), these results demonstrate considerable overlap between “successes” and “failures” in psychosocial functioning. In both samples, about 40% of the individuals in the high functioning group were considered treatment failures, and about 60% of treatment failures were in the average to high functioning groups.

### **Discussion**

Across two independent samples of individuals undergoing treatment for AUD, we find that a substantial portion of these individuals (53.14% – 56.76%) are considered treatment failures according to the current 4+/5+ heavy drinking cutoff. Amongst these so-called treatment failures, we found significant heterogeneity on a variety of psychosocial functioning indicators. Specifically, we found three subgroups of individuals in both the COMBINE and Project MATCH samples. The smallest class demonstrated the worst psychosocial functioning, suggesting that a small minority of these heavy drinkers suffered

the brunt of the negative outcomes attributed to treatment failure. Further, a substantial number of “failures” exhibited relatively healthy psychosocial functioning including few alcohol-related consequences, low mental health symptoms, and high quality of life across multiple domains. Indeed, in both samples, the relatively large high-functioning classes of treatment “failures” had equal or better psychosocial functioning than the sub-sample of treatment “successes” on nearly all outcomes, with the exception of drinking-related consequences.

In the research context, our findings question whether the 4+/5+ heavy drinking endpoint is sensitive to detecting individuals who are actually at increased risk of impaired functioning, or whether some heavy drinkers can still have significant improvements in functioning. These results are consistent with recent research showing that some heavy drinking, particularly heavy drinking in combination with low risk drinking, is associated with similar consequences and health care costs up to 1 and 3 years following treatment as abstinence or low risk drinking outcomes (36,37).

The 4+/5+ heavy drinking cutoff is also less relevant or useful in the context of treatment, whereby clinicians aim to improve the quality of life and psychosocial functioning of their AUD clients. Level of consumption may be a target of many treatments, but more often a clinician is interested in clinical benefit, regardless of whether a client is exceeding a 4+/5+ heavy drinking cutpoint. In the context of a wide range of harm reduction strategies, reducing consumption may be sufficient to reduce harm, but may not always be necessary.

While we do not believe that any consumption-based primary endpoint can reliably delineate treatment “successes” from treatment “failures,” we do suggest alternate approaches that may move the field forward. One possible approach is to measure outcome in terms of a drop in severity of AUD based on DSM-5 (e.g., a drop from AUD – severe to AUD – moderate, a drop from moderate to mild, or a drop from mild to “in remission”). In a similar vein, and in keeping with how the US FDA measures success with respect to interventions for obesity, high blood pressure, and diabetes, outcome could be defined by a clinically meaningful reduction in some indicator of symptomatology (e.g., reductions in alcohol-related problems). In the rare situation that consumption-based cutoffs were deemed necessary, we suggest that the only way to derive truly valid cutoffs for individuals would be to employ an ecological momentary assessment approach or real time objective alcohol monitoring (e.g., transdermal alcohol monitoring, (38)) that can actually map negative consequences to specific levels of episodic drinking.

Despite the strengths of using two large datasets with clinical populations, there are a number of limitations. First, we used different indicators of psychosocial functioning in our LPAs at specific time points and across studies based on availability of the indicators of psychosocial functioning and when they were assessed. Another limitation of this study was the inability to use the calendar-based data to obtain information regarding the duration of drinking episodes. The most recent definition supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA, 2004) defines binge drinking as consuming 4+/5+ drinks for women/men *within a two-hour period*, presumably based on the notion that this leads to a blood alcohol level around 0.08 g/dL. Given that we are unable to account for duration of



drinking episodes, one could argue that our definition loses some of the sensitivity of a time-specific measure. Finally, the current study did not examine various time windows (e.g., 30 days versus 60 days) for evaluating presence or absence of heavy drinking days. Future research could consider varying time windows.

