Summer 2023

Validation of the Addictions Neuroclinical Assessment Among Diverse Individuals with Alcohol Use Disorder

Elena R. Stein
University of New Mexico

Elena Stein, M.S.
Candidate

Psychology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Katie Witkiewitz, Ph.D., Chairperson

Barbara McCrady, Ph.D.

Eric Claus, Ph.D.

Yu-Yu Hsiao, Ph.D.

VALIDATION OF THE ADDICTIONS NEUROCLINICAL ASSESSMENT AMONG DIVERSE INDIVIDUALS WITH ALCOHOL USE DISORDER

by

ELENA R. STEIN

B.A., Neuroscience, Amherst College, 2012
M.S. Psychology, University of New Mexico, 2018

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
Psychology

The University of New Mexico
Albuquerque, New Mexico

August, 2023

ACKNOWLEDGMENTS

I am surrounded by a large village of people who gave me invaluable support throughout the experience of completing graduate school. None of this would have been possible, of course, without the unwavering support and guidance of my mentor, Dr. Katie Witkiewitz. I am grateful for the opportunities Katie made possible for me, and all of the ways in which she has impacted my growth into the psychologist I hope to become. I would also like to acknowledge the support of my dissertation committee, Drs. Barbara McCrady, Eric Claus, and Yu-Yu Hsiao, all of whom gave me hugely important feedback, as well as their time and effort. They have guided me through the dissertation process and in many other capacities: clinical work, master's thesis, and important coursework. Furthermore, I thank Drs. McCrady, Claus, and Witkiewitz in their role as multiple PIs for letting me use data from their two parent R01s. I greatly appreciate the many study staff at CASAA and MRN who worked so hard to collect study data. Likewise, I thank NIAAA for funding these two parent grants and my own F31 fellowship.

I am grateful to my friends, labmates, and colleagues— overlapping categories— for their constant encouragement, guidance, humor, and deep wells of compassion. I am especially thankful for my dear friend and labmate of five years, Tori Votaw, who made every single part of graduate school better. I am grateful to have learned from many other UNM faculty members who have supported me along the way. Drs. Terri Moyers and Elizabeth Yeater provided invaluable clinical supervision, lessons which will guide me for years to come. Dr. Kamilla Venner gave me impactful guidance conceptualizing this dissertation and many other topics. I am also deeply grateful for the clients and participants I have had the honor of working with in clinical and research capacities. I am thankful for their trust and willingness to approach challenges head-on. Finally, I am grateful for the love and support from my family— thank you.

VALIDATION OF THE ADDICTIONS NEUROCLINICAL ASSESSMENT AMONG DIVERSE INDIVIDUALS WITH ALCOHOL USE DISORDER

by

Elena R. Stein

B.A., Neuroscience, Amherst College
M.S. Psychology, University of New Mexico
Ph.D., Psychology, University of New Mexico

ABSTRACT

Individuals with alcohol use disorder (AUD) are heterogeneous in terms of etiology, maintenance, symptoms, and recovery, yet current diagnostic categories fail to adequately capture this heterogeneity. Corresponding to the neurobiological addiction cycle, the Alcohol and Addictions Research Domain Criteria proposes a framework of three core domains disrupted in AUD: negative emotionality, incentive salience, and executive function. The Addictions Neuroclinical Assessment (ANA) is a hypothesized multimodal assessment battery of these three domains, which may better characterize AUD heterogeneity. The current study validated the ANA in a sample of drinkers (N=245) who were diverse with respect to ethnicity and alcohol treatment-seeking status. This ANA model demonstrated measurement invariance over time, across sex, and across Hispanic and non-Hispanic white ethnicity. Drinking, incentive salience, and negative emotionality decreased over time, but only change in negative emotionality was associated with change in drinking. Clinical implications, measurement considerations, and future directions for precision medicine are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Alcohol use disorder (AUD) is highly prevalent in the United States, with approximately 29% of adults meeting criteria for AUD at some point in their lifetime (Grant et al., 2015). Understanding the mechanisms that drive AUD prevalence are critically important for accurately diagnosing and effectively treating AUD, particularly now as heavy drinking and alcohol-related deaths have increased substantially during the COVID-19 pandemic (Rossow et al., 2021; White et al., 2022). A diagnosis of AUD requires that an individual meets at least two of the eleven AUD criteria described in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (American Psychiatric Association, 2013). These eleven criteria represent aspects of impaired control over alcohol use, social consequences, risky use, and physiological criteria. Symptom counts of these eleven criteria are meant to represent a spectrum of severity of a unidimensional AUD phenomenon, although empirical literature has not supported this unidimensional model (Linden-Carmichael et al., 2019; Watts et al., 2021). Indeed it is not surprising that the DSM-5 AUD diagnosis fails to represent a unidimensional construct given that there are 2,048 possible combinations of AUD criteria resulting in diagnosis (Lane & Sher, 2015), a situation in which it is possible for two individuals who both meet diagnostic criteria for AUD to share no symptom overlap with one another.

Alcohol Use Disorder Heterogeneity & AARDoC

Among individuals who meet criteria for AUD, there is substantial heterogeneity in the etiology, maintenance, clinical course, and recovery (Litten et al., 2015; Tucker et al., 2020). This heterogeneity is not well-captured by the DSM-5 diagnostic framework, which is primarily based on consequences of alcohol use that may be multidetermined, not necessarily

reflective of etiologic variation, and non-specific to AUD with substantial overlap with other categories of psychopathology (Boness et al., 2021).

To address these concerns, researchers at the National Institute on Alcohol Abuse and Alcoholism proposed the Alcohol Addiction Research Domain Criteria (AARDoC) (Litten et al., 2015). This theoretical framework is modeled on the National Institute of Mental Health's Research Domain Criteria (RDoC), which integrates data across levels (e.g., genetic, behavioral, environmental) to elucidate core functional domains underlying psychopathology (Insel et al., 2010). The general RDoC model proposes six core functional domains: positive and negative valence systems, cognitive systems, arousal/regulatory systems, sensorimotor systems, and social processes (National Advisory Mental Health Council Workgroup on Changes to the Research Domain Criteria Matrix, 2018). Intended to provide further clarity around AUD heterogeneity, AARDoC was envisioned as an extension of RDoC specifically aimed at characterizing AUD phenotypes that could be useful in personalized medicine— that is, matching individuals to treatments meant to target their own personal characteristics or AUD subtype (Becker, 2015; Hutchison, 2008; Krystal & O'Malley, 2015; Litten et al., 2015).

Addictions Neuroclinical Assessment: Operationalization of AARDoC

The AARDoC has been operationalized by the Addictions Neuroclinical Assessment (ANA), a proposed assessment battery that examines three core functional domains in AUD: incentive salience, negative emotionality, and executive function (Kwako et al., 2016). These three constructs, fewer in number than the broader RDoC model for general psychopathology, were identified as uniquely important to AUD because of their correspondence to the Koob and Le Moal addiction cycle (Koob & Le Moal, 1997).

Koob & Le Moal addiction cycle. The addiction cycle posits three progressive and cyclical stages reflecting acquired neuroadaptations that occur with chronic compulsive substance use: binge/intoxication, withdrawal/negative affect, and preoccupation/anticipation. The binge/intoxication stage involves consuming a substance heavily and experiencing its rewarding, positive effects. This stage is primarily characterized by positive reinforcement learning and engagement of the mesocorticostriatal dopaminergic reward pathways in the brain (Koob & Volkow, 2016). Continued substance use in the binge/intoxication stage influences reward valuation, including conditioned cues related to the substance. As reward thresholds shift higher, there may be an increase in negative emotional states associated with withdrawal, when access to the rewarding substance is removed (Koob & Le Moal, 2001). Thus, the withdrawal/negative affect stage is characterized by both acute negative emotional states and chronic or protracted negative emotional states characterized by dysphoria, stress, and hypohedonia (Koob, 2021). Substance use in the withdrawal/negative affect stage may be driven by negative reinforcement processes and neuroadaptations in the nucleus accumbens, extended amygdala, and the hypothalamic-pituitary-adrenal axis (Koob & Volkow, 2016). The preoccupation/anticipation stage describes disruptions in executive functioning and subsequent behavioral self-regulation in the context of craving, mediated by prefrontal cortical circuitry. These three stages of the addiction cycle feed back into one another and intensify. The state of addiction is therefore thought of as progressive attempts to experience rewarding effects, relieve negative emotional states, and experience repeated failures of self-regulation, representing each stage of the cycle.

The ANA proposes a multimodal assessment battery to examine three domains that correspond to the Koob and Le Moal addiction cycle model: incentive salience, negative emotionality, and executive function (Kwako et al., 2016). Importantly, in these models, the three ANA domains are thought of as acquired states resulting from heavy substance use and subsequent neuroadaptations. Yet at the same time, individuals may vary with respect to baseline vulnerabilities or dispositional traits related to incentive salience, negative emotionality, and executive function, adding to heterogeneity. The ANA was proposed as a heuristic framework to advance our ability to characterize heterogeneity in AUD from a mechanistic perspective. In other words, not all individuals with heavy alcohol use will express impairments or even similar characteristics across all three domains, but these three domains represent core functional mechanisms through which addictive processes can develop in individuals.

Incentive salience is described as the psychological process of attributing motivational reward value to stimuli, mediated through dopaminergic mesocorticostriatal pathways (Kwako et al., 2016). Previously neutral stimuli can acquire reward value through incentive motivation and conditioned reinforcement processes (Berridge, 2012). Incentive salience may be best understood as a "wanting," rather than a "liking" phenomenon (Berridge & Robinson, 2016). With respect to substance use, evidence from animal models (Cofresí et al., 2019) as well as human laboratory and clinical studies (Zilverstand et al., 2018) shows substance-related cues and context take on high incentive salience value. In the ANA, incentive salience is expressed as stimuli or cues associated with alcohol use becoming increasingly attractive with continued alcohol use; phenomenologically, this may be perceived as increased craving for alcohol when exposed to salient cues. Incentive salience

and related constructs are related to AUD treatment outcomes. Greater cue-reactivity to alcohol-related stimuli is associated with higher probability of a return to alcohol use after a period of abstinence (Grüsser et al., 2004; Kvamme et al., 2019; Sinha et al., 2011), while a tendency to drink for positive reinforcement (i.e., reward drinking phenotype) has been shown to moderate AUD pharmacological treatment response such that those with higher reward drinking respond better to naltrexone than placebo (Mann et al., 2018; Witkiewitz et al., 2019).

Negative emotionality represents increased negative affective states, both acutely and chronically, with heavy alcohol consumption and withdrawal (Kwako et al., 2016). Negative emotional states range from anxiety, depression, hypohedonia, and alexithymia to and other dysphoric states. Individuals with AUD tend to report more negative moods generally (Grant et al., 2004). In addition, it has been widely documented that those with AUD demonstrate larger negative emotional responses to alcohol cues and stress cues than those without AUD (Sinha et al., 2009). Using alcohol for the purpose of reducing negative emotional states (i.e., negative reinforcement learning) is a key part of the addiction cycle (Koob & Le Moal, 1997), as well as other prominent addiction theories, such as the self-medication hypothesis (Luciano et al., 2022), and drinking to cope in the motivational model of substance use (Cooper et al., 2015). Negative emotionality has been shown to be predictive of AUD treatment outcomes in behavioral psychotherapy approaches (Swan et al., 2020). In secondary analyses of clinical trials examining the efficacy of medications for AUD, those who endorsed greater tendency to drink alcohol to relieve negative affect (i.e., relief drinking) had better drinking outcomes with acamprosate than placebo (Roos et al., 2017), although this finding was not replicated in a new sample (Mann et al., 2018).

Executive function broadly describes the cognitive and behavioral self-regulation processes involved in executing future-oriented goals (Bickel et al., 2012). There are numerous subdomains of executive function, yet those that fall under the umbrella of *cognitive control* seem to be most relevant in the context of heavy alcohol use and addictive disorders (Wilcox et al., 2014). These include attention, response inhibition, planning, working memory, behavioral flexibility, and valuation of future events (Kwako et al., 2016). The ANA executive function domain attempts to measure acquired dysfunction in these processes that are mediated by top-down frontal cortical circuits that regulate (or fail to regulate) incentive salience and negative emotionality by glutamatergic connections to the basal ganglia and the extended amygdala, respectively (Kwako et al., 2017). There is some evidence that cognitive control processes predict or mediate worse alcohol treatment outcomes (Wilcox et al., 2014).

Current ANA Literature

Broadly, the ANA literature to date is characterized by initial support of the three factor AARDoC model by several independent research groups, but also notable methodological weaknesses. The originally proposed ANA battery by Kwako and colleagues (2016) consisted of self-report, behavioral, and task-based neuroimaging assessments for each of the three domains. This comprehensive battery is estimated to take 10 hours to complete (Ghitza, 2017). In order for the ANA to be more widely implemented in research and clinical settings, a more concise or perhaps modifiable battery must be established.

Since the original ANA battery was proposed, several research groups have attempted to validate the ANA model using secondary data analysis of pre-existing datasets. This approach has benefits and limitations. Secondary data analysis enables researchers to

evaluate the factor structure of the ANA much more quickly and efficiently, and some existing studies already incorporate commonly used assessments that correspond to the incentive salience, negative emotionality, and executive function constructs. In addition, it may be advantageous to demonstrate that the three factor structure can be replicated using sets of different assessments because that implies the reliability of the underlying latent constructs. On the other hand, an accumulation of secondary data analytic evidence for the three factor ANA model— each using different assessment batteries that were not specifically designed to measure ANA constructs— leaves us no closer to knowing which assessments have optimal discriminant validity or utility for precision medicine. Although there is a need for primary research studies specifically designed to answer questions about the ANA, much value can still be gleaned from careful secondary data analytic efforts.

The first published study validating the ANA was by Kwako and colleagues with a sample that included individuals across the alcohol use spectrum, from healthy controls who did not meet criteria for AUD, to those seeking treatment for AUD (Kwako et al., 2019). This sample was notably diverse with respect to racial identity (greater than half of the sample was African American) and alcohol use severity, but not with respect to ethnic identity (less than 6% of the total sample was Hispanic/Latino/a/x). These authors used exploratory and confirmatory methods to validate the three factor ANA model, multiple indicator - multiple cause (MIMIC) analysis to examine predictors of the ANA factors, and receiver operating characteristic curve analyses to assess the ability of each factor to predict AUD. Although the authors did replicate the three factor model with factors that distinguished well between individuals with and without AUD, the measures used as factor indicators in the model were not ideal matches to the theoretical aims of the ANA. For example, the factor described as

negative emotionality consisted of many trait-like indicators with significant loadings, including the NEO-Personality Inventory neuroticism scale (positive) and extraversion and agreeableness scales (negative), as well as trait anxiety, aggression, and positive urgency; but depression did not significantly load on this negative emotionality factor. Personality traits, aggression, and positive urgency do not align closely with negative emotionality as conceptualized as an acquired affective state following chronic, negatively reinforced alcohol use. Although personality traits may show a pattern of normative change over the life course (Roberts et al., 2006), they are nonetheless thought to be relatively stable constructs and thus personality assessments are likely not the most sensitive tool to capture alcohol-related affective change. Additionally, all indicators of the executive function and incentive salience factors are self-report, which may not be the optimal approach to capture information related to these domains. Self-report measures of executive function have not been shown to correlate to performance-based measures of executive function (Buchanan, 2016).

Following this initial validation study, DeMartini et al. sought to replicate this model with similar measures and analytic methods. These authors conducted a secondary data analysis of a sample of non-treatment seeking drinkers who were primarily white (56%) and evenly split between male and female (51%) participants. The participants in this study reported relatively lighter or less problematic drinking, with only about a quarter of the sample meeting criteria for a current or lifetime AUD diagnosis. Similar to Kwako et al. (2019), this study likewise found evidence for a three factor model corresponding to ANA domains using self-report measures of impulsivity (Barratt Impulsiveness Scale and UPPS subscales) for the executive function factor; self-report measures of habitual drinking behavior, enhancement and coping motives, and a single item for craving for the incentive

salience factor; and self-report measures of depression, anxiety, and obsessive-compulsive symptoms for negative emotionality. These self-report measures have similar problems as described above, namely that self-report measures of incentive salience and executive function may not be the most valid or sensitive approach to capturing variance in these factors. Furthermore, coping motives (e.g., drinking to relieve negative affective states, a negative reinforcement process) appeared to be the strongest indicator of what was described as the incentive salience factor in exploratory factor analysis findings (confirmatory factor analysis item loadings were not reported). However, this may not be a good theoretical fit given that incentive salience and the binge intoxication phase of the addiction cycle are described as being mediated by positive reinforcement processes. Despite these methodological concerns, impairment in these ANA factors were all related to current AUD diagnosis and were associated with predicators such as family history of AUD, age of alcohol use onset, and trauma history, similar to findings from Kwako et al., (2019).

