HHS Public Access

Author manuscript

*Biol Psychiatry Cogn Neurosci Neuroimaging*. Author manuscript; available in PMC 2026 July 01.

# Divergent Pathways Taken in Adolescence Predict Embracing or Resisting Moderate to Heavy Drinking in Young Adulthood

**Camila González, Ph.D.**[a], **Tam T. Nguyen-Louie, Ph.D.**[b], **Edith V. Sullivan, Ph.D.**[a], **Adolf Pfefferbaum, M.D.**[a,c], **James Klo, B.Sc.**[c], **Joseph Dehoney, M.Sc.**[e], **Qingyu Zhao, Ph.D.**[d], **Ehsan Adeli, Ph.D.**[a,e], **Susan F. Tapert, Ph.D.**[b], **Kilian M. Pohl, Ph.D.**[a,f,*]

[a]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

[b]Department of Psychiatry, University of California, San Diego, San Diego, CA, USA

[c]Center for Health Sciences, SRI International, Menlo Park, CA, USA

[d]Weill Medical College, Cornell University, New York, NY, USA

[e]Department of Computer Science, Stanford University, Stanford, CA, USA

[f]Department of Electrical Engineering, Stanford University, Stanford, CA, USA

## Abstract

**Background:** Heavy alcohol drinking during adolescence is a major public health concern and a primary risk factor for developing alcohol use disorder (AUD) in adulthood. Identifying robust, modifiable predictors that distinguish adolescents who initiate heavy drinking from those who maintain low consumption can guide targeted prevention strategies.

**Methods:** We analyzed longitudinal data from 285 participants (144 male, 141 female) of the National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA) study. All participants were no-to-low drinkers at age 15 years, remained in the study within a year of turning 21, and were followed annually, making this neuroimaging study the largest of its kind. Each visit consisted of 240 measurements (spanning cognitive, environmental, neuroimaging, and psychosocial domains), which were used to train a deep learning model to predict drinking levels and learn an embedding space capturing key risk factors. Individual developmental profiles were then clustered into distinct pathways across ages 15 to 21.

**Results:** Six unique drinking pathways were identified. Two pathways leading to heavy drinking were characterized by early exposure to peer drinking, positive alcohol expectancies, and higher sensation-seeking, predominantly among males. Three moderate-drinking pathways showed gradual increases in consumption while maintaining lower peer drinking exposure. One

[*]**Corresponding author:** Kilian M. Pohl, Ph.D., Stanford University School Medicine, Department of Psychiatry and Behavioral Sciences, 401 Quarry Road, Stanford, CA 94305-5717, kpohl@stanford.edu.

**Disclosures**: The authors report no biomedical financial interests or potential conflicts of interest.

pathway maintained no-to-low drinking, marked by negative expectancies toward alcohol and low social reinforcement.

**Conclusions:** Distinct developmental pathways highlight socially driven motivators as key modifiable risk factors underlying adolescent drinking. Targeting peer influence and alcohol-related expectancies could help divert youth from trajectories leading to heavy drinking and reduce future burden of AUD.

### Keywords

adolescent drinking; alcohol use disorder; deep learning; longitudinal clustering; drinking pathways; risk and resilience factors

## Introduction

Heavy drinking in adolescence is a major public health concern and a risk factor for developing alcohol use disorder (AUD) in adulthood with its attendant long-term health and socioeconomic consequences (1). Drinking at this early age is shaped by a complex interplay of multiple categories of factors. For example, demographics are linked to drinking patterns, with males two to three times more likely than females to develop AUD (2,3), although recently this gap has narrowed (4). Heightened probability of heavy alcohol use is also linked to personality traits, notably, extroversion and disinhibition (5,6). Expectations of positive effects from drinking alcohol, such as tension reduction and social facilitation, are associated with greater drinking in youth (2,7). Cognitive performance indicators, including low academic achievement at age 12 years, increase AUD risk in adulthood (5). Cross-sectional studies have further identified environmental factors – e.g., family dynamics (2,8), home alcohol access (9), and peer relationships (10) – as predictors of subsequent drinking. Mental health problems (11), including anxiety and depression and the use of non-alcohol psychoactive substances (12), can also portend heavy drinking. Similarly, neural underpinnings that indicate early problematic drinking include thinner cortices, being linked to higher alcohol consumption and binge drinking (4,13), smaller brain volume, and poorer cognitive function, which precede initiating alcohol use in adolescents with a family history of substance use disorder (14–16).

Longitudinal studies further show that early drinking trajectories often persist into young adulthood, with heavier adolescent drinking predicting higher risk for later alcohol misuse (17,18). Studies therefore have focused on drinking trajectories during adolescence, which are typically grouped into broad categories such as stable low drinkers, escalating drinkers, and early heavy users, using growth mixture modeling or latent class approaches (19–21). Although the exact groupings differ across publications, heterogeneous developmental patterns of alcohol use are consistently reported and involve subgroups differing in onset timing and escalation rate of drinking (22,23).