## Conclusion

Overall, our findings call into question the use of the current 4+/5+ heavy drinking definition as it is applied to clinical populations. We believe this application can have negative impacts on how researchers and clinicians define treatment success and treatment failure. Any alcohol consumption based cutoff will provide a single benchmark for all clients irrespective of their baseline use and a host of other factors that may contribute to whether they experience the negative consequences and impairments in psychosocial functioning that are a hallmark of AUD. Further, these cutoffs fail to account for meaningful improvements made by clients who are considered treatment “failures.”

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

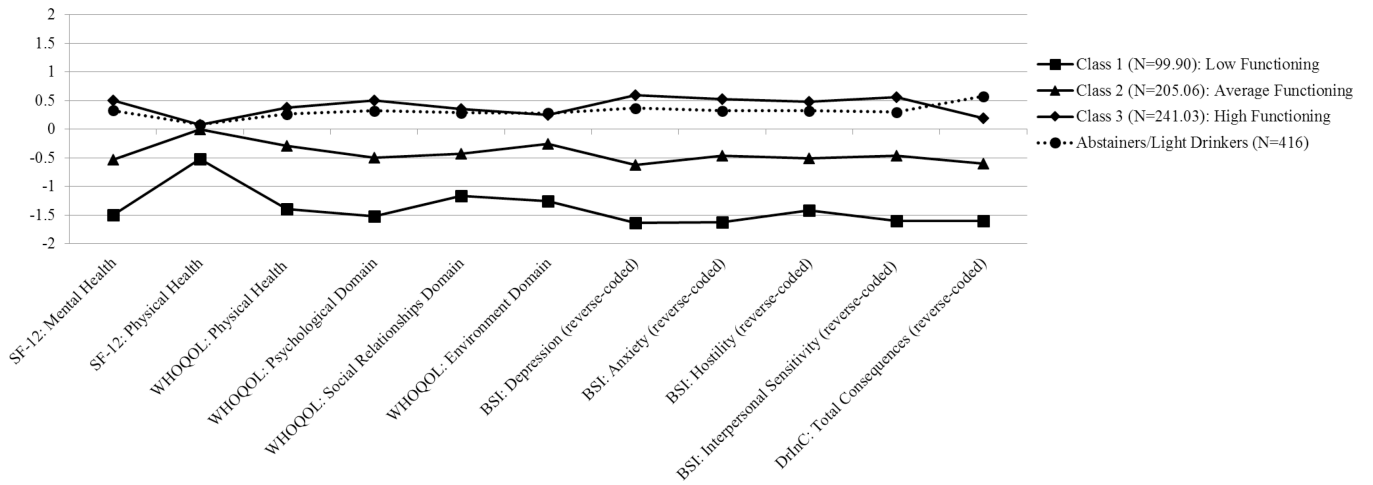
This secondary data analysis was supported by a grant from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) awarded to Dr. Witkiewitz (R01-AA022328). MRP is supported by a career development grant from the NIAAA (K01-AA023233).