Additional efforts to validate individual ANA domains include several secondary data analysis studies of the Relapse Replication and Extension Project (RREP) (Lowman et al., 1996). The original RREP study recruited individuals seeking alcohol treatment in community treatment programs. The sample was primarily white (67%) and male (59%). A study by Votaw et al. focused on the negative emotionality domain and demonstrated good fit of a unidimensional confirmatory factor analysis model that utilized a brief battery of measures assessing negative affective consequences of drinking and symptoms of depression and anxiety (Votaw et al., 2020). The highest loading factors were the Beck Depression Inventory and Beck Anxiety Inventory, with relatively lower factor loadings from the State Trait Anger Expression Inventory and individual affective consequence items from the

Drinker Inventory of Consequences. Notably, the Beck Depression Inventory and Beck Anxiety Inventory are state measures that were initially proposed by Kwako et al. (2016). Votaw et al. showed that their negative emotionality factor was measurement invariant by gender and over time (baseline and up to 12-months during or after treatment). This negative emotionality domain was positively associated with more frequent and heavier drinking as well as drinking to regulate negative affect. Additional evidence suggested the association between baseline negative emotionality and drinking outcomes was mediated by higher coping motives. Thus, negative emotionality may have important predictive validity.

In another study by the same research group using secondary analyses of the RREP data, Stein et al. validated the incentive salience domain using confirmatory factor analytic methods with a selection of self-report items from the Alcohol Dependence Scale and Impaired Control Scale (Stein et al., 2021). The items used to assess a latent incentive salience factor were about intensity of urges or thoughts about drinking, as well as perceived difficulty resisting or limiting drinking. The incentive salience factor was shown to be measurement invariant by sex and positively associated with drinking frequency and intensity at baseline and 12-month follow-up. These finding suggest that incentive salience may be modeled using self-report measures and shows relevant predictive validity. However, questions remain about the relative advantages of assessing incentive salience with self-report versus behavioral assessments of related constructs, such as cue reactivity and incentive motivation, given that the literature has consistently shown that behavioral tasks can capture incentive salience below the level of conscious awareness (Wiers et al., 2002).

A fourth research group has published two secondary data analysis studies validating the ANA model among heavy drinkers (Nieto et al., 2021) and individuals who use

methamphetamine (Nieto & Ray, 2022). The sample used to validate the ANA among drinkers was primarily male (67%) and characterized as Black/other racial identity (75%). Exploratory factor analytic methods showed evidence for a four factor solution: three factors that corresponded to the ANA domains, and a fourth factor that was described as an alcohol-related consequences domain. These findings suggest that the ANA domains capture AUD heterogeneity above and beyond AUD phenomenology (e.g., loss of control over drinking, alcohol withdrawal, dependence criteria). Items used for the incentive salience factor were from the Alcohol Dependence Scale (obsessive subscale), Obsessive-Compulsive Drinking Scale, and Penn Alcohol Craving Scale with significant cross loadings from alcohol use severity measures. The negative emotionality factor was determined by state measures (Beck Depression Inventory, Beck Anxiety Inventory), as well the State-Trait Anxiety Inventory. The executive function factor had significant factor loadings from measures of attention and memory (Digit Span) and delay discounting (Monetary Choice Questionnaire). Of note, self-report measures of trait impulsivity did not load significantly onto the executive function factor. In this sample, greater dysfunction on ANA domains was associated with drinking severity and older age. Male sex was associated with higher scores on the incentive salience and alcohol-related consequences domains.

Although largely congruent, the second study by Nieto and colleagues validating the ANA among persons who used methamphetamine suggested some slight differences in ANA factor measurement compared to their first study of heavy drinkers (Nieto & Ray, 2022). The participants in this study were majority male (72%) and evenly split between non-Hispanic white and Latino (approximately 40% each). The negative emotionality factor similarly included the Beck Depression Inventory and Beck Anxiety Inventory, but also included

negative affective items from the Methamphetamine Withdrawal Questionnaire and Barratt Impulsiveness Scale subscales (constraint, impulsivity). A selection of items from the Methamphetamine Urge Questionnaire that were conceptually similar to items used in prior studies (e.g., related to craving, impaired control over use) comprised the incentive salience factor. The executive function factor had significant loadings from Digit Span, but unlike the previous study, delay discounting failed to load significantly on this factor. A measure of risk attitudes also loaded on the executive function factor, but a stop signal task did not. Stimulant use disorder symptoms counts were significantly positively associated with both the incentive salience and negative emotionality factors. For both studies by Nieto and colleagues, the samples consisted of a minority of white participants (25-40%), but the other participants' racial and ethnic identities were not described in detail. It is unclear whether the ANA factors the authors identified in these studies were measurement invariant across these diverse identity groups.

Taken together, there are several methodological and analytic issues to consider with respect to the existing ANA studies. In several of the studies described above, the statistical approaches had several limitations. Confirmatory factor analysis approaches resulted in three ANA factors with high inter-factor correlations (e.g. ranging from .76 to .90 in Kwako et al., 2019 and ranging from .45 to .71 in DeMartini et al., 2021), whereas the Nieto work (Nieto et al., 2021; Nieto & Ray, 2022) used exploratory factor analysis that allowed all items to load on all factors and Nieto et al (2021) forced zero correlations between factors via orthogonal rotation. Confirmatory factor analysis factor loadings were not reported by DeMartini et al. (2021). The confirmatory factor analysis approaches may be problematic because highly correlated ANA factors likely have inadequate discriminant validity if they share

considerable substantive overlap, which detracts from the original intention of the ANA approach to differentiate AUD phenotypes.

In addition, the measures used as ANA factor indicators in prior studies as described above demonstrate inconsistent factor loadings between studies. For example, a depression measure loaded significantly on the negative emotionality factor in one study (DeMartini et al., 2021), while it loaded on the incentive salience factor in another (Kwako et al., 2019). Likewise, the Barratt Impulsiveness Scale has shown inconsistent factor loadings, with significant loadings on the executive function factor in some studies (DeMartini et al., 2021; Kwako et al., 2019), but not loading on any factor in others (Nieto et al., 2021). Potential explanations for these inconsistencies may be that ANA factors are capturing very broad constructs that do not exactly correspond to the specific acquired impairments of the addiction cycle. Furthermore, it is theoretically possible that certain items contribute variance to more than one ANA factor and that there is some association between ANA factors, but the statistical methods used thus far (confirmatory factor analysis and exploratory factor analysis with orthogonal rotation, respectively) cannot account for this.

Like any theoretical model of psychological constructs, it is critically important to examine whether the ANA constructs and measurements used to assess them are valid in diverse populations. Previous work on the ANA has been mixed when it comes to the racial and ethnic diversity of research participants, as well as other facets of diversity. Several studies have examined race as a predictor ANA domains with the understanding that non-white race is not associated with any inherent characteristics, but rather it is likely a proxy for experiences of racism, a factor that impacts health and well-being (DeMartini et al., 2021; Kwako et al., 2019; Nieto et al., 2021; Yearby, 2018). In these studies, there have been

mixed findings about whether white racial identity is associated with better (Nieto et al., 2021) or worse (DeMartini et al., 2021; Kwako et al., 2019) executive function, and some evidence that being white is associated with lower incentive salience scores (Kwako et al., 2019). However, examining the binary comparison of white and non-white (alternatively described as Black/Other) racial identity oversimplifies the complexity of racial identity and obscures the heterogeneity within the large swath of the population that is not white. Furthermore, given that the Hispanic/Latino/a/x population is the second most populous ethnic group in the United States and growing (U.S. Census Bureau, 2017), it will be especially important to validate the ANA within this group specifically. To be clear, the ANA proposes neurobiologically derived domains, and given that race is a social construct, not a biological one, there is no evidence to suggest underlying neurobiological differences among racial/ethnic groups. However, structural inequities and sociocultural factors influence the prevalence and presentation of addictive disorders and may impact task performance on measures recommended for AARDoC or RDoC assessments (Dodell-Feder et al., 2020). Thus, it is critical to establish construct validity of the ANA across diverse racial and ethnic groups and, to date, no ANA studies have examined measurement invariance across racial or ethnic groups.

It is also important to examine measurement invariance by sex, which has been done only by two prior ANA studies that found measurement invariance of the incentive salience and negative emotionality domains in the same sample of treatment seeking individuals with AUD (Stein et al., 2021; Votaw et al., 2020). These studies did not examine measurement invariance of executive function by sex. Unlike ethnicity, there are some well-documented neurobiological sex differences in addiction (Becker & Koob, 2016) that may be relevant to

the measurement of ANA domains, including in the brain's reward network regions (Sawyer et al., 2017) and amygdala pathways involved in stress-motivated alcohol use (Mineur et al., 2022). Particularly in cases in which there may be sex differences in underlying mechanisms, it is crucial to establish that ANA measurement is equivalent by sex so that any group differences can be properly interpreted as true group differences, and not artifacts of measurement non-equivalence between males and females. Of note, in this study we refer to the biological construct of sex assigned at birth rather than gender, which is better understood as a social construct and does not necessarily align with sex assigned at birth.

Considering the existing literature on the ANA to date, there are several key gaps and remaining questions that the current study aimed to address. First, as noted above, prior studies have shown high inter-factor correlations, and there is a need to employ multimodal assessments and analytic techniques that increase discriminant validity of the factors. Second, there is a need to understand more about the longitudinal characteristics of the ANA factors and how they relate to changes in drinking behavior over time. Finally, there is a need for additional evidence about whether this framework is valid among diverse demographic groups.

Current Study

This secondary data analysis study used methods balancing exploratory and confirmatory approaches to test a three-factor ANA model using a sample of heavy drinkers recruited from the community. To assess negative emotionality, incentive salience, and executive function, we proposed a battery that included self-report, behavioral, and functional neuroimaging measures. We examined the validity of a three-factor ANA model in several ways: testing whether the assessment battery was measurement invariant over

time, by sex, and by ethnicity, examining the relationships between change in drinking and change in ANA factors over time, and investigating direct and moderated associations between sex, ethnicity, ANA factors, and drinking. Based on prior research, we hypothesized that we would find strong support for a three factor ANA model at all time points. Next, we hypothesized that at baseline, greater drinking intensity would be associated with higher negative emotionality (Votaw et al., 2021), higher incentive salience (Nieto et al., 2021; Stein et al., 2021), and more disrupted executive function (Kwako et al., 2019). Given that there have been inconsistent findings in the ANA literature, we did not make directional hypothesis about whether sex or ethnicity would be associated with the three ANA factors. In terms of longitudinal relationships, we hypothesized that reductions in drinking over time would be associated with decreases in negative emotionality, decreases in incentive salience, and improvements in executive function.

Materials and Methods

Data and Participants

The current study was a secondary data analysis combining data from two primary longitudinal studies, *ABQ DrinQ* and *ABQ Treat*, designed to examine neurobiological, behavioral, and cognitive characteristics associated with alcohol use and alcohol behavior change over an 18-month study period (ABQ DrinQ: R01AA023665; ABQ Treat: R01AA025762). Both primary studies recruited heavy drinking adult participants from the Albuquerque, New Mexico metropolitan area. ABQ DrinQ recruited nontreatment-seeking hazardous drinkers, while ABQ Treat recruited treatment-seeking individuals with AUD.

Participants (N=245) from ABQ DrinQ (n=189) and ABQ Treat (n=56) were recruited from the community using flyers, advertisements on radio, television, and online,

and at community events such as beer and wine festivals. To be eligible for study enrollment, participants had to be ages 22-85 years, self-identify as a "heavy/binge/weekly drinker" (both studies) or "moderate drinker" (ABQ DrinQ only), demonstrate "hazardous and harmful alcohol use" as indicated by an Alcohol Use Disorder Identification Test (AUDIT; Saunders et al., 1993) score > 8 for men and > 7 for women, engage in drinking within the past 30 days, produce an alcohol level of 0.00 by saliva test strip at in-person screening in order to give consent to study procedures, and be right handed. An additional inclusion criterion for ABQ Treat was that participants must be explicitly seeking treatment to reduce their drinking at the time of enrollment and be willing to participate in 12 weekly psychotherapy sessions. The following exclusion criteria were applied in both studies in order to examine individual alcohol use trajectories and to account for magnetic resonance imaging (MRI) contraindications: history of brain injury or neurological diagnoses, evidence of or lifetime schizophrenia, or bipolar disorder (ABQ DrinQ) or current psychosis (ABQ Treat), substance use disorder other than nicotine or cannabis, evidence of recent illicit drug use (other than cannabis) on a baseline urine toxicology screen, contraindications for MRI (e.g., metal in the body), pregnancy as verified by a urine pregnancy screen prior to MRI, estimated IQ < 80, inability to read or speak English fluently, history of severe alcohol withdrawal (e.g. seizures, tremors, delirium tremens). An additional exclusion criterion for ABQ DrinQ was that participants could not be currently (or within the past one month) receiving treatment or participating in any form of mutual help to reduce drinking. An additional exclusion for ABQTreat was that participants could not be concurrently enrolled in other treatment for alcohol use at the time of study enrollment.

Procedures

These two studies utilized nearly identical assessment procedures and measures. Following a telephone screen to determine initial eligibility, participants were invited to complete a baseline assessment session, at which time informed consent was obtained from all participants. The baseline assessment included self-report, behavioral, cognitive, and functional MRI measures, detailed below, and was often divided across several days. Participants enrolled in ABQ Treat were randomized at the end of the baseline assessment to receive either Cognitive Behavioral Treatment or Mindfulness Based Treatment for 12 weekly psychotherapy sessions. Participants in both ABQ DrinQ and ABQ Treat were then invited to complete a 3-month assessment session comprised of the same self-report, behavioral, cognitive, and functional MRI measures as baseline. Both primary studies involved several assessment sessions through 18-months post-baseline. However, this current study only includes data from the baseline and 3-month assessment sessions. All study procedures were approved by the University of New Mexico Institutional Review Board.

Self-Report and Behavioral Assessments

Demographic Questionnaire. A self-report questionnaire assessing various demographic factors was given to participants at baseline. Demographic variables assessed included biological sex, age, ethnicity, racial identity, and others.

Timeline Followback (TLFB). We used the TLFB, a calendar and event-based cue method, to collect daily drinking data. At the baseline assessment session, participants reported their daily drinking for the previous 90 days. At the 3-month assessment session, participants reported their daily drinking since the previous assessment date. We calculated two summary measures of drinking patterns based on these daily data: percent heavy drinking days (PHDD; defined as greater than 4/5 daily drinks for women men (National

Institute on Alcohol Abuse and Alcoholism, 2022) and drinks per drinking day (DDD) to assess drinking frequency and intensity, respectively. For the latent change score analyses, we used PHDD and DDD at each time point to model a latent drinking variable to model change in drinking frequency and intensity over time. A large body of research has demonstrated the validity and reliability of the TLFB measure (Sobell et al., 2001; Sobell & Sobell, 1996).

Structured Clinical Interview for DSM-5 (SCID). Diagnoses of Alcohol Use Disorder were established at the baseline assessment session using the SCID Module E, Substance Use Disorders (First et al., 2016).

Alcohol Use Disorder Identification Test (AUDIT). The AUDIT is a 10 item questionnaire assessing hazardous alcohol consumption. The AUDIT was administered during the initial phone screen to determine eligibility (Saunders et al., 1993).

Indicators of Addictions Neuroclinical Assessment domains

NIH Toolbox Emotion Battery. The NIH Toolbox Emotion Battery for Ages 18+ was administered at baseline and 3-month assessment sessions using an iPad application. This battery consists of self-report measures assessing the following constructs: Positive Affect, General Life Satisfaction, Emotional Support, Friendship, Loneliness, Perceived Rejection, Perceived Hostility, Self-Efficacy, Sadness, Perceived Stress, Fear, and Anger (Salsman et al., 2013). Each of these measures included 5 to 12 self-report items delivered either as a Fixed Form, or as a Computer Adaptive Test. Based on consistency with ANA negative emotionality domain measurement from previously published literature, we included age-corrected standard scores for the measures of Positive Affect, Loneliness, Self-Efficacy, Sadness, Perceived Stress, Fear, and Anger as indicators for this domain.

NIH Toolbox Cognition Battery. The NIH Toolbox Cognition Battery for Ages 18+ was administered at baseline and 3-month assessment sessions using an iPad application (Weintraub et al., 2013). This battery consists of brief tests a variety of cognitive constructs, each taking between approximately 3-7 minutes to complete. For this study, we included age-corrected standard scores for three measures: Flanker Inhibitory Control and Attention Test (construct: attention & executive functioning), List Sorting Working Memory Test (construct: working memory), and Dimensional Change Card Sort Test (construct: executive function) as indicators for the ANA executive function domain.

Monetary Choice Questionnaire (MCQ). The MCQ is a 27-item self-report measure of delay discounting that asks participants whether they would prefer to have a smaller amount of money immediately or a larger amount of money after a delayed period of time (e.g., "Would you rather have $11 today or $30 in 7 days?") (Kirby et al., 1999). The extent to which an individual discounts delayed rewards (*k*) is estimated, where larger *k* values represent steeper rates of discounting (Kaplan et al., 2016). The natural log transform of *k* (ln*k*) was included in the current study as an indicator of the ANA executive function domain.

Impaired Control Scale (ICS). The ICS is a self-report questionnaire consisting of 44 items designed to measure impairment in control over alcohol use in the past six months (Heather et al., 1993). Three subscales are generated: attempted control, failed control, and perceived control. For this study, the failed control (FC) subscale, which measures participants' frequency of impaired restraint over drinking in the past six months, was included as an indicator of the ANA incentive salience domain. The ICS has been shown to have good discriminant and predictive validity in prior studies (Heather et al., 1998).