Despite the array of risk factors empirically identified as associated with heavy alcohol consumption in youth and the heterogeneity in adolescent drinking trajectories, it remains unclear which are primary drivers of drinking behavior and whether they can predict engaging in or refraining from hazardous drinking on an individual level. One likely reason

is that heterogeneous constellations of factors are antecedents of multiple diverse pathways for engaging in or refraining from heavy drinking. For instance, individuals who attend 4-year residential colleges were shown to be more likely to initiate heavy drinking earlier than those taking different post-high school paths (24) and that demographic and social factors, including parental education and peer environments, can influence the drinking trajectory (25). Further, proactive family monitoring during adolescence was linked to less future alcohol misuse (26). While previous studies explored bivariate relations between isolated factors and alcohol use, our aim was to identify which factor constellations predicted specific drinking patterns for individuals from adolescence to early adulthood.

Accordingly, we hypothesized that separate pathways towards specific drinking behaviors and levels are each driven by unique sets of a few predictors. To test this hypothesis, this study designed a novel deep learning approach that reduced 240 potential predictors of alcohol drinking to many fewer significant factors aggregating individuals into drinking pathways that emerged between ages 15 and 21 years. Measurements in the dimensions of demography, personality, cognition, mental health, substance use, environment, and brain imaging were acquired annually in 285 participants of the National Consortium on Alcohol and Neurodevelopment in Adolescence (27) (NCANDA) study. Critically, all participants chosen for this analysis were no-to-low drinkers at age 15 years and continued participation in the study within a year of their 21st birthday, making this investigation the largest annual follow-up MRI study for this age range.

To build drinking pathways, we required a representational space that distilled the large number of measurements into a compact but informative form. Such a space needed to retain the factors most relevant for distinguishing drinking behavior, while being low-dimensional enough to enable meaningful distance calculations between individuals. Towards this goal, a longitudinal Deep Neural Network (28) was first trained to predict the number of monthly drinks for each yearly visit. The encoding of each yearly visit was then used to map each visit to a point in this space. Linking successive visits yielded individualized drinking trajectories. Applying time-series k-means clustering (29) to these trajectories revealed characteristic drinking pathways driven by the longitudinal progression of predictive factors.

To identify the factors driving these pathways, we quantified the influence of each original measurement on the representational space by permuting individual inputs and assessing their effect on the model. This analysis revealed that the predictors most relevant to shaping trajectories were not uniform across the cohort but instead varied by pathway, with distinct sets of factors statistically driving the divergence between developmental patterns. In this way, we could move beyond identifying generic risk factors and instead uncover pathway-specific drivers of resilience and vulnerability in adolescent drinking behavior. Importantly, many of the elements underlying the heavy drinking pathways are modifiable and can be addressed with targeted public health strategies. Our trajectory-based approach, which clusters individuals within a representational space learned via deep learning, not only pinpoints actionable prevention targets for reducing heavy drinking initiation but also offers a scalable blueprint for mapping risk and resilience in other developmental health contexts.

## Methods and Materials

### Study Design and Participants

The NCANDA (27) study enrolled 831 youth recruited between 2013 and 2015 across five sites: Oregon Health & Science University (OHSU), SRI International, University of California, San Diego (UCSD), University of Pittsburgh Medical Center (UMPC), and Duke University Medical Center (DUMC). The cohort is representative of sex and racial/ethnic groups in the United States. Over half of the participants had at least one risk factor associated with heavy drinking at recruitment, such as a family history of AUD.

Data collected yearly included substance use and mental health interviews, questionnaires covering alcohol expectancies, personality, pubertal development, and peer and family relationships; and structural, functional, and diffusion magnetic resonance imaging (MRI). Participants under 18 years old provided assent, and adult participants and legal guardians provided written informed consent for participation in the study after receiving a complete description of the procedures, risks, and benefits. Institutional Review Boards from each acquisition site approved the study protocols.

The sample used in this study comprised visits from the Year 9 Data Release of NCANDA made publicly available via NIH NDAR (collection 4513). From the initial NCANDA cohort ($N=831$), we excluded 23 participants with baseline MRI anomalies (14) (resulting in $N=808$) and 24 who changed acquisition sites ($N=784$). Demographic characteristics (sex and age) of the analyzed sample were comparable to those reported for the full NCANDA cohort. We limited our sample to visits where participants were between 15 and 21 years of age (selected as this age window maximized the number of longitudinal visits available among participants meeting our inclusion criteria, see Supplemental Figure 1), and with data recorded at age 15 years and within a year of their 21$^{st}$ birthday ($N=310$) to ensure sufficient longitudinal coverage for meaningful trajectory characterization. We further excluded subjects who exceeded no-to-low drinking (30) criteria at age 15 years, defined using the adjusted Cahalan score (30) as adolescents already engaging in moderate or heavy drinking at baseline would conflate the alcohol initiation with escalation processes and thus reduce our ability to isolate early predictors of divergence. Our final sample comprised $N=285$ study participants (Table 1).