## References

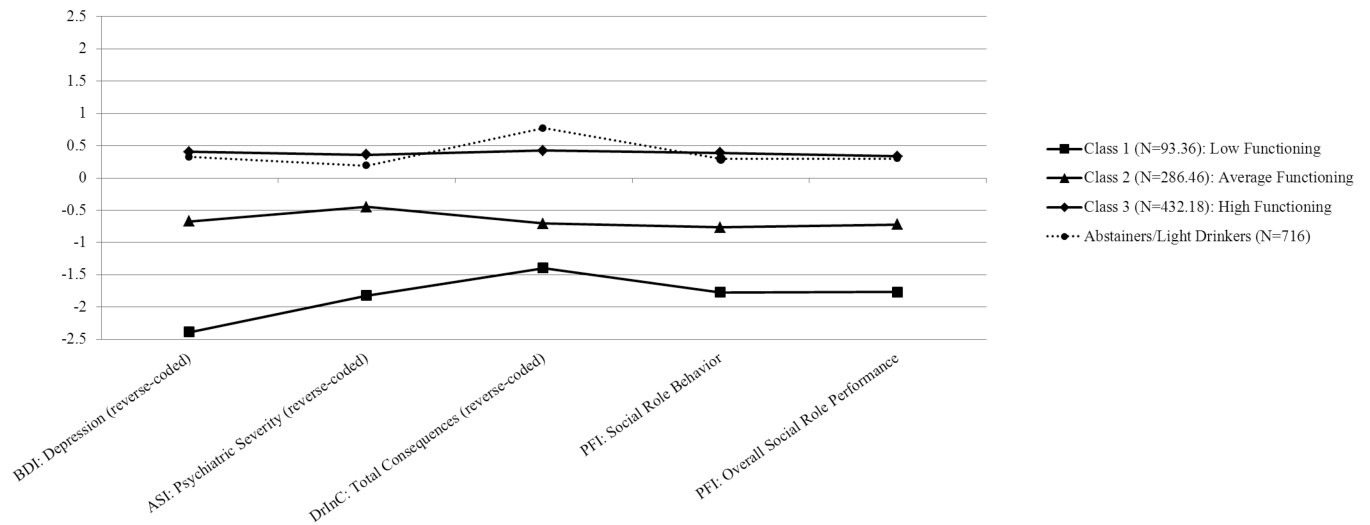
1. Maisto SA, Clifford PR, Stout RL, Davis CM. Moderate drinking in the first year after treatment as a predictor of three-year outcomes. *J Stud Alcohol Drugs*. 2007; 68(3):419–427. [PubMed: 17446982]
2. Administration F and D. Medical Review of Vivitrol: 21–897. Rockville, MD: 2006.
3. Administration F and D. Alcoholism: Developing Drugs for Treatment (No. PDA D-0152-001). Silver Spring, MD: Food and Drug Administration; 2015.
4. European Medicines Agency. Guideline on the development of medicinal products for the treatment of alcohol dependence. 2010 Feb.:1–17.
5. Wechsler H, Dowdall GW, Davenport A, Rimm EB. A gender-specific measure of binge drinking among college students. *Am J Public Health*. 1995; 85(7):982–985. [PubMed: 7604925]
6. Pearson MR, Kirouac M, Witkiewitz K. Questioning the validity of the binge or heavy drinking criterion in college and clinical populations. *Addiction*. 2015
7. Wechsler H, Nelson TF. Binge drinking and the American college student: What's five drinks? *Psychol Addict Behav*. 2001; 15:287–291. [PubMed: 11767258]
8. Wechsler H, Davenport A, Dowdall G, Moeykens B, Castillo S. Health and behavioral consequences of binge drinking in college: A national survey of students at 140 campuses. *J Am Med Assoc*. 1994; 272(1672–1677):1672–1677.
9. Wechsler H, Kuo M. College students define binge drinking and estimate its prevalence: Results of a national survey. *Jounal Am Coll Heal*. 2000 Sep.49:57–64.
10. Cisler RA, Zweben A. Development of a Composite Measure for Assessing Alcohol Treatment Outcome: Operationalization and Validation. *Alcohol Clin Exp Res* [Internet]. 1999; 23(2):263–271. Available from: <http://dx.doi.org/10.1111/j.1530-0277.1999.tb04109.x>.