Penn Alcohol Craving Scale (PACS). The PACS is a 5-item self-report questionnaire assessing frequency, intensity, and duration of craving over the past week. This measure has shown good concurrent and predictive validity in prior studies (Flannery et al., 1999). In this current sample, the PACS showed good internal consistency (McDonald's Omega (Ω) reliability = .89, .91 at baseline, three months) and was included as an indicator of the incentive salience domain.

Short Inventory of Problems (SIP). The SIP is a 15-item self-report measure that assesses various consequences of alcohol use (Kiluk et al., 2013). In this current study, two items were used as indicators of the ANA executive function domain (item 5: "I have taken foolish risks when I have been drinking;" item 6: "When drinking, I have done impulsive things that I regretted later"). These items were selected because they capture impulsive and risk-taking behavior suggesting lack of behavioral control in the context of drinking. Prior ANA studies described above have used self-report assessments of general impulsivity such as the Barratt Impulsiveness Scale and the UPPSP. Thus, these two SIP items assess a similar impulsivity construct, but in a more specific drinking context.

Functional Magnetic Resonance Imaging Assessments

Visual Cue Task. This task involved two 12.5 minute runs in which participants were shown 150 visual cues total, comprised of 50 images each category: negative, neutral, and alcohol. Each image was presented for 2300 ms with a variable duration fixation cross shown between images (1380, 1840, or 2300 ms). Images were presented in a pseudorandom order such that sequences of six images (two from each category) were shown in mini-blocks, with longer fixation periods (4600 - 8280 ms) between mini-blocks.

For the negative emotionality factor, we examined the activation differences between negative and neutral image presentation by calculating percent signal change in the Blood Oxygenated Level Dependent (BOLD) responses in seed regions of interest in the bilateral amygdala, given its well-document role in emotion processing and the negative emotionality AARDoC domain (Voon et al., 2020; Yang et al., 2020).

For the incentive salience factor, we examined the BOLD percent signal change between alcohol and neutral image presentation in seed regions of interest in the left anterior cingulate cortex, left posterior cingulate cortex, right caudate, and right medial orbitofrontal gyrus. These regions were selected based on prior work demonstrating cue reactivity task activation in these areas (Kim et al., 2014; Schacht et al., 2013) and correspondence to the AARDoC incentive salience domain (Voon et al., 2020).

Stop Signal Task. This task was 5.7 minutes of Go and Stop trials. In each trial, an empty circle was presented on screen for 500 ms, and then a circle with either a left- or right-pointing arrow was presented for up to 1000 ms. Participants were instructed to respond by indicating which direction the arrow was pointing by either pressing a button with their pointer finger (to indicate left) or their middle finger (to indicate right). These constituted the Go trials. For 25% of the trials, the area within the circle would turn red after a variable delay, indicating to the participant that they were to withhold their responding and as such, these were Stop trials. Participants were told to respond as quickly and accurately as they could.

For the executive function factor, we examined the BOLD percent signal change between correct inhibition (i.e., correct withholding of button-press when the red circle appeared) vs. correct Go trials (i.e., correct button-pressing indicating right and left arrows).

We interpret this contrast as representing accurate response inhibition, a construct subsumed under the AARDoC executive function domain. We selected seed regions of interest including the right inferior frontal gyrus opercular part, right insula, and right median cingulate and paracingulate gyri based on their theoretical correspondence with the AARDoC executive function domain and their specific correspondence to response inhibition, as demonstrated by prior work in this area (Zhang et al., 2017).

MRI acquisition and data analysis. Scans were acquired on a Siemens 3T Trio TIM scanner with a 32-channel head coil. Additional details about MRI data acquisition, quality control, and preprocessing have been published elsewhere (For information about the visual cue task, see: Al-Khalil et al., 2021; for information about the stop signal task, see: Swartz et al., 2021).

Statistical Analyses

All analyses were conducted in Mplus, version 8.7 (Muthén & Muthén, 2017) using maximum likelihood estimation with robust standard errors, a robust method for handling missing data under the assumption that data are missing at random (Hallgren & Witkiewitz, 2013).

Exploratory Structural Equation Modeling (ESEM). To test the a priori latent factor structure of the ANA using a novel set of multimodal indicators, we used ESEM, a method that allows for flexible modeling of a pre-specified, theoretically-derived factor structure in which indicators are free to load on more than one factor. This accommodates items that may contribute variance to more than one factor, which may be theoretically sound when items are multidetermined. Allowing cross-loadings is a notable feature of this analytic method because it reduces the artificial inflation of factor correlations that typically occur when

cross-loadings are set to zero in confirmatory factor analysis, thus improving discriminant validity of the factors when using ESEM. In contrast to exploratory factor analysis, we pre-specified the number of factors to be three. Model fit was evaluated at both baseline and three-month time points with χ^2 test, absolute fit indices (Root mean square error of approximation (RMSEA) < .06), and incremental fit indices (Comparative fit index (CFI) > .95) (Hu & Bentler, 1999). To assess concurrent validity of the ANA domains, we included drinking covariates (PHDD, DDD) in the baseline ESEM model. Additionally, we examined age and study (ABQ DrinQ, ABQ Treat) as covariates since prior research has shown associations between age and executive functioning constructs (Swartz et al., 2021) and clinical indicators of alcohol use severity and negative emotionality (Votaw et al., 2020).

Measurement Invariance Testing. In order to compare latent ANA factors over time (baseline, 3-month follow-up) and to establish the validity of the ANA factor structure in diverse demographic groups (male and female sex; Hispanic and non-Hispanic white ethnicity[1]), a series of measurement invariance models were tested in order of least to most restrictive. First, we tested configural invariance models to examine the overall factor structure at different time points and among demographic groups. These models provided a test of whether the factor structure itself could be considered the same across groups (e.g., same number of factors and items). Next, we tested metric invariance models, also known as "weak" invariance, by constraining factor loadings to equality. For a model to be metric invariant, this suggests that at the same level of the latent factor, the associations between the latent factor and the indicator items are equivalent across groups. We then examined scalar invariance, also known as "strong" invariance, by constraining factor loadings as well as

[1] Non-Hispanic white was selected as the comparison group, as opposed to non-Hispanic regardless of race, because this reduced heterogeneity in the comparison group.

intercepts to equality. Scalar measurement invariance is critical for comparing factor means between groups. Achieving scalar invariance means that at the same level of the latent factor, the association between the latent factor and the items as well as item-level endorsement is equivalent across groups. Residual measurement invariance, also known as "strict" invariance, is often viewed as a highly restrictive form of invariance that constrains loadings, intercepts, and residual invariances to be equal. Achieving scalar invariance means that at the same level of the latent factor, the associations between the latent factor and the items, item-level endorsement, and item-level residual variances are equivalent across groups. For all invariance tests (metric versus configural; scalar versus metric; strict versus scalar), we used well-established model comparison tools in Mplus to examine decrement in model fit based on a priori thresholds (e.g., a significant chi-square difference test, a negative change in CFI ≥.01, Cheung & Rensvold, 2002, and a positive change in RMSEA ≥.015, Chen, 2007). For all sets of measurement invariance tests (over time, sex, and ethnicity), we tested whether metric invariance models did not fit worse than configural models, whether scalar invariance models did not fit worse than metric models, and strict invariance models did not fit worse than scalar models. If there was adequate evidence that strict models did not fit worse than scalar models, or that scalar models did not fit worse than metric models, we then moved on to subsequent analyses that compared means across these groups.

Latent Change Score Analysis. In order to examine the relationships between change in drinking and change in ANA factors from baseline to three months, we used a series of Latent Change Score (LCS) models. The two-wave LCS (2W-LCS) approach is an optimal statistical method when researchers want to learn about both between- and within-person change processes, and model the measurement error by specifying multiple-indicator latent

variables. More specifically, we estimated the latent variable for each ANA construct of interest at baseline and three-month follow-up and a latent change factor for each construct. We also estimated a latent drinking variable at both time points and a latent change factor for the drinking latent construct. The rationale for estimating a latent drinking variable was to capture information related to drinking frequency as well as intensity; the latent drinking factor was comprised of PHDD and DDD parameters calculated from the Timeline Followback at each time point.

Using LCS, we estimated relationships among the latent change factors for multiple constructs. The first assumption in an LCS approach is that constructs must be measurement invariant at the two time points. Thus, in addition to testing measurement invariance of ANA factors over time as described above, we also tested measurement invariance of the univariate latent drinking factor over time in the same manner. Prior to fitting multivariate LCS models, we fit univariate LCS models to examine the change process for each construct (e.g., ΔNegativeEmotionality, ΔIncentiveSalience, ΔExecutiveFunction) and whether the mean or variance of its latent change factor was significantly different from zero, which provides information about degree of heterogeneity in within-person change. A univariate LCS model for the drinking factor could not be identified given that it involves a two-indicator latent factor. Thus, we fit three univariate models (negative emotionality, incentive salience, executive function) and three bivariate LCS models (negative emotionality and drinking; incentive salience and drinking; executive function and drinking; e.g., $\beta_{\Delta Drinking, \Delta NegativeEmotionality}$, $\beta_{\Delta Drinking, \Delta IncentiveSalience}$, $\beta_{\Delta Drinking, \Delta ExecutiveFunction}$). For each LCS model, fit was evaluated with the χ^2 test, absolute fit indices (Root mean square error of approximation (RMSEA) < .06), and incremental fit indices (Comparative fit index (CFI) > .95) (Hu &

Bentler, 1999). Finally, we examined age and study as covariates in each of the three LCS models.

Results

Sample Characteristics

The sample included a total of 245 participants, 56 of whom were treatment-seeking and enrolled in the ABQ Treat study, and 189 of whom were not seeking-treatment when they enrolled in the ABQ DrinQ study. Recruitment for the ABQ Treat study was ongoing at the time of analyses for this dissertation, thus the relatively smaller sample size. In the total sample, 46.5% of participants were female; 48.6% were Hispanic, 36.3% non-Hispanic white, 19.7% American Indian/Alaska Native, 3.7% African American, 3.7% Asian, and 0.4% Native Hawaiian. Of note, participants could indicate multiple ethnic or racial identity categories. They had an average age of 35.8 years (SD = 10.5, range 22 to 69). In terms of other sociodemographic characteristics, 44.1% of participants were never married, 24.5% married, 13.9% divorced, 9.8% living with a non-married partner, 3.6% engaged, 2.8% separated, and 1.2% widowed. Participants' annual income was such that 11.8% earned from $0-$9,999, 15.9% from $10,000-$19,999, 17.2% from $20,000-$29,999, 13.8% from $30,000-$39,999, 7.3% from $40,000-$49,999, 9.4% from $50,000-$59,999, and 24.1% over $60,000. Participants in ABQ Treat were significantly older than those in ABQ DrinQ, but did not vary on any other demographic variable. Sociodemographic characteristics of the full sample as well as descriptive statistics by study are presented in Table 1.

Baseline drinking characteristics were computed using past 90 day daily drinking behavior obtained using the Timeline Followback (TLFB). Across the full sample, the mean drinks per drinking day (DDD) at baseline was 6.2 and the percent of heavy drinking days

(PHDD; defined by >4/5 drinks per day for women/men) at baseline was 32.3%. On average, PHDD was significantly higher for those in ABQ Treat (45.8%) than those in ABQ DrinQ (28.7%, $p < .001$), although DDD was not significantly different between studies. The sample size at three months was 201, accounting for 44 participants who did not complete the three-month follow-up assessment. A higher proportion of participants enrolled in the ABQ Treat study completed the three-month follow-up assessment (94.6%, n=53) compared to those enrolled in the ABQ DrinQ study, 78.3% of whom completed the three-month follow-up (n=148). Individuals who completed the three-month assessment were more likely to be older (M = 36.7) than those who did not complete the three-month assessment (M = 31.9 ; p = .003), but not did not differ on other demographic variables or baseline drinking variables. Baseline clinical characteristics are presented in Table 1 for the total sample and by parent study. Supplemental Table 1 shows clinical characteristics for the total sample and by parent study over time.

Exploratory Structural Equation Models

Baseline measurement model with neuroimaging, self-report, and behavioral indicators. We fit an exploratory structural equation model (ESEM) to the full set of self-report, behavioral, and neuroimaging hypothesized indicators at baseline. This model fit the data poorly (Model $\chi 2$ (295) = 633.529, p < 0.001; RMSEA = 0.068; CFI = 0.850). For each hypothesized ANA factor, the hypothesized neuroimaging indicators did not load onto the same factor as the self-report and behavioral items. Factor loadings for this first iteration of an ANA baseline model are presented in Table 2.

Baseline measurement model with neuroimaging indicators only. Given that the full ESEM model with all hypothesized indicators did not fit the data well and there were low

correlations between the neuroimaging indicators and self-report/behavioral indicators within each hypothesized factor, we attempted to re-estimate the three-factor model in two ways: (1) with neuroimaging indicators only, and (2) with self-report and behavioral indicators only. For the neuroimaging only model, we found mixed evidence for model fit, (Model $\chi 2$ (42) = 74.449, p = 0.002; RMSEA =.069; CFI =.925). Although the model fit indices were slightly worse than our a priori thresholds, they represented an improvement from the first model and were quite close to the model fit thresholds. All indicators significantly loaded onto their hypothesized ANA factor, and one executive function indicator (right insula, stop signal task response inhibition) had a low but significant cross-loading on the incentive salience factor. Factor loadings for this second iteration of an ANA baseline model are presented in Table 3.

Baseline measurement model with self-report and behavioral indicators only. We also re-estimated the baseline ESEM model with self-report and behavioral indicators only and found excellent fit to the data (Model $\chi 2$ (61) = 73.311, p = 0.134; RMSEA = 0.029; CFI = 0.993). Because this model clearly fit the data best, we accepted this as the final ANA ESEM model with which to move forward to subsequent analyses. We identified three factors with loadings that were consistent with the ANA framework. The first factor could be described as a negative emotionality factor because all NIH Toolbox Emotion measures of negative affectivity (i.e., sadness, anger affect, fear affect, loneliness, anger hostility, and perceived stress) loaded positively onto this factor, while the NIH Toolbox measure of positive affect loaded negatively. The second factor could be described as an incentive salience factor, as the ICS Failed Control subscale and the PACS summary score loaded positively onto this factor. In addition, the two individual SIP items 5 (“I have taken foolish risks when I have been drinking”) and 6 (“When drinking, I have done impulsive things that I

regretted later") loaded positively onto this incentive salience factor, suggesting shared variance between items assessing craving and impaired behavioral control in relation to alcohol. The third factor can be described as an executive function factor with positive loadings by the delay discounting measure, and negative loadings of the NIH Toolbox cognition measures (i.e., flanker, list sorting working memory, and dimensional change card sort). Thus, higher executive function factor scores represent great impairment in executive function. The two individual SIP items 5 and 6 did not load significantly onto the executive function factor. Factor loadings and associations between baseline ANA factors and baseline drinking variables for this final ANA baseline model are presented in Table 4.

Covariates with the baseline ANA ESEM model. We added baseline drinking covariates of alcohol consumption frequency (PHDD) and intensity (DDD) to the ESEM model. Negative emotionality was significantly associated with baseline PHDD ($r = 0.183$, $p = 0.030$) but not with baseline DDD ($r = -0.088$, $p = 0.346$). Incentive salience was significantly associated with PHDD ($r = 0.516$, $p < 0.001$) but not with DDD ($r = 0.056$, $p = 0.325$). Executive function was significantly associated with DDD ($r = -0.274$, $p = 0.002$) but not with PHDD ($r = 0.054$, $p = 0.492$). We next added study as a covariate, where study (ABQ DrinQ=0, ABQ Treat =1) represents non-treatment-seeking and treatment-seeking status, respectively. Negative emotionality was significantly associated with study ($r = 0.213$, $p < .001$) such that negative emotionality was higher among treatment-seekers. Incentive salience was also significantly associated with study ($r = 0.217$, $p = .022$) such that incentive salience was higher among treatment-seekers. Executive function was not associated with study. Finally, we added age as a covariate and found that incentive salience was significantly associated with age ($r = 0.234$, $p = .001$), such that older age was associated

with higher incentive salience. Negative emotionality and executive function were not associated with age.

Measurement Invariance Testing

Longitudinal measurement invariance. We examined whether the final ANA ESEM model (reported in Table 4) was measurement invariant across time from baseline to three months. First, we fit the ESEM model in the three-month follow-up sample and found good model fit to the data (Model $\chi 2$ (61) = 60.459, p = 0.496; RMSEA < 0.001; CFI = 1.000). Results from fitting configural, metric, scalar, and residual measurement invariance models indicated that the configural model fit the data well and the metric model did not fit significantly worse than the configural model ($\Delta\chi 2$ ($\Delta 36$) = 33.520, p = .587; ΔRMSEA = -0.003, ΔCFI = -0.001). There was some mixed evidence about the fit of the scalar and residual models. Although the $\chi 2$ difference test indicated that the scalar model fit the data significantly worse than the metric model and the residual model fit the data significantly worse than the scalar model, all other a priori model comparison indices suggested that there was not a large decrement in model fit from metric to scalar ($\Delta\chi 2$ ($\Delta 12$) = 42.282, p < .001; ΔRMSEA = 0.003, ΔCFI = 0.008) and from scalar to residual ($\Delta\chi 2$ ($\Delta 15$) = 48.811, p < .001; ΔRMSEA = 0.002, ΔCFI = 0.009) (a negative change in CFI $\geq$.01, Cheung & Rensvold, 2002; and a positive change in RMSEA $\geq$.015, Chen, 2007). Thus, we determined that there was adequate evidence for scalar measurement invariance in the ANA ESEM model over time to proceed to additional longitudinal analyses. Full measurement invariance results for the ANA ESEM model are presented in Table 5.

Sex measurement invariance. We tested whether the ANA ESEM model was measurement invariant by sex. Results showed that the metric invariance model did not fit

significantly worse than the configural model ($\Delta\chi2$ ($\Delta36$) = 38.746, p = .347; ΔRMSEA = 0.007, ΔCFI = 0.001), the scalar model did not fit significantly worse than the metric model ($\Delta\chi2$ ($\Delta12$) = 3.809, p = .987; ΔRMSEA = 0.006, ΔCFI = 0.005), and the residual model did not fit significantly worse than the scalar model ($\Delta\chi2$ ($\Delta18$) = 14.615, p = .688; ΔRMSEA = 0.002, ΔCFI = 0.009). Thus, the ANA ESEM model is measurement invariant by male and female sex (see Table 5).

Hispanic and non-Hispanic white measurement invariance. Next, we tested whether the ANA ESEM model was measurement invariant by ethnicity, specifically comparing Hispanic and non-Hispanic white participants. The configural model could not be estimated normally due to non-positive definite residual covariance matrices in the Hispanic subgroup; specifically, the impaired control scale indicator was estimated to have a factor loading greater than 1 and a negative residual variance in the Hispanic, but not the non-Hispanic white group. Thus, the model fit indices of the configural model are not trustworthy. However, when we added the additional model constraints required to estimate the metric invariance model, we found good model fit, (Model $\chi2$ (161) = 178.475, p = 0.164; RMSEA =.032; CFI =.989). Although we cannot say definitively that the metric model does not show significant decrement in model fit compared to the configural model, the metric model alone demonstrates good model fit and therefore we moved on to the next most restrictive model comparison. The scalar model did not fit significantly worse than the metric model ($\Delta\chi2$ ($\Delta12$) = 3.809, p = .987; ΔRMSEA = 0.006, ΔCFI = 0.005). The residual model fit significantly worse than the scalar model ($\Delta\chi2$ ($\Delta18$) = 49.945, p = <.001; ΔRMSEA = 0.032, ΔCFI = 0.021), which suggests that residual variances in the ANA ESEM model are not equivalent among Hispanic and non-Hispanic white participants. However,

taken together, there is sufficient evidence that the ANA ESEM model is scalar measurement invariant by Hispanic and non-Hispanic white ethnicity, which is a sufficient condition to move on to additional analyses testing group differences by ethnicity (see Table 5).

Latent Change Score Models

Measurement invariance of the drinking latent construct. Prior to testing the two-wave latent change score (2W-LCS) models, we first examined the measurement invariance of a two-indicator latent drinking factor model across time. Note that a two-indicator model is not identified, however the test of the two-indicator model across time is identified and measurement invariance was examined for the two-indicator model across time using robust maximum likelihood estimation (MLR). With MLR estimation, we compared models using the Satorra-Bentler Scaled Chi-Square difference test (Satorra & Bentler, 2010). The metric model did not fit significantly worse than the configural model ($p = .374$). The scalar model did fit somewhat worse than the metric model based on the chi-square difference test ($p = .040$). However, the magnitude of the differences in indicator intercepts was quite small between the unconstrained metric model and the constrained scalar model, and was reflective of the fact that on average, the mean PHDD decreased to a greater extent from baseline to three months, while the mean DDD decreased to a smaller extent[2]. The more restrictive residual model did not fit significantly worse than the scalar model that constrains residual variances of items ($p = .333$). Although there was evidence for scalar non-invariance, we found sufficient evidence for strict measurement invariance of the drinking latent factor over time. These measurement invariance results are presented in full in Table 6.

[2] In the metric model, unstandardized intercepts of the indicators were $PHDD_{baseline}=0.326$, $PHDD_{3\text{-}month}=0.237$, $DDD_{baseline}=6.154$, $DDD_{3\text{-}month}=5.671$. In the scalar model, unstandardized intercepts of the indicators constrained to equality over time were PHDD=0.281, DDD=5.963.

Univariate models. Prior to fitting bivariate 2W-LCS models, we fit univariate LCS models to each construct individually in order determine whether the mean or variance of the latent change factor was significantly different than zero. Thus, we fit univariate LCS models to examine latent change factors for negative emotionality, incentive salience, and executive function. We did not fit a univariate model for the drinking construct because with only two indicators (PHDD, DDD), this model could not be identified. Since the goal of the univariate LCS models is to estimate means and variances of latent change factors for each ANA construct, we adopted a model identification strategy of fixing the mean and variance of the latent ANA constructs to 0 and 1 respectively at baseline, and fixing both the mean and variance to 0 at three-month follow-up.

The univariate LCS model for negative emotionality fit the data adequately, $\chi 2(88) = 187.408$, $p < .001$, RMSEA = .068, CFI = .958. The mean latent negative emotionality change score was -0.444 ($p < 0.001$), suggesting that on average, individuals decreased in their negative emotionality over time, with significant variability among individuals, $\sigma^2 = 53.768$, SE = 7.432, $p < .001$. Baseline levels of negative emotionality were negatively associated with latent negative emotionality change score (r=-0.393, $p < 0.001$) such that those who had higher negative emotionality at baseline demonstrated a smaller amount of change in negative emotionality over time.

The univariate LCS model for incentive salience fit the data well, $\chi 2(4) = 7.382$, $p = .171$, RMSEA = .059, CFI = .991. The mean latent incentive salience change score was -0.538 ($p < 0.001$), suggesting that on average, individuals decreased in their incentive salience over time, with significant variability among individuals, $\sigma^2 = 0.177$, SE = 0.048, $p < 0.001$. Baseline levels of incentive salience were negatively associated with latent incentive

salience change score (r=-0.501, $p < 0.001$) such that those who had higher incentive salience at baseline demonstrated a smaller amount of change in incentive salience over time.

The univariate LCS model for executive function fit the data well, $\chi2(25) = 45.089$, $p < .01$, RMSEA = .057, CFI = .968. The mean latent executive function change score was = -0.560 (p = 0.056), suggesting that on average, individuals did not significantly change in their executive function score, with non-significant variability among individuals, $\sigma^2 = 0.008$, SE = 0.008 $p = 0.299$. Baseline levels of executive function were not associated with latent executive function change score (r = -0.164, $p = 0.448$). Taken together, the acceptable model fit indices for all three ANA factor's univariate LCS models warranted moving on to bivariate LCS analyses that examined the dynamic relationship between changes in drinking and changes in ANA factors.

Changes in negative emotionality and drinking. We fit a 2W-LCS model to determine whether changes in drinking from baseline to three-month follow-up predicted changes in negative emotionality over the same time period. This model fit the data well: $\chi2(145) = 272.819$, $p < .001$, RMSEA = .060, CFI = .952. In this bivariate model, the mean of the drinking latent change score factor was negative and significant, $\mu_{\Delta drink} = -.444$, $p < .001$, indicating that overall there were significant reductions in drinking over time in this sample. Of note, there was significant variability in latent drinking change, $\sigma^2 = .041$, $p < .001$, which suggests variation in how much individuals reduced their drinking from baseline to three-month follow-up. The mean of the negative emotionality latent change score factor was negative and significant, $\mu_{\Delta NE} = -.340$, $p < .001$, suggesting that overall there were significant reductions in negative emotionality over time. Similar to the latent drinking change score, there was significant variability in latent negative emotionality change score,

σ^2= 50.452, p < .001, indicating variability in how much individuals reduced their negative emotionality over time. Latent change scores in drinking predicted latent change scores in negative emotionality, b = .247, SE = .092, p = 0.016, such that greater reductions in drinking over time were related to greater reductions in negative emotionality over time. Figure 1 shows a scatterplot of latent change scores of drinking on the x-axis and latent change scores of negative emotionality on the y-axis.

When study was included as a covariate, we found that study was not associated with the drinking latent change score factor (r = -0.017, p = 0.789) nor the negative emotionality latent change score factor (r = -0.095, p = 0.279). When age was included as a covariate, we found that age was not associated with the drinking latent change score factor (r = 0.001, p = 0.988) nor the negative emotionality latent change score factor (r = -0.035, p = 0.668).

Changes in incentive salience and drinking. Next, we fit a 2W-LCS model to determine whether changes in drinking from baseline to three-month follow-up predict changes in incentive salience over the same time period. This model fit the data well: χ2(21) = 38.281, p = . 012, RMSEA = .058, CFI = .978. In this bivariate model, the mean of the drinking latent change score factor was negative and significant, $\mu_{\Delta drink}$ = -.435, p < .001, indicating that overall there were significant reductions in drinking over time. There was significant variability in latent drinking change, σ^2= .044, p < .001, suggesting variation in how much individuals reduced their drinking from baseline to three-month follow-up. The mean of the incentive salience latent change score factor was negative and significant, $\mu_{\Delta IS}$ = -.480, p < .001, suggesting that overall there were significant reductions in incentive salience over time. Similar to the latent drinking change score, there was significant variability in latent incentive salience change score, σ^2= .174, p < .001, indicating variation in how much

individuals reduced their incentive salience over time. Latent change scores in drinking did not predict latent change scores in incentive salience, b = .149, SE = .108, p = 0.177. Figure 2 shows a scatterplot of latent change scores of drinking on the x-axis and latent change scores of incentive salience on the y-axis.

When study was included as a covariate, we found that study was not associated with the drinking latent change score factor (r = -0.087, p = 0.181) nor the incentive salience latent change score factor (r = -0.035, p = 0.759). When age was included as a covariate, we found that age was not associated with the drinking latent change score factor (r = -0.045, p = 0.515) nor the incentive salience latent change score factor (r = 0.057, p = 0.529).

Changes in executive function and drinking. The final 2W-LCS model was fit to determine whether changes in drinking from baseline to three-month follow-up predict changes in executive function over the same time period. This model fit the data well: $\chi 2(58)$ = 95.038, p = .002, RMSEA = .051, CFI = .960. In this bivariate model, the mean of the drinking latent change score factor was negative and significant, $\mu_{\Delta drink}$ = -.450, p < .001, indicating that overall there were significant reductions in drinking over time in this sample. There was significant variability in latent drinking change, σ^2= .040, p < .001, which suggests that not all individuals reduced their drinking from baseline to three-month follow-up. The mean of the executive function latent change score factor was negative but nonsignificant, $\mu_{\Delta EF}$ = -.453, p = .078, suggesting that overall there was no significant change in executive function over time. There was not significant variability in latent executive function change score, σ^2= .008, p = .294, indicating that there were not significant between-person differences in latent executive function change over time. Latent change scores in drinking did not predict latent change scores in executive function, b = .277, SE = .180, p =

0.210. Figure 3 shows a scatterplot of latent change scores of drinking on the x-axis and latent change scores of executive function on the y-axis.

When study was included as a covariate, we found that study was not associated with the drinking latent change score factor (r = 0.001, p = 0.992) nor the executive function latent change score factor (r = -0.484, p = 0.065). When age was included as a covariate, we found that age was not associated with the drinking latent change score factor (r = 0.006, p = 0.938) nor the executive function latent change score factor (r = -0.226, p = 0.206).

Discussion

The Addictions Neuroclinical Assessment (ANA) framework was replicated in a sample of treatment-seeking and non-treatment seeking heavy drinkers using self-report, behavioral, and neuroimaging assessments of negative emotionality, incentive salience, and executive function. Each of these three domains was associated with alcohol use at baseline. On average, drinking, negative emotionality, and incentive salience all decreased significantly in this sample from baseline to three-month follow-up, although there was significant variability among individuals indicating that not all individuals changed their drinking, negative emotionality, and incentive salience in the same way. Executive function did not change significantly over time, and there was minimal variability in executive function change over time.

This study contributed to the ANA literature in several ways. Through an iterative model fitting process, we showed that a multimodal assessment battery combining self-report, behavioral, and functional neuroimaging measures may not be an appropriate strategy for assessing the three ANA domains, given poor fit of this model. The battery with solely neuroimaging measures may be adequate for capturing the three factor ANA model, but the

battery with self-report and behavioral measures had optimal model fit. We found that an ANA battery composed of self-report and behavioral measures was measurement invariant over time, and across both sex and ethnicity. Finally, our findings demonstrated that change in negative emotionality was associated with change in drinking, and neither changes in incentive salience nor executive function were related to change in drinking. Further, study (i.e., treatment-seeking status) and age were not significantly associated with change in drinking or ANA factors.

ANA Domains are Invariant

It is notable that this study found adequate measurement invariance of the three factor ANA model from baseline to three months, by male and female sex, and across Hispanic and non-Hispanic white ethnicity. Evidence for scalar and residual measurement invariance suggests that we are measuring the same constructs in the same way across time and various demographic groups. Testing the assumption of measurement invariance is a critical step in evaluating the psychometric properties of new assessments, and without testing measurement invariance, we can arrive at faulty conclusions in group comparisons, which has important implications for biological phenotypes of psychopathology (Moriarity et al., 2022). Given that we did find measurement invariance, it was acceptable to move on to subsequent analytic steps involving longitudinal models and group comparisons. We found that there were no main effects of sex or ethnicity on ANA domains at baseline, and the only significant interaction was that sex moderated the association between incentive salience and percent heavy drinking days (PHDD) such that there was a stronger positive association for men and a weaker association for women.

These findings are mostly congruent with prior ANA literature that has found no main effect of sex on incentive salience or executive function (DeMartini et al., 2021; Kwako et al., 2019; Nieto & Ray, 2022; Stein et al., 2021), and no main effect of sex on negative emotionality (DeMartini et al., 2021; Nieto et al., 2021). However, prior validation studies of the ANA domains have found main effects of sex on negative emotionality such that women scored higher (Votaw et al., 2020), which is consistent with the broader literature demonstrating that women with substance use disorders report higher levels of negative affect than men and that negative affect in women is particularly related to craving, alcohol use, and treatment outcomes (Guinle & Sinha, 2020). One potential explanation for why we did not find a main effect of sex on negative emotionality in this sample is that the majority of the sample was non-treatment seekers. Treatment-seeking status was associated with higher negative emotionality, so it is possible that the impact of sex on negative emotionality was obscured in the full sample if women exhibit higher negative emotionality only at higher levels of AUD severity.

Change in ANA Domains and Drinking Over Time

In terms of longitudinal findings, we fit latent change score models to examine the relationships between change over time in each of the three ANA latent factors and a drinking latent factor. We found significant reductions from baseline to three months for drinking, negative emotionality, and incentive salience, and change in negative emotionality was associated with change in drinking. However, the latent change factors for drinking, negative emotionality, and incentive salience all showed significant between-person variability, so it is clear that not all people change on these factors in the same way, if at all. This is consistent with previous literature demonstrating variation in drinking change

trajectories, including among those who do not seek alcohol treatment (Witkiewitz et al., 2014), as well as variability in how individuals themselves seek to change their alcohol use and define recovery (Tucker et al., 2020).

Executive function, unlike the other latent factors, did not change over time nor was there significant between-person variance in this pattern. Although there are well documented cognitive changes related to chronic heavy alcohol use (Crowe et al., 2019), and even long-term moderate alcohol use (Topiwala et al., 2017), it may be the case that three months is too short of a time period in which to observe measurable differences in cognitive performance in a sample of drinkers who, on average, are continuing to drink at high risk levels despite overall reductions in drinking. In previous studies, small effect size improvements in executive function have been observed in as little as six weeks for those entering treatment with predominantly abstinence goals (Bates et al., 2005). There is also evidence in the literature for individual variation in longer term change trajectories following addiction treatment, with some people showing delayed gain, continuous gain, or loss of gain in various executive function subdomains (Bates et al., 2013), so it is notable that in this study we found neither improvements nor decrements in our executive function factor during this time period. It may be that the changes in drinking observed in the current sample are too small to impact executive function. The lack of observed change in the executive function latent factor may also be explained by the sensitivity of tests used. The executive function factor, as we modeled it, influenced performance on a flanker task of attention and inhibitory control, a dimensional change card sort task of cognitive flexibility, a list sorting task of immediate recall and working memory sequencing, and self-report measure of delay discounting. Executive function is a broad construct and the tests we used tap into an array of

sub-domains, each with relatively low factor loadings (ranging in magnitude from 0.27 to 0.42). It may be the case that trying to model a unitary executive function construct provides a less sensitive tool to detect changes that may be happening in individual sub-domains. When thinking about precision medicine approaches, evidence-based treatments for AUD may target specific executive function domains and not others. It may be that more granular characterization of executive function sub-domains could be more sensitive phenotypic markers. For example, Cognitive Behavioral Therapy may target cognitive flexibility and pre-treatment cognitive flexibility may impact treatment response (Goodkind et al., 2016; Nagata et al., 2018). Mindfulness-based approaches might specifically impact attention (Chiesa et al., 2011) and inhibition (Millett et al., 2021).