11. Maisto SA, Kirouac M, Witkiewitz K. Alcohol Use Disorder Clinical Course Research: Informing Clinicians' Treatment Planning Now and in the Future. *J Stud Alcohol Drugs*. 2014; 75(5):799–807. [PubMed: 25208198]
12. Donovan DM, Bigelow GE, Brigham GS, Carroll KM, Cohen AJ, Gardin JG, et al. Primary outcome indices in illicit drug dependence treatment research: Systematic approach to selection and measurement of drug use end-points in clinical trials. *Addiction*. 2012; 107(4):694–708. [PubMed: 21781202]
13. Association, AP. Diagnostic and statistical manual of mental disorders. 5th. Washington, DC: American Psychiatric Association; 2013.
14. Anton RF, O'Malley SS, Ciraulo Da, Cisler Ra, Couper D, Donovan DM, et al. Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence. *J Am Med Assoc*. 2006; 295(17):2003–2017.
15. Group PMR. Project MATCH: Rationale and methods for a multisite clinical trial matching patients to alcoholism treatment. *Alcohol Clin Exp Res*. 1993; 17:1130–1145. [PubMed: 8116822]
16. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol* [Internet]. 2010; 6:109–138. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20192788>.
17. Wilk P, Vingilis E, Bishop JEH, He W, Braun J, Forchuk C, et al. Distinctive trajectory groups of mental health functioning among assertive community treatment clients: An application of growth mixture modelling analysis. *Can J Psychiatry*. 2013; 58(12):670–678. [PubMed: 24331286]
18. Wu J, Witkiewitz K. Network Support for Drinking: An Application of Multiple Groups Growth Mixture Modeling to Examine. 2003
19. Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, et al. Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence. *J Am Med Assoc*. 2006; 295(17):2003–2017.
20. Miller WR, Del Boca FK. Measurement of drinking behavior using the Form 90 family of instruments. *J Stud Alcohol Suppl*. 1994; 12(12):112–118. [PubMed: 7722987]
21. Miller, WR.; Tonigan, JS.; Longabaugh, R. The drinker inventory of consequences (DrInC) [Internet]. The Drinker Inventory of Consequences (DrInC): An Instrument for Assessing Adverse Consequences of Alcohol Abuse. National Institute on Alcohol Abuse and Alcoholism. 1995. Available from: <http://pubs.niaaa.nih.gov/publications/drinc.pdf>
22. Ware JE Jr, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996; 34:220–233. [PubMed: 8628042]
23. WHOQoL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med*. 1998; 28:551–558. [PubMed: 9626712]
24. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. *Psychological medicine*. 1983:595–605. [PubMed: 6622612]
25. McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, et al. The Fifth Edition of the Addiction Severity Index. *J Subst Abuse Treat* [Internet]. 1992; 9:199–213. Available from: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list\\_uids=8634377044579492629http://www.ncbi.nlm.nih.gov/pubmed/1334156](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=8634377044579492629http://www.ncbi.nlm.nih.gov/pubmed/1334156).
26. Feragne MA, Longabaugh R, Stevenson JF. The Psychosocial Functioning Inventory. *Eval Health Prof*. 1983; 6(1):25–48. [PubMed: 10259949]
27. Muthen, LK.; Muthen, BO. Mplus User's Guide. Los Angeles, CA: Muthen & Muthen;
28. Marsh HW, Lüdtke O, Trautwein U, Morin AJS. Classical Latent Profile Analysis of Academic Self-Concept Dimensions: Synergy of Person- and Variable-Centered Approaches to Theoretical Models of Self-Concept. *Structural Equation Modeling: A Multidisciplinary Journal*. 2009:191–225.
29. Henson JM, Reise SP, Kim KH. Detecting Mixtures From Structural Model Differences Using Latent Variable Mixture Modeling: A Comparison of Relative Model Fit Statistics. *Struct Equ Model A Multidiscip J*. 2007; 14(2):202–226.
30. Akaike H. A new look at the statistical model identification. *IEE Trans Autom Control*. 1974; 19:716–723.

31. Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978; 6:461–464.
32. Lo Y, Mendell NR, Rubin DB. Testing the Number of Components in a Normal Mixture. *Biometrika* [Internet]. 2001; 88(3):767–778. Available from: <http://biomet.oxfordjournals.org/content/88/3/767.short\papers2://publication/uuid/82C048A3-1680-4CB7-8914-C352485B7DA8>.
33. Vuong QH. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica.* 1989; 57(2):307–333.
34. Asparouhov T, Muthén BO. Computationally efficient estimation of multilevel high-dimensional latent variable models. *Proc 2007 Jt Stat Meet* [Internet]. 2007:2531–2535. Available from: <http://www.statmodel2.com/download/JSM2007000746.pdf>.
35. Clark SL, Muthen B. Relating latent class analysis results to variables not included in the analysis. 2009 Available from: <http://www.statmodel.com/download/relatinglca.pdf>.
36. Aldridge AP, Zarkin Ga, Dowd WN, Bray JW. The Relationship Between End-of-Treatment Alcohol Use and Subsequent Healthcare Costs: Do Heavy Drinking Days Predict Higher Healthcare Costs? *Alcohol Clin Exp Res* [Internet]. 2016; 40(5):1122–1128. Available from: <http://doi.wiley.com/10.1111/acer.13054>.
37. Witkiewitz K. “Success” Following Alcohol Treatment: Moving Beyond Abstinence. 2013 Jan. 37:9–13.
38. Dougherty DM, Charles NE, Acheson A, John S, Furr RM, Hill-Kapturczak N. Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking. *Exp Clin Psychopharmacol* [Internet]. 2012 Oct; 20(5):373–381. [cited 2013 Mar 30] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22708608>.



**Figure 1.** Depiction of the three latent classes and abstainers/light drinkers sample defined by pattern of standardized means on psychosocial variables in COMBINE. SF-12=Short Form Health Survey, WHOQOL=World Health Organization Quality of Life Assessment-BREF, BSI=Brief Symptom Inventory, DrInC=Drinker Inventory of Consequences. We reversed coded the BSI and DrInC variables such that higher scores indicate healthier psychosocial functioning (e.g., higher scores indicates lower depressive symptoms). Standardized scores were created taking into account abstainers/light drinkers (i.e., whole sample).



**Figure 2.**

Depiction of the three latent classes and abstainers/light drinkers sample defined by pattern of standardized means on psychosocial variables in Project Match. DrINC=Drinker Inventory of Consequences, ASI=Addiction Severity Index, BDI=Beck Depression Inventory, PFI=Psychosocial Functioning Inventory. We reversed coded the BDI, ASI, and DrInC variables such that higher scores indicate healthier psychosocial functioning (e.g., higher scores indicates lower depressive symptoms). Standardized scores were created taking into account abstainers/light drinkers (i.e., whole sample).

**Table 1**

Fit statistics for 1 through 7 class solutions for Latent Profile Analysis (LPA) across COMBINE and Project Match.

Fit Statistics	Number of Classes- COMBINE						
	1	2	3	4	5	6	7
AIC	36068.64	33780.72	<b>33169.90</b>	32963.54	32804.06	32662.03	32571.84
BIC	36163.30	33927.01	<b>33367.82</b>	33213.09	33105.24	33014.84	32976.29
Adjusted BIC	36093.46	33819.08	<b>33221.80</b>	33028.98	32883.03	32754.54	32688.90
Relative Entropy	----	0.926	<b>0.892</b>	0.850	0.874	0.873	0.877
Smallest <i>n</i>	546	221	<b>100</b>	66	34	25	17
LRT	----	<i>p</i> < .001	<b><i>p</i> = .002</b>	<i>p</i> = .208	<i>p</i> = .460	<i>p</i> = .467	<i>p</i> = .543

Fit Statistics	Number of Classes- Project MATCH						
	1	2	3	4	5	6	7
AIC	15297.63	14099.42	<b>13760.28</b>	13590.25	13512.62	13281.97	13188.93
BIC	15344.62	14174.61	<b>13863.67</b>	13721.84	13672.41	13469.95	13405.10
Adjusted BIC	15312.87	14123.80	<b>13793.81</b>	13632.92	13564.44	13342.93	13259.03
Relative Entropy	----	0.837	<b>0.821</b>	0.861	0.841	0.895	0.900
Smallest <i>n</i>	812	270	<b>94</b>	96	17	36	24
LRT	----	<i>p</i> < .001	<b><i>p</i> = .048</b>	<i>p</i> = .109	<i>p</i> = .143	<i>p</i> = .102	<i>p</i> = .792

Note. AIC= Akaike Information Criterion, BIC= Bayesian Information Criterion, LRT=Lo-Mendell-Rubin Adjusted Likelihood Ratio Test. The preferred class solution is highlighted in boldtype face for emphasis.