Covariate Effects of Age and Treatment-Seeking Status

In each of the bivariate LCS models, we examined study (i.e., treatment-seeking status) and age as covariates to see whether these individual-level characteristics were related to the change relationships we observed. Consistently across all three LCS models, we found that neither study nor age was significantly associated with change in drinking or ANA factors. It may be the case that our sample size limited statistical power to detect the impact of treatment-seeking status, since the treatment seeking group (n=56) was substantially smaller than the non-treatment seeking group (n=189). Another factor that may influence these findings is the potential of assessment reactivity, or changing one's drinking in response to participating in a research study assessing alcohol use (Clifford et al., 2007). In this sense, even the non-treatment seeking study participants may be changing their drinking to a greater extent than the general population of non-treatment seeking drinkers, and therefore they may resemble the treatment-seekers with respect to drinking changes.

However, these findings may also suggest that even though treatment-seeking status is related to higher initial level of negative emotionality and incentive salience, individuals may change in these core ANA domains and drinking to a similar extent, whether or not they seek formal treatment. Age has been inconsistently related to ANA domains across the literature. In one sample, age was significantly associated with impairment in all three ANA factors (Nieto et al., 2021), while in another study age was not related to the ANA factors (Kwako et al., 2019), yet in a third, age was only associated with higher incentive salience (DeMartini et al., 2021), which was consistent with the findings of this study. Given these discrepancies and the possible confounding variables, it is difficult to draw firm conclusions about the association between age and ANA domains. In this study, age was associated with higher incentive salience, but not associated with change in drinking or ANA factors over time. Given that alcohol consumption is increasing among older adults in particular (Grant et al., 2017), it will be important to continue examining the validity and usefulness of the ANA as an alcohol precision medicine tool across age groups.

Clinical Implications

When examining the change in ANA factors over time, only negative emotionality change was associated with drinking change. This finding maps on to a large body of literature examining the role of negative emotionality as a predictor of alcohol outcomes, particularly alcohol treatment outcomes. Two recent secondary data analysis studies validating the ANA/AARDoC framework have examined the predictive validity of the negative emotionality domain. In a prospective observational study following individuals as they enter alcohol treatment, baseline negative emotionality was indirectly associated with greater drinking intensity one year later through higher coping motives (e.g., drinking to

relieve negative affective states) at six months (Votaw et al., 2021). Another study that looked at longitudinal outcomes from Project MATCH and the COMBINE Study found that baseline negative emotionality was associated with greater drinking intensity and frequency one year later, as well as worse psychosocial functioning outcomes three year later (Witkiewitz et al., under review). Taken together, these prior findings and the findings from this study suggest that the ANA negative emotionality domain may have important predictive value and relevance for precision medicine efforts.

At baseline, the treatment-seeking group had higher negative emotionality than the non-treatment seeking group, yet study (i.e., treatment-seeking status) was not associated with change in negative emotionality over time. This may seem counterintuitive given that alcohol treatments such as Cognitive Behavioral Therapy and Mindfulness-Based Relapse Prevention directly target negative affective factors, so one might expect greater change in negative emotionality among treatment-seekers. It may be that the relationship between treatment-seeking status and change in negative emotionality would be observable over a longer period of time than three-months. Three months represents the time period in which participants were receiving treatment, not post-treatment, and there is some evidence to support a "sleeper effect" or delayed onset of improvements following treatment, suggesting implementation of treatment coping skills after the initial treatment phase (Carroll et al., 1994; Roos et al., 2020). It is also possible that we simply do not have the statistical power to detect a greater change in negative emotionality among the treatment-seekers due to smaller sample size. Although non-significant, the regression coefficient of latent negative emotionality change on study (coded 0 = non-treatment-seeking study ABQ DrinQ, 1 = treatment-seeking study ABQ Treat) was negative (r = -0.095, p = 0.283), suggesting that the

relationship was in the direction of greater negative emotionality reduction (the direction of latent change) among treatment seekers. Based on this effect size, we would need a sample size of greater than 800 people for power of .80 to detect a significant effect of treatment on change in negative emotionality.

Precision medicine, or the practice of matching people to treatments that best address their individual characteristics, may be an important avenue for improving AUD treatment outcomes. For example, prior precision addiction medicine work has demonstrated that naltrexone is more effective for individuals whose drinking is primarily driven by positive reinforcement or reward (Mann et al., 2018; Witkiewitz et al., 2019), and acamprosate may be more effect for those who drink primarily for negative reinforcement (Roos et al., 2017). Considering the importance of negative emotionality in predicting drinking outcomes, future precision medicine research should address how to best harness assessment of the ANA negative emotionality domain in matching individuals with especially high negative emotionality to treatments that effectively address both substance use and negative affectivity. There are numerous evidence-based treatments that have already been developed and tested that target substance use and negative emotionality directly or indirectly. Examples include Cognitive Behavior Therapy for Alcohol Problems (Epstein & McCrady, 2009), Mindfulness-Based Relapse Prevention (Bowen et al., 2021), Community Reinforcement Approach (Myers & Smith, 1995), Concurrent Treatment of PTSD and Substance Use Disorders using Prolonged Exposure (Back et al., 2014), and Mindfulness-Oriented Recovery Enhancement (Garland, 2013). Across many clinical trials, negative emotionality has been shown to be a predictor and moderator of treatment outcomes, a

mechanism of treatment, and a specific treatment outcome target in and of itself (for a review of recent clinical trials of these treatments, see: Swan et al., 2020).

In fact, the relevance of negative emotionality as a critical predictor of difficulty in substance use outcomes can be traced back to Marlatt's foundational work on relapse prevention (Marlatt, 1978). At the time, proponents of the disease model of addiction, the prevailing model of that time, put forth that alcohol relapse was driven by a compulsive and physiological need for the drug, therefore carrying the assumption that relapse was inevitable. In a study with individuals who went through AUD treatment and later returned to drinking, Marlatt documented the reasons and situations that preceded a return to drinking and found that negative emotional states were the most common proximal cause. This represented a turning point in the field of addiction treatment towards developing more hopeful, evidence-based treatments that target negative emotionality as a modifiable risk factor for returning to drinking (Donovan & Witkiewitz, 2012). At present, we have clearly identified that negative emotionality predicts relevant outcomes in addiction treatment and we have treatments specifically geared towards treating these symptoms. However, significant research and treatment gaps still exist. Only about 25% of those who identify alcohol as a problem seek treatment or community resources (Tucker & Witkiewitz, 2022) and among those who do enter treatment, higher baseline rates of negative emotionality are associated with prematurely leaving treatment (Syan et al., 2020). Thus, there is a need to better identify those individuals who experience both high negative emotionality and problematic alcohol use, including outside of traditional mental health treatment settings, and elucidate how we can better facilitate effective treatment delivery.

Future AUD precision medicine efforts may also go beyond the simple practice of matching individuals to treatments based on one-time baseline assessment of characteristics such as negative emotionality, reward/relief drinking, or other ANA domains. As we learn more about the dynamic temporal relationships between ANA domains, drinking, as well as important contextual factors, precision medicine should also include adaptive treatments based on more frequent assessment of relevant brain or behavioral changes (Lenze et al., 2020). In this current study we were limited by examining only two time points, albeit with thorough assessment measures at each. However, it is feasible to imagine a targeted ANA battery that could be delivered at frequent, relevant time intervals before, during, and after treatment such that specific treatment components could be delivered to the right person at the right time and monitored for their effectiveness. Lenze et al. provide a thorough review of precision clinical trial methodology and a proposed research agenda (Lenze et al., 2021), and findings from this current study provide preliminary support for the ANA framework to be well-suited for these types of approaches.

Multimodal Data Integration: Strengths and Limitations of Current Study

The ANA was originally proposed as a heuristic framework and assessment strategy that would integrate data at multiple levels of analysis in order to capture greater information about the heterogeneity of addictive disorders, and specifically, the acquired impairments in addiction cycle domains (Kwako et al., 2016). As hypothesized in the ANA, combining neuroscience-based assessments with self-report and behavioral assessments would generate greater individual phenotypic information. However, the findings of this study suggest that the reality of multimodal data integration is more complicated than the theory. Our original baseline exploratory structural equation model that combined functional neuroimaging with

self-report and/or behavioral indicators for each ANA domain did not fit the data well. For each ANA domain, the neuroimaging parameters did not load onto the same factors as the corresponding self-report or behavioral measures. The only exception was that list sorting working memory had a weak but significant factor loading on the same factor as the stop signal neuroimaging indicators. We were not able to locate other studies that have successfully derived latent factor structures using functional neuroimaging, self-report, and behavioral indicators together.

Although we ultimately settled on a baseline ANA model that included only self-report and behavioral indicators because it showed the best fit to the data, we found adequate model fit of an ANA battery composed of task-based functional neuroimaging indicators only. These were percent BOLD signal change in regions of interest for a cue reactivity alcohol > neutral image contrast for incentive salience, negative > neutral image contrast for negative emotionality, and response inhibition in a stop signal task for executive function. This finding validates recent calls in the literature that propose a neuroimaging-based battery to examine underlying mechanisms in AUD that contribute to heterogeneity and to inform precision medicine targets (Voon et al., 2020). However, given the large body of literature on alterations in brain structure and function as a consequence of alcohol use, substantially larger data sets than this one would be needed in order to compare the relative utility of multiple different task-based, structural, and resting state functional connectivity measures in achieving these goals.

The lack of fit of a model that combines neuroimaging indicators with self-report and behavioral measures is consistent with a pattern in the broader literature of low to no correlations between self-report and/or behavioral measures and neuroimaging measures of

the same construct. In an example that is relevant to this particular study, many studies have found low to no correspondence between self-reported negative affectivity and amygdala reactivity to emotional stimuli (MacDuffie et al., 2019). A recent study combined brain imaging data from approximately 50,000 participants and found that brain-behavior phenotype effect sizes for a wide range of constructs are much smaller than previously thought, and require enormous sample sizes to generate reproducible, reliable findings. There are many potential reasons for the lack of correspondence between self-reported and neurobiological measurements of what is supposedly the same trait, including psychometric inconsistencies between the assumptions of each type of measure, self-report measures that measure too broad a psychological construct rather than a more specific trait theorized to correspond to a neurobiological feature, and lack of consideration for the contextual nature of many psychological and neurobiological phenomenon (for a thorough review of ten methodological and conceptual problems that contribute to the gap between self-report and neuroscientific measures, see Brandt & Mueller, 2022).

One limitation of this study in particular that may have contributed to non-correspondence between self-report/behavioral and neuroimaging ANA indicators was that we only included neuroimaging variables from the non-treatment-seeking ABQ DrinQ sample. This was due to time constraint, as ABQ DrinQ enrollment and data collection was complete at the time of the current study's data analysis, while enrollment and data collection was ongoing for the treatment-seeking ABQ Treat sample and thus neuroimaging data was not available because it had not yet undergone the full processing pipeline. It is possible that had we included neuroimaging variables from the ABQ Treat sample, which was slightly more clinically severe than the ABQ DrinQ sample (e.g., higher percent heavy drinking

days), we might have been able to detect a clearer signal for an ANA model incorporating neuroimaging indicators. However, given the broader literature reviewed above and the relatively small differences in AUD severity between the two parent study samples, it is unlikely this would have changed the overall results.

Additional work should be done to effectively overcome the current methodological obstacles to integrating self-report, behavioral, and neuroimaging measures. Yet at the same time, the current study suggests that we have practical ways to assess ANA domains with self-report and behavioral measures that are associated with drinking, measured equivalently across sex and ethnicity, and, thus, may be sufficient for making progress in addiction precision medicine.

Limitations

There are several important limitations to this study. As a secondary data analysis study, the assessment measures were not specifically selected with the ANA in mind. Although the two parent studies, funded prior to the proposal of the ANA, were designed to test the original addiction cycle model (Koob & Volkow, 2016) from which the ANA was derived and we selected measures that had good theoretical fit with all three ANA constructs and good psychometric properties, we did not use empirical methods to evaluate which measures are optimal for achieving the goals of the ANA. Furthermore, although the measures we selected may indeed assess acquired traits as hypothesized by the addiction cycle, they may also capture dispositional vulnerabilities. Large scale, longitudinal study designs that account for pre-alcohol use initiation characteristics, such as the ABCD study (Saragosa-Harris et al., 2022), may be better equipped to truly assess acquired addiction cycle impairments and disentangle them from premorbid factors.

In addition, the sample size for this analytic approach was small and we were unable to test the measurement model using a split half replication. Small sample size was due in part to recruitment delays because of the COVID-19 pandemic. Another limitation to this work was that we used data from two timepoints only and a relatively short longitudinal window of three months. Most cutting-edge longitudinal analytic methods require three or four time points at minimum in order to draw more specific conclusions about the nature of change and change relationships (e.g., non-linear, quadratic) (Grimm et al., 2011). With only two time points, the findings reported here can be considered preliminary and require further validation. This is especially true given some of the drawbacks of using a latent drinking factor in longitudinal analysis that may not be fully scalar invariant, even though there was some evidence that it was residual invariant. Given the other analytic options available with two timepoints (e.g., two wave panel models), two wave latent change score analysis is the optimal method for modeling error, generating both within- and between-person information, and differentiating between intraindividual change and residualized change (Henk & Castro-Schilo, 2016). Furthermore, the three-month time period is relevant, though it is relatively short. Many common evidence-based treatments for AUD are intended to be delivered over a three-month span, so it would be reasonable to expect that core features associated with heavy alcohol use could change substantially in this time period. Likewise, prior research examining non-treatment seekers' alcohol use trajectories suggests that subsets of this population decrease either gradually or drastically during short time periods such as six months (Witkiewitz et al., 2014).

In addition, the incentive salience domain may not be completely well characterized by the measures used in this study. The final ESEM ANA model in this study relied on

measures of alcohol craving and self-reported failed control over alcohol behavior for the factor that we describe as representing incentive salience. This is consistent with how prior studies have utilized self-report measures of craving and impairments in alcohol self-regulation in the face of alcohol stimuli (Nieto et al., 2021; Stein et al., 2021). However, these self-report measures may not adequately capture incentive salience, which is the neurobiologically-mediated process of attributing motivational value to stimuli. The construct of incentive salience is supported by a large body of animal and human laboratory research showing that mesocorticolimbic dopamine systems underlie the process of substance and substance-related cues taking on higher motivational value (Cofresí et al., 2019; Zilverstand et al., 2018). At a certain level of sensitization, this "wanting" of substance-related stimuli can occur even in the face of not "liking" the substance (Berridge & Robinson, 2016). Although the neurobiological process of incentive salience likely corresponds to the phenomenological experience of increased craving and loss of behavioral control when faced with relevant stimuli, it is not clear that neurobiological changes in incentive salience rise to the level of conscious awareness to the extent that they can be described with self-report measures. Future work may find that neuroimaging or behavioral measures may be a more accurate way to capture the ANA incentive salience domain, and a more accurate incentive salience domain may show closer relationships to drinking change over time.

Conclusions

This study provides additional evidence for the validity and utility of the Addictions Neuroclinical Assessment. We identified a battery of self-report and behavioral measures that well capture the core ANA domains of negative emotionality, incentive salience, and

executive function. We also found that a neuroimaging battery, in the absence of self-report and behavioral measures, may also provide a reasonable battery for the ANA framework. By testing measurement invariance of the self-report and behavioral ANA battery, we contributed to the literature by establishing the ANA framework can be measured equivalently by female and male sex, Hispanic and non-Hispanic white ethnicity, and over time. We further demonstrated that these domains are associated with drinking cross-sectionally, that negative emotionality and incentive salience may change over a three-month time period while executive function may not, and that change in negative emotionality is related to change in drinking over time. Age and treatment-seeking status were not associated with these relationships. Future research should continue to study the utility of the ANA domains as a tool to characterize heterogeneity in AUD and to advance treatment matching based on individual differences and dynamic changes in these domains. In the emerging field of precision medicine for alcohol treatment, the ANA provides a potentially fruitful way for identifying those with particularly high negative emotionality and suggests that we should continue to develop treatments that adequately address the underlying mechanisms that drive negative emotionality.

Table 1. Sample characteristics for the full sample and by study

Measures	Total Sample (N=245)		Treatment-Seeking Participants (ABQ Treat, N=56)		Non-Treatment-Seeking Participants (ABQ Drinq, N=189)		*p*
	N	%/Mean (SD)	N	%/Mean (SD)	N	%/Mean (SD)	
Female	114	46.5%	32	57.1%	82	43.4%	.070
Male	131	53.5%	24	42.9%	107	56.6%	.070
Age	-	35.8 (10.5)	-	41.8 (10.2)	-	35.1 (10.0)	<.001
Hispanic	119	48.6%	25	44.6%	94	49.7%	.253
Non-Hispanic white	89	36.3%	24	42.9%	65	34.4%	.249
American Indian/Alaska Native	48	19.6%	9	16.1%	39	20.6%	.442
Asian	9	3.7%	1	1.8%	8	4.2%	.395
Black or African American	9	3.7%	0	-	9	4.8%	.096
Native Hawaiian or Other Pacific Islander	1	0.4%	0	-	1	0.5%	.586
Baseline past 90 day percent heavy drinking days (PHDD)	-	32.6% (29.5)	-	45.8% (34.1)	-	28.7% (26.9)	<.001
Baseline past 90 day drinks per drinking days (DDD)	-	6.2 (3.7)	-	6.6 (3.1)	-	6.0 (3.8)	.320

Table 2. Factor loadings for the first iteration of a three-factor baseline ESEM ANA model with neuroimaging, self-report, and behavioral indicators

Measures	Factor 1	Factor 2	Factor 3
Hypothesized Negative Emotionality Measures			
Sadness[a]	0.916*	-0.109	0.066
Anger Affect[a]	0.767*	-0.055	-0.085
Fear Affect[a]	0.904*	-0.057	-0.032
Loneliness[a]	0.798*	-0.042	0.025
Anger Hostility[a]	0.598*	-0.092	0.003
Positive Affect[a]	-0.732*	0.008	-0.030
Perceived Stress[a]	0.834*	-0.036	-0.072
Amygdala (left)[b]	0.237	0.812*	0.040
Ventromedial prefrontal cortex (left)[b]	0.138	0.277	0.133
Amygdala (right)l[b]	0.121	0.793*	0.118
Ventromedial prefrontal cortex (right)[b]	0.036	0.512*	0.077
Hypothesized Incentive Salience Measures			
ICS Failed Control	0.348*	-0.137	-0.078
Penn Alcohol Craving Scale	0.493*	-0.079	-0.043
Anterior cingulate cortex (left)[c]	0.091	0.075	0.242
Posterior cingulate cortex 1 (left) [c]	0.109	0.142	0.151
Posterior cingulate cortex 2 (left)c	0.040	0.092	0.239
Caudate (right)[c]	-0.054	0.006	0.059
Medial orbitofrontal gyrus (right)[c]	-0.008	0.104	0.113
Hypothesized Executive Function Measures			
Delay Discounting[d]	0.042	-0.096	-0.175
Flanker[a]	-0.034	0.134	0.103
List sorting working memory[a]	0.066	0.288*	0.269*
Dimensional change card sort[a]	0.032	0.113	0.071
SIP Item 5: I have taken foolish risks when I have been drinking.	0.279*	-0.164	0.025
SIP Item 6: When drinking, I have done impulsive things that I regretted later.	0.386*	-0.107	-0.081
Inferior frontal gyrus, opercular part 1 (right)[e]	0.000	-0.114	0.826*
Inferior frontal gyrus, opercular part 2 (right) [e]	-0.086	-0.075	0.657*
Insula (right) [e]	0.058	-0.085	0.794*
Median cingulate and paracingulate (right)[e]	0.031	-0.097	0.797*

* $p < .05$

[a]NIH Toolbox Emotion Battery and Cognition Battery

[b]fMRI Visual Cue Task, Negative > Neutral contrast

[c]fMRI Visual Cue Task, Alcohol > Neutral contrast

[d]Monetary Choice Questionnaire

[e]fMRI Stop Signal Task, response inhibition

Table 3. Factor loadings for the second iteration of a three-factor baseline ESEM ANA model with only neuroimaging indicators

Indicators	Negative Emotionality	Incentive Salience	Executive Function
Hypothesized Negative Emotionality Measures			
Amygdala (left)[a]	0.890*	-0.067	-0.031
Ventromedial prefrontal cortex (left) [a]	0.247*	0.216	0.000
Amygdala (right)l [a]	0.793*	-0.021	0.062
Ventromedial prefrontal cortex (right) [a]	0.476*	0.076	-0.040
Hypothesized Incentive Salience Measures			
Anterior cingulate cortex (left)[b]	0.064	0.567*	-0.021
Posterior cingulate cortex 1 (left) [b]	0.041	0.580*	-0.008
Posterior cingulate cortex 2 (left) [b]	-0.010	0.637*	0.105
Caudate (right) [b]	-0.080	0.298*	-0.027
Medial orbitofrontal gyrus (right) [b]	0.010	0.497*	-0.038
Hypothesized Executive Function Measures			
Inferior frontal gyrus, opercular part 1 (right)[c]	-0.007	-0.004	0.802*
Inferior frontal gyrus, opercular part 2 (right) [c]	-0.014	0.120	0.608*
Insula (right) [c]	0.038	-0.157*	0.851*
Median cingulate and paracingulate (right) [c]	-0.019	0.094	0.752*

* $p < .05$

[a]fMRI Visual Cue Task, Negative > Neutral contrast

[b]fMRI Visual Cue Task, Alcohol > Neutral contrast

[c]fMRI Stop Signal Task, response inhibition

Table 4. Factor loadings for the third and final iteration of a three-factor baseline ESEM ANA model with self-report and behavioral indicators only

Measures	Negative Emotionality	Incentive Salience	Executive Function
Hypothesized Negative Emotionality Measures			
Sadness[a]	0.925*	0.020	-0.084
Anger Affect[a]	0.779*	-0.036	0.118
Fear Affect[a]	0.908*	0.010	-0.026
Loneliness[a]	0.855*	-0.078	-0.110
Anger Hostility[a]	0.629*	-0.092	0.282
Positive Affect[a]	-0.675*	-0.158*	0.175*
Perceived Stress[a]	0.863*	-0.060	0.075
Hypothesized Incentive Salience Measures			
ICS Failed Control	-0.102*	0.972*	0.094
Penn Alcohol Craving Scale	0.156*	0.740*	0.006
Hypothesized Executive Function Measures			
Delay Discounting[b]	-0.012	0.042	0.406*
Flanker[a]	-0.044	0.037	-0.270*
List sorting working memory[a]	0.033	0.052	-0.390*
Dimensional change card sort[a]	0.038	0.046	-0.421*
SIP Item 5: I have taken foolish risks when I have been drinking.	0.110	0.351*	0.168
SIP Item 6: When drinking, I have done impulsive things that I regretted later.	0.219*	0.341*	0.150
Drinking Covariates			
Baseline past 90 days percent heavy drinking days[c]	0.183*	0.516*	0.054
Baseline past 90 days drinks per drinking day[c]	-0.088	0.056	-0.274*

* $p < .05$

[a]NIH Toolbox Emotion Battery and Cognition Battery

[b]Monetary Choice Questionnaire

[c]Timeline Followback

Table 5. Measurement invariance testing of the ANA model over time (baseline to three months), by sex (male, female), and by ethnicity (Hispanic, non-Hispanic white)

Model	$\chi 2$	df	RMSEA	ΔRMSEA	CFI	ΔCFI	$\chi 2$ difference test
Time							
Configural	538.622	327	0.051	-	0.943	-	-
Metric	572.142	363	0.048	-.003	0.944	-.001	33.52 (36) $p = 0.5871$
Scalar	614.424	375	0.051	.003	0.936	.008	42.282 (12), $p < 0.001$
Residual	663.235	390	0.053	.002	.927	.009	48.811 (15) $p < 0.001$
Sex							
Configural	167.885	122	0.055	-	0.973	-	-
Metric	202.805	158	0.048	-.007	0.974	.001	38.746 (36) $p = 0.347$
Scalar	206.736	170	0.042	-.006	0.979	.005	3.809 (12) $p = 0.987$
Table Residual	221.351	188	0.038	-.004	0.981	,002	14.615 (18) $p = 0.688$
Ethnicity							
Configural	115.220	122	0.000	-	1.000	-	-
Metric	167.734	158	0.024	.024	0.994	.006	53.283 (36) $p = 0.0318$
Scalar	171.757	170	0.010	-.014	0.999	.005	7.556 (12) $p = 0.8188$
Residual	221.702	188	0.042	.032	0.978	-.021	49.945 (18) $p < .001$

Table 6. Measurement invariance testing of the latent drinking factor models over time (baseline to three months)

Model	Loglikelihood	Scaling correction	k	p-value compared to configural	p-value compared to metric	p-value compared to scalar
Configural	-1140.34	3.5561	13	-	-	-
Metric	-1141.788	3.9349	11	0.374	-	-
Scalar	-1150.516	4.2321	9	0.040	0.035	-
Residual	-1152.888	4.8248	7	0.060	0.053	0.333

Figure 1. Scatterplot of the association between latent negative emotionality change and latent drinking change.

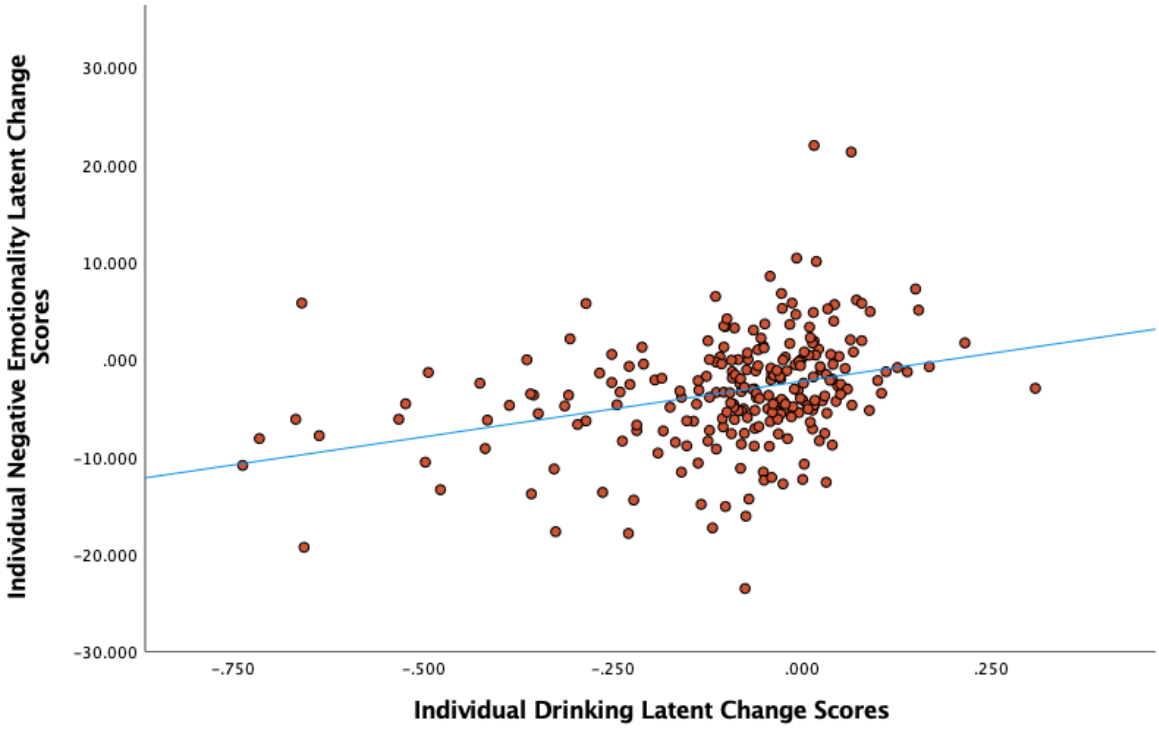


Figure 2. Scatterplot of the association between latent incentive salience change and latent drinking change.

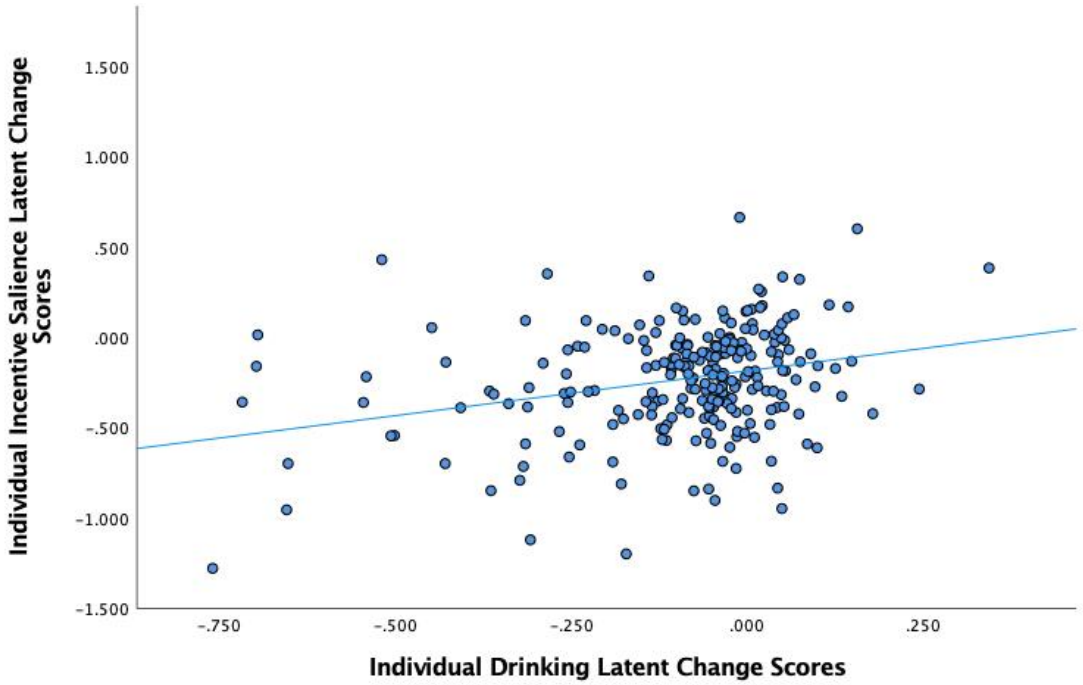
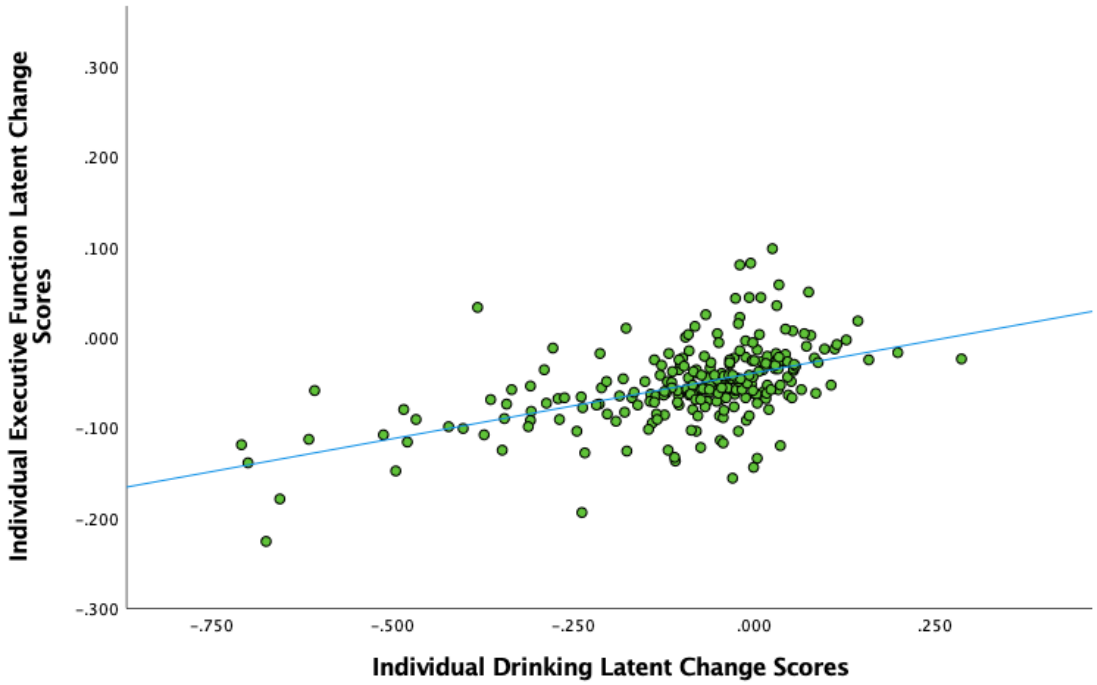


Figure 3. Scatterplot of the association between latent executive function change and latent drinking change.



References

Al-Khalil, K., Vakamudi, K., Witkiewitz, K., & Claus, E. D. (2021). Neural correlates of alcohol use disorder severity among nontreatment-seeking heavy drinkers: An examination of the incentive salience and negative emotionality domains of the alcohol and addiction research domain criteria. *Alcoholism: Clinical and Experimental Research*, *45*(6), 1200–1214. https://doi.org/10.1111/acer.14614

American Psychiatric Association. (2013). *Diagnostic and statistical manual ofmental disorders (DSM–5)*.

Back, S. E., Foa, E. B., Killeen, T. K., Mills, K. L., Teesson, M., Cotton, B. D., Carroll, K. M., & Brady, K. T. (2014). *Concurrent Treatment of PTSD and Substance Use Disorders Using Prolonged Exposure (COPE)*. Oxford University Press. https://doi.org/10.1093/med:psych/9780199334537.001.0001

Bates, M. E., Buckman, J. F., Voelbel, G. T., Eddie, D., & Freeman, J. (2013). The mean and the individual: Integrating variable-centered and person-centered analyses of cognitive recovery in patients with substance use disorders. *Frontiers in Psychiatry*, *4*(DEC), 11–14. https://doi.org/10.3389/fpsyt.2013.00177

Bates, M. E., Voelbel, G. T., Buckman, J. F., Labouvie, E. W., & Barry, D. (2005). Short-term neuropsychological recovery in clients with substance use disorders. *Alcoholism: Clinical and Experimental Research*, *29*(3), 367–377. https://doi.org/10.1097/01.ALC.0000156131.88125.2A

Becker, H. C. (2015). Challenges and exciting new opportunities to advance personalized treatment for alcohol use disorder. *Alcoholism: Clinical and Experimental Research*, *39*(4), 587–588. https://doi.org/10.1111/acer.12664

Becker, J. B., & Koob, G. F. (2016). Sex Differences in Animal Models: Focus on Addiction. *Pharmacol Rev*, *68*(2), 242–263. https://doi.org/10.1124/pr.115.011163

Berridge, K. C. (2012). From prediction error to incentive salience: Mesolimbic computation of reward motivation. *European Journal of Neuroscience*, *35*(7), 1124–1143. https://doi.org/10.1111/j.1460-9568.2012.07990.x

Berridge, K. C., & Robinson, T. (2016). Liking, wanting and the incentive salience theory of addiction. *American Psychologist*, *71*(8), 670–679.