**Table 2**

Mean comparisons between classes on outcomes at 12 months in COMBINE based on most-likely class membership

	Raw Scores			
	Class 1: Low Functioning “Failures”	Class 2: Average Functioning “Failures”	Class 3: High Functioning “Failures”	Abstainers/ Light Drinkers “Successes”
Sample sizes	97 – 100	202 – 206	232 – 236	412 – 415
<i>SF-12</i> Mental Health	-1.767 <sub>a</sub>	-0.757 <sub>b</sub>	0.502 <sub>c</sub>	0.136 <sub>d</sub>
<i>SF-12</i> Physical Health	-0.260 <sub>a</sub>	0.178 <sub>b</sub>	0.248 <sub>b</sub>	0.252 <sub>b</sub>
<i>WHOQOL-BREF</i> Physical Health	22.878 <sub>a</sub>	26.677 <sub>b</sub>	30.575 <sub>c</sub>	30.104 <sub>c</sub>
<i>WHOQOL-BREF</i> Psychological Domain	16.700 <sub>a</sub>	20.850 <sub>b</sub>	24.861 <sub>c</sub>	24.127 <sub>d</sub>
<i>WHOQOL-BREF</i> Social Relationships Domain	7.903 <sub>a</sub>	9.766 <sub>b</sub>	11.728 <sub>c</sub>	11.560 <sub>c</sub>
<i>WHOQOL-BREF</i> Environment Domain	24.354 <sub>a</sub>	29.858 <sub>b</sub>	32.658 <sub>c</sub>	32.815 <sub>c</sub>
<i>BSI</i> Depression	73.315 <sub>a</sub>	62.192 <sub>b</sub>	48.754 <sub>c</sub>	51.290 <sub>d</sub>
<i>BSI</i> Anxiety	68.977 <sub>a</sub>	56.034 <sub>b</sub>	45.159 <sub>c</sub>	47.410 <sub>d</sub>
<i>BSI</i> Hostility	63.719 <sub>a</sub>	54.766 <sub>b</sub>	45.071 <sub>c</sub>	46.600 <sub>c</sub>
<i>BSI</i> Interpersonal Sensitivity	68.847 <sub>a</sub>	57.048 <sub>b</sub>	46.422 <sub>c</sub>	49.090 <sub>d</sub>
<i>DrInC</i> Total Consequences	57.149 <sub>a</sub>	34.856 <sub>b</sub>	17.210 <sub>c</sub>	8.570 <sub>d</sub>
	Standardized Scores (z-scores)			
<i>SF-12</i> Mental Health	-1.507 <sub>a</sub>	-0.535 <sub>b</sub>	0.516 <sub>c</sub>	0.329 <sub>d</sub>
<i>SF-12</i> Physical Health	-0.519 <sub>a</sub>	-0.008 <sub>b</sub>	0.076 <sub>b</sub>	0.079 <sub>b</sub>
<i>WHOQOL-BREF</i> Physical Health	-1.406 <sub>a</sub>	-0.287 <sub>b</sub>	0.384 <sub>c</sub>	0.265 <sub>c</sub>
<i>WHOQOL-BREF</i> Psychological Domain	-1.533 <sub>a</sub>	-0.482 <sub>b</sub>	0.507 <sub>c</sub>	0.319 <sub>d</sub>
<i>WHOQOL-BREF</i> Social Relationships Domain	-1.181 <sub>a</sub>	-0.425 <sub>b</sub>	0.366 <sub>c</sub>	0.286 <sub>c</sub>
<i>WHOQOL-BREF</i> Environment Domain	-1.265 <sub>a</sub>	-0.266 <sub>b</sub>	0.268 <sub>c</sub>	0.281 <sub>c</sub>
<i>BSI</i> Depression	1.636 <sub>a</sub>	0.618 <sub>b</sub>	-0.604 <sub>c</sub>	-0.364 <sub>d</sub>
<i>BSI</i> Anxiety	1.628 <sub>a</sub>	0.453 <sub>b</sub>	-0.529 <sub>c</sub>	-0.319 <sub>d</sub>
<i>BSI</i> Hostility	1.426 <sub>a</sub>	0.503 <sub>b</sub>	-0.485 <sub>c</sub>	-0.320 <sub>c</sub>
<i>BSI</i> Interpersonal Sensitivity	1.593 <sub>a</sub>	0.462 <sub>b</sub>	-0.562 <sub>c</sub>	-0.299 <sub>d</sub>
<i>DrInC</i> Total Consequences	1.577 <sub>a</sub>	0.604 <sub>b</sub>	-0.196 <sub>c</sub>	-0.575 <sub>d</sub>