Bickel, W. K., Jarmolowicz, D. P., Mueller, E. T., Gatchalian, K. M., & McClure, S. M. (2012). Are executive function and impulsivity antipodes? A conceptual reconstruction with special reference to addiction. *Psychopharmacology*, *221*(3), 361–387. https://doi.org/10.1007/s00213-012-2689-x

Boness, C. L., Watts, A. L., Moeller, K. N., & Sher, K. J. (2021). The Etiologic, Theory-Based, Ontogenetic Hierarchical Framework of Alcohol Use Disorder: A Translational Systematic Review of Reviews. In *Psychological Bulletin* (Vol. 147, Issue 10). https://doi.org/10.1037/bul0000333

Bowen, S., Chawla, N., Grow, J., & Marlatt, G. A. (2021). *Mindfulness-Based Relapse Prevention for Addictive Behaviors* (2nd ed.). Guilford Press.

Brandt, A., & Mueller, E. M. (2022). Negative affect related traits and the chasm between self-report and neuroscience. In *Current Opinion in Behavioral Sciences* (Vol. 43, pp. 216–223). https://doi.org/10.1016/j.cobeha.2021.11.002

Buchanan, T. (2016). Self-report measures of executive function problems correlate with personality, not performance-based executive function measures, in non-clinical samples. *Psychological Assessment*, *28*(4), 372–385.

Carroll, K. M., Rounsaville, B. J., Nich, C., Gordon, L. T., Wirtz, P. W., & Gawin, F. (1994). One-Year Follow-up of Psychotherapy and Pharmacotherapy for Cocaine Dependence: Delayed Emergence of Psychotherapy Effects. *Archives of General Psychiatry*, *51*(12), 989–997. https://doi.org/10.1001/archpsyc.1994.03950120061010

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiesa, A., Calati, R., & Serretti, A. (2011). Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. In *Clinical Psychology Review* (Vol. 31, Issue 3, pp. 449–464). Pergamon. https://doi.org/10.1016/j.cpr.2010.11.003

Clifford, P. R., Maisto, S. A., & David, C. M. (2007). Alcohol treatment research assessment exposure subject reactivity effects: Part I Alcohol use and related consequences. *Journal of Studies on Alcohol and Drugs*, *68*(4), 519–528. https://doi.org/10.15288/jsad.2007.68.519

Cofresí, R. U., Bartholow, B. D., & Piasecki, T. M. (2019). Evidence for incentive salience sensitization as a pathway to alcohol use disorder. *Neuroscience and Biobehavioral Reviews*, *107*(October), 897–926. https://doi.org/10.1016/j.neubiorev.2019.10.009

Cooper, M., Kuntsche, E., Levitt, A., Barber, L., & Wolf, S. (2015). Motivational Models of Substance Use: A Review of Theory and Research on Motives for Using Alcohol,

Marijuana, and Tobacco. In K. J. Sher (Ed.), *The Oxford Handbook of Substance Use Disorders* (pp. 375–421). https://doi.org/10.1093/oxfordhb/9780199381678.013.017

Crowe, S. F., Cammisuli, D. M., & Stranks, E. K. (2019). Widespread cognitive deficits in alcoholism persistent following prolonged abstinence: An updated meta-analysis of studies that used standardised neuropsychological assessment tools. *Archives of Clinical Neuropsychology*, *35*(1), 31–45. https://doi.org/10.1093/arclin/acy106

DeMartini, K. S., Gueorguieva, R., Pearlson, G., Krishnan-Sarin, S., Anticevic, A., Ji, L. J., Krystal, J. H., & O'Malley, S. S. (2021). Mapping data-driven individualized neurobehavioral phenotypes in heavy alcohol drinkers. *Alcoholism: Clinical and Experimental Research*, *012870*(February), 841–853. https://doi.org/10.1111/acer.14580

Dodell-Feder, D., Ressler, K. J., & Germine, L. T. (2020). Social cognition or social class and culture? on the interpretation of differences in social cognitive performance. *Psychological Medicine*, *50*(1), 133–145. https://doi.org/10.1017/S003329171800404X

Donovan, D., & Witkiewitz, K. (2012). Relapse prevention: From radical idea to common practice. *Addiction Research and Theory*, *20*(3), 204–217. https://doi.org/10.3109/16066359.2011.647133

Epstein, E. E., & McCrady, B. S. (2009). *A Cognitive-Behavioral Treatment Program for Overcoming Alcohol Use Problems*. Oxford University Press. https://doi.org/10.1093/med:psych/9780195322811.001.0001

First, M., Williams, J., Karg, R., & Spitzer, R. (2016). Structured Clinical Interview for DSM-5 Disorders, Clinician Version (SCID-5-CV). In *American Psychiatric Association*. https://doi.org/10.1002/9781118625392.wbecp351

Flannery, B. A., Volpicelli, J. R., & Pettinati, H. M. (1999). *Psychometric Properties of the*

Penn Alcohol Craving Scale. *23*(8), 1289–1295.

Garland, E. L. (2013). *Mindfulness-Oriented Recovery Enhancement for Addiction, Stress, and Pain* (1st ed.). NASW Press.

Ghitza, U. E. (2017). Commentary: Addictions Neuroclinical Assessment: A Neuroscience-Based Framework for Addictive Disorders. *Frontiers in Psychiatry*, *8*. https://doi.org/10.3389/fpsyt.2017.00002

Goodkind, M. S., Gallagher-Thompson, D., Thompson, L. W., Kesler, S. R., Anker, L., Flournoy, J., Berman, M. P., Holland, J. M., & O'Hara, R. M. (2016). The impact of executive function on response to cognitive behavioral therapy in late-life depression. *International Journal of Geriatric Psychiatry*, *31*(4), 334–339. https://doi.org/10.1002/gps.4325

Grant, B. F., Chou, S. P., Saha, T. D., Pickering, R. P., Kerridge, B. T., Ruan, W. J., Huang, B., Jung, J., Zhang, H., Fan, A., & Hasin, D. S. (2017). Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001-2002 to 2012-2013: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *JAMA Psychiatry*, *74*(9), 911–923. https://doi.org/10.1001/jamapsychiatry.2017.2161

Grant, B. F., Saha, T. D., Ruan, W. J., Goldstein, R. B., Chou, S. P., Jung, J., Zhang, H., Smith, S. M., Pickering, R. P., Huang, B., & Hasin, D. S. (2015). Epidemiology of DSM-5 Alcohol Use Disorder Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry*, *72*(8), 757–766. https://doi.org/10.1001/jamapsychiatry.2015.2132

Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, S. P., Dufour, M. C., Compton, W.,

Pickering, R. P., & Kaplan, K. (2004). Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: Results from the national epidemiologic survey on alcohol and related conditions. In *Archives of General Psychiatry* (Vol. 61, Issue 8, pp. 807–816). https://doi.org/10.1001/archpsyc.61.8.807

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*(5), 1357–1371. https://doi.org/10.1111/j.1467-8624.2011.01630.x

Grüsser, S. M., Wrase, J., Klein, S., Hermann, D., Smolka, M. N., Ruf, M., Weber-Fahr, W., Herta, F., Mann, K., Braus, D. F., & Heinz, A. (2004). Cue-induced activation of the striatum and medial prefrontal cortex is associated with subsequent relapse in abstinent alcoholics. *Psychopharmacology*, *175*(3), 296–302. https://doi.org/10.1007/s00213-004-1828-4

Guinle, M. I. B., & Sinha, R. (2020). The role of stress, trauma, and negative affect in alcohol misuse and alcohol use disorder in women. *Alcohol Research: Current Reviews*, *40*(2), 1–17. https://doi.org/10.35946/arcr.v40.2.05

Hallgren, K. A., & Witkiewitz, K. (2013). Missing data in alcohol clinical trials: A comparison of methods. *Alcoholism: Clinical and Experimental Research*, *37*(12), 2152–2160. https://doi.org/10.1111/acer.12205

Heather, N., Booth, P., & Luce, A. (1998). Impaired Control Scale: Cross-validation and relationships with treatment outcome. *Addiction*, *93*(5), 761–771.

Heather, N., Tebbutt, J. S., Mattick, R. P., & Zamir, R. (1993). Development of a scale for measuring impaired control over alcohol consumption: A preliminary report. *Journal of Studies on Alcohol*, *54*(6), 700–709. https://doi.org/10.15288/jsa.1993.54.700

Henk, C. M., & Castro-Schilo, L. (2016). Preliminary Detection of Relations Among Dynamic Processes With Two-Occasion Data. *Structural Equation Modeling*, *23*(2), 180–193. https://doi.org/10.1080/10705511.2015.1030022

Hutchison, K. E. (2008). Alcohol dependence: Neuroimaging and the development of translational phenotypes. *Alcoholism: Clinical and Experimental Research*, *32*(7), 1111–1112. https://doi.org/10.1111/j.1530-0277.2008.00691.x

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Kaplan, B. A., Amlung, M., Reed, D. D., Jarmolowicz, D. P., McKerchar, T. L., & Lemley, S. M. (2016). Automating Scoring of Delay Discounting for the 21- and 27-Item Monetary Choice Questionnaires. *Behavior Analyst*, *39*(2), 293–304. https://doi.org/10.1007/s40614-016-0070-9

Kiluk, B. D., Dreifuss, J. A., Weiss, R. D., Morgenstern, J., & Carroll, K. M. (2013). *The Short Inventory of Problems – Revised (SIP-R): Psychometric properties within a large, diverse sample of substance use disorder treatment seekers*. *27*(1), 307–314. https://doi.org/10.1037/a0028445.The

Kim, S. M., Han, D. H., Min, K. J., Kim, B. N., & Cheong, J. H. (2014). Brain activation in response to craving- and aversion-inducing cues related to alcohol in patients with alcohol dependence. *Drug and Alcohol Dependence*, *141*, 124–131. https://doi.org/10.1016/j.drugalcdep.2014.05.017

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount

rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General*, *128*(1), 78–87. https://doi.org/10.1037/0096-3445.128.1.78

Koob, G. F. (2021). Drug addiction: Hyperkatifeia/negative reinforcement as a framework for medications development. *Pharmacological Reviews*, *73*(1), 163–201. https://doi.org/10.1124/PHARMREV.120.000083

Koob, G. F., & Le Moal, M. (1997). Drug abuse: Hedonic homeostatic dysregulation. *Science*, *278*(5335), 52–58. https://doi.org/10.1126/science.278.5335.52

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, *24*(2), 97–129. https://doi.org/10.1016/S0893-133X(00)00195-0

Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: a neurocircuitry analysis. In *The Lancet Psychiatry* (Vol. 3, Issue 8, pp. 760–773). Elsevier Ltd. https://doi.org/10.1016/S2215-0366(16)00104-8

Krystal, J. H., & O'Malley, S. S. (2015). From translational neuroscience to personalized medicine. *Alcoholism: Clinical and Experimental Research*, *39*(4), 585–586. https://doi.org/10.1111/acer.12667

Kvamme, T. L., Pedersen, M. U., Overgaard, M., Rømer Thomsen, K., & Voon, V. (2019). Pupillary reactivity to alcohol cues as a predictive biomarker of alcohol relapse following treatment in a pilot study. *Psychopharmacology*, *236*(4), 1233–1243. https://doi.org/10.1007/s00213-018-5131-1

Kwako, L. E., Momenan, R., Grodin, E. N., Litten, R. Z., Koob, G. F., & Goldman, D. (2017). Neuropharmacology Addictions Neuroclinical Assessment : A reverse translational approach. *Neuropharmacology*, *122*, 254–264.

Kwako, L. E., Momenan, R., Litten, R. Z., Koob, G. F., & Goldman, D. (2016). Addictions Neuroclinical Assessment: A Neuroscience-Based Framework for Addictive Disorders. *Biological Psychiatry*, *80*(3), 179–189. https://doi.org/10.1016/j.biopsych.2015.10.024

Kwako, L. E., Schwandt, M. L., Ramchandani, V. A., Diazgranados, N., Koob, G. F., Volkow, N. D., Blanco, C., & Goldman, D. (2019). Neurofunctional Domains Derived From Deep Behavioral Phenotyping in Alcohol Use Disorder. *American Journal of Psychiatry*, *176*(9), 744–753. https://doi.org/10.1176/appi.ajp.2018.18030357

Lane, S. P., & Sher, K. J. (2015). Limits of current approaches to diagnosis severity based on criterion counts: An example with DSM-5 alcohol use disorder. *Clinical Psychological Science*, *3*(6), 819–835. https://doi.org/10.1177/2167702614553026

Lenze, E. J., Nicol, G. E., Barbour, D. L., Kannampallil, T., Wong, A. W. K., Piccirillo, J., Drysdale, A. T., Sylvester, C. M., Haddad, R., Miller, J. P., Low, C. A., Lenze, S. N., Freedland, K. E., & Rodebaugh, T. L. (2021). Precision clinical trials: A framework for getting to precision medicine for neurobehavioural disorders. *Journal of Psychiatry and Neuroscience*, *46*(1), 97–110. https://doi.org/10.1503/jpn.200042

Lenze, E. J., Rodebaugh, T. L., & Nicol, G. E. (2020). A Framework for Advancing Precision Medicine in Clinical Trials for Mental Disorders. *JAMA Psychiatry*, *77*(7), 663–664.

Linden-Carmichael, A. N., Dziak, J. J., & Lanza, S. T. (2019). Dynamic features of problematic drinking: Alcohol use disorder latent classes across ages 18-64. *Alcohol and Alcoholism*, *54*(1), 97–103. https://doi.org/10.1093/alcalc/agy074

Litten, R. Z., Ryan, M. L., Falk, D. E., Reilly, M., Fertig, J. B., & Koob, G. F. (2015). Heterogeneity of alcohol use disorder: Understanding mechanisms to advance

personalized treatment. *Alcoholism: Clinical and Experimental Research*, *39*(4), 579–584. https://doi.org/10.1111/acer.12669

Lowman, C., Allen, J., Stout, R. L., Connors, G., Longabaugh, R., Maisto, S. A., Miller, W. R., Rubin, A., Stour, R. L., Waldron, D., Westerberg, V. S., & Zywiak, W. H. (1996). Replication and extension of Marlatt's taxonomy of relapse precipitants: Overview of procedures and results. *Addiction*, *91*, S51–S71. https://doi.org/10.1111/j.1360-0443.1996.tb02327.x

Luciano, M. T., Acuff, S. F., Olin, C. C., Lewin, R. K., Strickland, J. C., McDevitt-Murphy, M. E., & Murphy, J. G. (2022). Posttraumatic stress disorder, drinking to cope, and harmful alcohol use: A multivariate meta-analysis of the self-medication hypothesis. *Journal of Psychopathology and Clinical Science*. https://doi.org/10.1037/abn0000764

MacDuffie, K. E., Knodt, A. R., Radtke, S. R., Strauman, T. J., & Hariri, A. R. (2019). Self-rated amygdala activity: an auto-biological index of affective distress. *Personality Neuroscience*, *2*, 1–7. https://doi.org/10.1017/pen.2019.1

Mann, K., Roos, C. R., Hoffmann, S., Nakovics, H., Leménager, T., Heinz, A., & Witkiewitz, K. (2018). Precision Medicine in Alcohol Dependence: A Controlled Trial Testing Pharmacotherapy Response among Reward and Relief Drinking Phenotypes. *Neuropsychopharmacology*, *43*(4), 891–899. https://doi.org/10.1038/npp.2017.282

Marlatt, G. A. (1978). Craving for alcohol, loss of control, and relapse: A cognitive-behavioral analysis. In *Alcoholism: New directions in research and treatment* (pp. 271–314). Plenum Press.

Millett, G., D'Amico, D., Amestoy, M. E., Gryspeerdt, C., & Fiocco, A. J. (2021). Do group-based mindfulness meditation programs enhance executive functioning? A systematic

review and meta-analysis of the evidence. *Consciousness and Cognition*, *95*(April), 103195. https://doi.org/10.1016/j.concog.2021.103195

Mineur, Y. S., Garcia-Rivas, V., Thomas, M. A., Soares, A. R., McKee, S. A., & Picciotto, M. R. (2022). Sex differences in stress-induced alcohol intake: a review of preclinical studies focused on amygdala and inflammatory pathways. *Psychopharmacology*. https://doi.org/10.1007/s00213-022-06120-w

Moriarity, D. P., Joyner, K. J., Slavich, G. M., & Alloy, L. B. (2022). Unconsidered issues of measurement noninvariance in biological psychiatry: A focus on biological phenotypes of psychopathology. *Molecular Psychiatry*, *June*, 1–5. https://doi.org/10.1038/s41380-021-01414-5

Muthén, L. K., & Muthén, B. O. (2017). *Mplus users guide (Version 8)*. Muthén & Muthén.