Note. SF-12=Short Form Health Survey, WHOQOL-BREF=World Health Organization Quality of Life Assessment-BREF, BSI=Brief Symptom Inventory, DrINC=Drinker Inventory of Consequences. Means sharing a subscript in a row indicate means that are not significantly different from each other based on Tukey post-hoc comparisons. Standardized scores were created in the full sample.

**Table 3**

Mean comparisons between classes on outcomes at 15 months in Project MATCH based on most-likely class membership

Raw Scores				
	Class 1: Low Functioning “Failures”	Class 2: Average Functioning “Failures”	Class 3: High Functioning “Failures”	Abstainers/ Light Drinkers “Successes”
Sample Sizes:				
<i>DrInC</i> Total Consequences	66.859 <sub>a</sub>	51.151 <sub>b</sub>	21.084 <sub>c</sub>	12.362 <sub>d</sub>
<i>ASI</i> Psychiatric Severity	0.480 <sub>a</sub>	0.216 <sub>b</sub>	0.062 <sub>c</sub>	0.094 <sub>d</sub>
<i>BDI</i> Depression	28.065 <sub>a</sub>	13.755 <sub>b</sub>	4.490 <sub>c</sub>	5.211 <sub>d</sub>
<i>PFI</i> Social Behavior Role	2.527 <sub>a</sub>	3.003 <sub>b</sub>	3.585 <sub>c</sub>	3.536 <sub>d</sub>
<i>PFI</i> Overall Social Role Performance	2.581 <sub>a</sub>	3.350 <sub>b</sub>	4.158 <sub>c</sub>	4.127 <sub>c</sub>
Standardized Scores (z-scores)				
<i>DrInC</i> Total Consequences	1.367 <sub>a</sub>	0.751 <sub>b</sub>	-0.429 <sub>c</sub>	-0.771 <sub>d</sub>
<i>ASI</i> Psychiatric Severity	1.861 <sub>a</sub>	0.456 <sub>b</sub>	-0.362 <sub>c</sub>	-0.190 <sub>d</sub>
<i>BDI</i> Depression	2.395 <sub>a</sub>	0.692 <sub>b</sub>	-0.410 <sub>c</sub>	-0.325 <sub>d</sub>
<i>PFI</i> Social Behavior Role	-1.766 <sub>a</sub>	-0.793 <sub>b</sub>	0.395 <sub>c</sub>	0.295 <sub>d</sub>
<i>PFI</i> Overall Social Role Performance	-1.769 <sub>a</sub>	-0.740 <sub>b</sub>	0.339 <sub>c</sub>	0.298 <sub>c</sub>

Note. *DrInC*=Drinker Inventory of Consequences, *ASI*=Addiction Severity Index, *BDI*=Beck Depression Inventory, *PFI*=Psychosocial Functioning Inventory. Means sharing a subscript in a row indicate means that are not significantly different from each other based on Tukey post-hoc comparisons. Standardized scores were created in the full sample.