Myers, R. J., & Smith, J. E. (1995). *Clinical Guide to Alcohol Treatment: The Community Reinforcement Approach* (1st ed.). Guilford Press.

Nagata, S., Seki, Y., Shibuya, T., Yokoo, M., Murata, T., Hiramatsu, Y., Yamada, F., Ibuki, H., Minamitani, N., Yoshinaga, N., Kusunoki, M., Inada, Y., Kawasoe, N., Adachi, S., Oshiro, K., Matsuzawa, D., Hirano, Y., Yoshimura, K., Nakazato, M., … Shimizu, E. (2018). Does cognitive behavioral therapy alter mental defeat and cognitive flexibility in patients with panic disorder? UMIN000022693 UMIN. *BMC Research Notes*, *11*(1). https://doi.org/10.1186/s13104-018-3130-2

National Advisory Mental Health Council Workgroup on Changes to the Research Domain Criteria Matrix. (2018). *RDoC Changes to the Matrix (CMAT) Workgroup Update: Addition of the Sensorimotor Domain*.

National Institute on Alcohol Abuse and Alcoholism. (2022). *Alcohol Facts and Statistics*.

https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/alcohol-facts-and-statistics

Nieto, S. J., Grodin, E. N., Green, R. J., & Ray, L. A. (2021). Evaluation of the Addictions Neuroclinical Assessment (ANA) framework through deep phenotyping of problem drinkers. *Drug and Alcohol Dependence*, *221*(September 2020), 108603. https://doi.org/10.1016/j.drugalcdep.2021.108603

Nieto, S. J., & Ray, L. A. (2022). Applying the Addictions Neuroclinical Assessment to derive neurofunctional domains in individuals who use methamphetamine. *Behavioural Brain Research*, *427*(March), 113876. https://doi.org/10.1016/j.bbr.2022.113876

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Roos, C. R., Carroll, K. M., Nich, C., Frankforter, T., & Kiluk, B. D. (2020). Short- and long-term changes in substance-related coping as mediators of in-person and computerized CBT for alcohol and drug use disorders. *Drug and Alcohol Dependence*, *212*(December 2019). https://doi.org/10.1016/j.drugalcdep.2020.108044

Roos, C. R., Mann, K., & Witkiewitz, K. (2017). Reward and relief dimensions of temptation to drink: construct validity and role in predicting differential benefit from acamprosate and naltrexone. *Addiction Biology*, *22*(6), 1528–1539. https://doi.org/10.1111/adb.12427

Rossow, I., Bye, E. K., Moan, I. S., Kilian, C., & Bramness, J. G. (2021). Changes in alcohol consumption during the covid-19 pandemic—small change in total consumption, but increase in proportion of heavy drinkers. *International Journal of Environmental Research and Public Health*, *18*(8). https://doi.org/10.3390/ijerph18084231

Salsman, J. M., Butt, Z., Pilkonis, P. A., Cyranowski, J. M., Zill, N., Hendrie, H. C., Kupst, M. J., Kelly, M. A. R., Bode, R. K., Choi, S. W., Lai, J.-S., Griffith, J. W., Stoney, C. M., Brouwers, P., Knox, S. S., & Cella, D. (2013). Emotion assessment using the NIH Toolbox. *Neurology*, *80*(11 Suppl 3), S76-86. https://doi.org/10.1212/WNL.0b013e3182872e11

Saragosa-Harris, N. M., Chaku, N., MacSweeney, N., Guazzelli Williamson, V., Scheuplein, M., Feola, B., Cardenas-Iniguez, C., Demir-Lira, E., McNeilly, E. A., Huffman, L. G., Whitmore, L., Michalska, K. J., Damme, K. S., Rakesh, D., & Mills, K. L. (2022). A practical guide for researchers and reviewers using the ABCD Study and other large longitudinal datasets. *Developmental Cognitive Neuroscience*, *55*, 101115. https://doi.org/10.1016/j.dcn.2022.101115

Satorra, A., & Bentler, P. M. (2010). Ensuring possitiveness of the scaled chi-square test statistic. *Psychometrika*, *75*(2), 243–248.

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

Sawyer, K. S., Oscar-Berman, M., Barthelemy, O. J., Papadimitriou, G. M., Harris, G. J., & Makris, N. (2017). Gender dimorphism of brain reward system volumes in alcoholism. *Psychiatry Research - Neuroimaging*, *263*, 15–25. https://doi.org/10.1016/j.pscychresns.2017.03.001

Schacht, J. P., Anton, R. F., & Myrick, H. (2013). Functional neuroimaging studies of alcohol cue reactivity: A quantitative meta-analysis and systematic review. *Addiction*

Biology, *18*(1), 121–133. https://doi.org/10.1111/j.1369-1600.2012.00464.x

Sinha, R., Fox, H. C., Hong, K. A., Bergquist, K., Bhagwagar, Z., & Siedlarz, K. M. (2009). Enhanced negative emotion and alcohol craving, and altered physiological responses following stress and cue exposure in alcohol dependent individuals. *Neuropsychopharmacology*, *34*(5), 1198–1208. https://doi.org/10.1038/npp.2008.78

Sinha, R., Fox, H. C., Hong, K. I. A., Hansen, J., Tuit, K., & Kreek, M. J. (2011). Effects of adrenal sensitivity, stress- and cue-induced craving,and anxiety on subsequent alcohol relapse and treatment outcomes. *Archives of General Psychiatry*, *68*(9), 942–952. https://doi.org/10.1001/archgenpsychiatry.2011.49

Sobell, L. C., Agrawal, S., Annis, H., Ayala-Velazquez, H., Echeverria, L., Leo, G. I., Rybakowski, J. K., Sandahl, C., Saunders, B., Thomas, S., & Ziólkowski, M. (2001). Cross-cultural evaluation of two drinking assessment instruments: Alcohol timeline followback and inventory of drinking situations. *Substance Use and Misuse*, *36*(3), 313–331. https://doi.org/10.1081/JA-100102628

Sobell, L. C., & Sobell, M. B. (1996). *Timeline Followback user's guide: A calendar method for assessing alcohol and drug use*. Addiction Research Foundation.

Stein, E. R., Votaw, V. R., Swan, J. E., & Witkiewitz, K. (2021). Validity and measurement invariance of the Addictions Neuroclinical Assessment incentive salience domain among treatment-seekers with alcohol use disorder. *Journal of Substance Abuse Treatment*, *122*(May 2020), 108227. https://doi.org/10.1016/j.jsat.2020.108227

Swan, J. E., Votaw, V. R., Stein, E. R., & Witkiewitz, K. (2020). The Role of Affect in Psychosocial Treatments for Substance Use Disorders. *Current Addiction Reports*, *7*(2), 108–116. https://doi.org/10.1007/s40429-020-00304-0

Swartz, M., Burton, F., Vakamudi, K., Al-Khalil, K., Witkiewitz, K., & Claus, E. D. (2021). Age dependent neural correlates of inhibition and control mechanisms in moderate to heavy drinkers. *NeuroImage: Clinical*, *32*, 102875. https://doi.org/10.1016/j.nicl.2021.102875

Syan, S. K., Minhas, M., Oshri, A., Costello, J., Sousa, S., Samokhvalov, A. V., Rush, B., & MacKillop, J. (2020). Predictors of premature treatment termination in a large residential addiction medicine program. *Journal of Substance Abuse Treatment*, *117*(July), 108077. https://doi.org/10.1016/j.jsat.2020.108077

Topiwala, A., Allan, C. L., Valkanova, V., Zsoldos, E., Filippini, N., Sexton, C., Mahmood, A., Fooks, P., Singh-Manoux, A., Mackay, C. E., Kivimäki, M., & Ebmeier, K. P. (2017). Moderate alcohol consumption as risk factor for adverse brain outcomes and cognitive decline: longitudinal cohort study. *BMJ*, *142*(16), j2353. https://doi.org/10.1136/bmj.j2353

Tucker, J. A., Chandler, S. D., & Witkiewitz, K. (2020). Epidemiology of Recovery from Alcohol Use Disorder. *Alcohol Research: Current Reviews*, *40*(3). https://doi.org/https://doi.org/10.35946/arcr.v40.3.0

Tucker, J. A., & Witkiewitz, K. (2022). Historical and Contemporary Perspectives on Pathways to Recovery from Alcohol Use Disorder. In J. A. Tucker & K. Witkiewitz (Eds.), *Dynamic Pathways to Recovery from Alcohol Use Disorder* (pp. 3–21). Cambridge University Press.

U.S. Census Bureau. (2017). *2017 National Population Projections Tables: Main Series*. https://www.census.gov/data/tables/2017/demo/popproj/2017-summary-tables.html

Voon, V., Grodin, E., Mandali, A., Morris, L., Doñamayor, N., Weidacker, K., Kwako, L.,

Goldman, D., Koob, G. F., & Momenan, R. (2020). Addictions NeuroImaging Assessment (ANIA): Towards an integrative framework for alcohol use disorder. *Neuroscience and Biobehavioral Reviews*, *113*(March), 492–506. https://doi.org/10.1016/j.neubiorev.2020.04.004

Votaw, V. R., Pearson, M. R., Stein, E. R., & Witkiewitz, K. (2020). The Addictions Neuroclinical Assessment Negative Emotionality Domain among Treatment-Seekers with Alcohol Use Disorder: Construct Validity and Measurement Invariance. *Alcoholism: Clinical and Experimental Research*, *44*(3), 679–688. https://doi.org/10.1111/acer.14283

Votaw, V. R., Stein, E. R., & Witkiewitz, K. (2021). A Longitudinal Mediation Model of Negative Emotionality, Coping Motives and Drinking Intensity Among Individuals Receiving Community Treatment for Alcohol Use Disorder. *Alcohol and Alcoholism*, *56*(5), 573–580. https://doi.org/10.1093/alcalc/agab012

Watts, A. L., Boness, C. L., Loeffelman, J. E., Steinley, D., & Sher, K. J. (2021). Does Crude Measurement Contribute to Observed Unidimensionality of Psychological Constructs? A Demonstration With DSM–5 Alcohol Use Disorder. *Journal of Abnormal Psychology*, *130*(5), 512–524. https://doi.org/10.1037/abn0000678.supp

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(11 Suppl 3), S54-64. https://doi.org/10.1212/WNL.0b013e3182872ded

White, A. M., Castle, I.-J. P., Powell, P. A., Hingson, R. W., & Koob, G. F. (2022). Alcohol-

Related Deaths During the COVID-19 Pandemic. *Jama*, 1–3. https://doi.org/10.1001/jama.2022.4308

Wiers, R. W., van Woerden, N., Smulders, F. T. Y., & de Jong, P. J. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology*, *111*(4), 648–658. https://doi.org/10.1037/0021-843X.111.4.648

Wilcox, C. E., Dekonenko, C. J., Mayer, A. R., Bogenschutz, M. P., & Turner, J. A. (2014). Cognitive control in alcohol use disorder: deficits and clinical relevance. *Review Neuroscience*, *25*(1), 1–24. https://doi.org/10.1515/revneuro-2013-0054.Cognitive

Witkiewitz, K., Dearing, R. L., & Maisto, S. A. (2014). Alcohol Use Trajectories Among Non – Treatment-Seeking Heavy Drinkers. *Journal of Studies on Alcohol and Drugs*, *75*(3), 415–422.

Witkiewitz, K., Roos, C. R., Mann, K., & Kranzler, H. R. (2019). Advancing Precision Medicine for Alcohol Use Disorder: Replication and Extension of Reward Drinking as a Predictor of Naltrexone Response. *Alcoholism: Clinical and Experimental Research*, *43*(11), 2395–2405. https://doi.org/10.1111/acer.14183

Witkiewitz, K., Stein, E. R., Votaw, V. R., Hallgren, K. A., Gibson, B. C., Boness, C. L., Pearson, M. R., & Maisto, S. A. (*under review*). Constructs Derived from the Addiction Cycle Predict Alcohol Use Disorder Treatment Outcomes and Recovery Three Years Following Treatment.

Yang, M., Tsai, S. J., & Li, C. S. R. (2020). Concurrent amygdalar and ventromedial prefrontal cortical responses during emotion processing: a meta-analysis of the effects of valence of emotion and passive exposure versus active regulation. *Brain Structure and Function*, *225*(1), 345–363. https://doi.org/10.1007/s00429-019-02007-3

Yearby, R. (2018). Racial Disparities in Health Status and Access to Healthcare: The Continuation of Inequality in the United States Due to Structural Racism. *American Journal of Economics and Sociology*, *77*(3–4), 1113–1152. https://doi.org/10.1111/ajes.12230

Zhang, R., Geng, X., & Lee, T. M. C. (2017). Large-scale functional neural network correlates of response inhibition: an fMRI meta-analysis. *Brain Structure and Function*, *222*(9), 3973–3990. https://doi.org/10.1007/s00429-017-1443-x

Zilverstand, A., Huang, A. S., Alia-Klein, N., & Goldstein, R. Z. (2018). Neuroimaging Impaired Response Inhibition and Salience Attribution in Human Drug Addiction : A Systematic Review. *Neuron*, *98*(5), 886–903. https://doi.org/10.1016/j.neuron.2018.03.048

Supplementary Table 1. Clinical and ANA variables for the full sample, by study, and over time.

	Total Sample (N=245)		ABQ Treat (n=56)		ABQ DrinQ (n=189)	
	baseline (n=245)	3-month (n= 201)	baseline (n=56)	3-month (n= 53)	baseline (n=189)	3-month (n= 148)
Clinical Measures						
Percent heavy drinking days (PHDD)	32.6% (29.5)	24% (24.6)	45.8% (34.1)	30.1% (29.4)	28.7% (26.9)	21.7% (22.3)
Drinks per drinking day (DDD)	6.2 (3.7)	5.7 (3.6)	6.6 (3.1)	5.94 (4.0)	6.0 (3.8)	5.7 (3.5)
ANA Factor Scores						
Negative Emotionality Factor Score	0.00 (10.3)	-3.29 (9.4)	3.5 (7.9)	-1.6 (8.0)	-1.0 (10.7)	-3.8 (9.7)
Incentive Salience Factor Score	0.00 (0.8)	-0.23 (0.7)	0.4 (0.7)	0.1 (0.6)	-0.1 (0.8)	-0.3 (0.7)
Executive Function Factor Scores	0.00 (0.2)	-0.05 (0.2)	-0.1 (0.2)	-0.1 (0.2)	0.0 (0.2)	0.0 (0.2)
ANA Indicator Measures						
Sadness	52.63 (11.9)	50.01 (12.0)	57.6 (9.5)	52.82 (13.2)	51.3 (12.1)	49.5 (11.7)
Anger Affect	52.82 (10.0)	50.07 (10.6)	55.7 (9.1)	50.6 (9.3)	52.0 (10.2)	50.0 (10.9)
Fear Affect	55.83 (10.7)	51.88 (10.7)	59.7 (8.5)	53.2 (10.2)	54.8 (11.0)	51.6 (10.9)
Loneliness	57.57 (10.0)	54.13 (10.1)	60.5 (8.2)	55.8 (10.6)	56.8 (10.3)	53.8 (10.0)
Anger Hostility	53.24 (10.1)	51 (10.5)	53.7 (9.5)	51.4 (9.6)	53.1 (10.3)	50.9 (10.7)
Positive Affect	46.31 (9.5)	46.89 (8.6)	42.3 (8.3)	44.1 (8.9)	47.4 (9.6)	47.5 (8.5)
Perceived Stress	54.16 (10.1)	50.51 (10.1)	56.1 (8.2)	50.4 (8.2)	53.6 (10.6)	50.5 (10.5)
ICS Failed Control	15.01 (8.8)	13.22 (7.3)	-	-	15.01 (8.8)	13.22 (7.3)
Penn Alcohol Craving Scale	10.79 (6.3)	9.14 (6.2)	14.9 (6.0)	11.9 (6.9)	9.6 (5.9)	8.2 (5.7)
Delay Discounting	-4.44 (1.6)	-4.59 (1.5)	-4.5 (1.6)	-4.6 (1.5)	-4.4 (1.6)	-4.6 (1.5)
Flanker	86.56 (13.0)	89.67 (23.4)	91.1 (13.1)	98.8 (15.3)	85.3 (12.7)	88.0 (12.4)
List sorting working memory	99.71 (13.1)	103.26 (14.0)	103.6 (13.5)	106.7 (15.3)	99.1 (13.0)	102.6 (13.7)
Dimensional change card sort	99.75 (16.8)	100.66 (16.2)	102.1 (16.5)	109.2 (18.1)	98.7 (16.7)	99.1 (15.4)

SIP Item 5: I have taken foolish risks when I have been drinking.	0.84 (0.7)	0.68 (0.7)	0.9 (0.7)	0.6 (0.8)	0.8 (0.6)	0.7 (0.7)
SIP Item 6: When drinking, I have done impulsive things that I regretted later.	0.82 (0.6)	0.62 (0.6)	1.0 (0.7)	0.6 (0.7)	0.7 (0.6)	0.6 (0.6)