### Measurements and regression target

The selection of most of the 68 non-imaging measurements (see Supplemental Table 1) was based on an earlier NCANDA study predicting the onset of alcohol drinking (9). The measurements involved five dimensions: *demographics* (sex, age and pubertal development (31)); *personality* assessed through, the Ten Item Personality Inventory (32) (TIPI), Urgency-Premeditation-Perseverance-Sensation Seeking (UPPS) Impulsive Behavior Scale (33,34) and Alcohol Expectancy Questionnaire (35) (AEQ), academic performance, career intentions, sleep health (36); *cognition* evaluated using the National Institutes of Health (NIH) cognitive battery (37); and *environmental factors* capturing perceived ease of access to alcohol (38,39), youth-parent relationships (e.g. parents' efforts to monitor their children), peer alcohol use behaviors, and the number of friends or romantic partners.

The current analysis expanded the earlier study with seven *mental health* subscales from the Youth Self-Report for participants under 18 years and Adult Self-Report otherwise to evaluate a range of psychological symptoms and behavioral problems in individuals (40), covering symptoms of anxiety, depression, somatic complaints, social withdrawal, thought problems, attention problems, and aggressive behavior. The analysis also included a non-alcohol *substance use* category, which included tobacco and cannabis use as both share common risk factors and co-occurrences with alcohol (12). Excluded from this category were the yearly reported number of drinks consumed in the last month, which was predicted by the deep learning model.

Imaging measurements included structural MRI data of 68 cortical and 25 subcortical measures derived from the Desikan-Killiany atlas in FreeSurfer. Regional lateral scores were averaged across both hemispheres. Diffusion tensor imaging (DTI) yielded fractional anisotropy (FA) and mean diffusivity (MD) for 28 bilateral brain regions. Functional MRI measurements include 23 resting-state functional connectivity efficiency scores (41). All scores were generated using the SIBIS pipeline (15,16,41).

To impute missing predictor values, we used three different strategies: if no measurements were available for a subject, we imputed the dataset mean; if the value at age 15 years was missing, we copied the value from the closest visit with a recorded measurement; and for missing values at later visits, we performed linear interpolation using available values. Note, this imputation was only performed for predictors as visits with missing drinking outcomes were ignored by the analysis.

To minimize confounding effects, a generalized linear model (GLM) residualized the data of the entire sample with respect to site across all measurements, scanner model for imaging data, and brain volume for structural measurements. Measurements were normalized in the −1 to 1 range. The log-scale of the number of drinks consumed in the last month was normalized in the 0 to 1 range and used as observation.

## Longitudinal Deep Learning Model

The deep learning analysis (Figure 1) consisted of first identifying the drinking pathways taken by participants in our cohort and then uncovering the predictors that drove each pathway. To identify pathways, a longitudinal deep learning model called Long Short-Term Memory (LSTM) network (28) was trained to predict the number of drinks consumed in the last month for each yearly visit (Figure 1, Step 1.A.). Importantly, this architecture utilizes previous but not subsequent time points to predict the output for each visit. Each participant was encoded by a time sequence of length seven (for ages 15 to 21 years), where each time point consisted of the 240 input measurements. The model (28,42) then predicted the monthly number of drinks at a visit based on input measurements of past and present but no future visits.

Nested, stratified 5-fold cross-validation was used to select the best model architecture. Stratifying participants across the 5 folds was done based on the highest adjusted Cahalan score (30) recoded of the subject during the observed 7-year range to ensure each fold was representative of the sample with regard to drinking levels. In the nested cross-validation

framework, the outer loop consisted of one fold held out as the testing set for evaluating model performance. Within the remaining four folds, the inner loop used three folds for training the model with different hyperparameters and one fold for validation to determine the optimal model (i.e., hyperparameter setting) for testing. Specifically, the Long Short-Term Memory network was trained on the three training folds using the Adam (43) optimizer and binary cross-entropy loss in PyTorch (44) based on different hyperparameter settings: number of Long Short-Term Memory layers (i.e., 1, 2, 3, or 4), number of units per layer (i.e., 8, 16, 32, or 64), starting learning rate (i.e., 1e-3 and 1e-4), augmentation scheme (masking input measurements, oversampling underrepresented classes(45), adding Gaussian noise to the inputs or outputs), and number of epochs (from 20 to 200 epochs, tested in 20 epoch intervals). For each configuration setting, the average mean absolute error (MAE) between the predicted and actual normalized monthly alcohol consumption was recorded on the validation fold. The model with the lowest MAE score (a two-layer LSTM with 32 hidden units and a regression head) was then applied to the test fold.

## 2D Maps

The optimal model architecture identified during cross-validation was also used to project the visits of each subject onto a 32-dimensional representational map so that the drinking behavior for visits in close proximity was predicted by similar measurements. This space was turned into a 2D map via t-SNE dimensionality reduction (46) for better visualization and for more robust clustering as distance calculation is more stable for lower-dimensional embeddings (47). Visits were now dots in this 2D and those adjacent visits from the same individual were connected via a line to define the *trajectory* for each study participants (Step 1.B. in Figure 1). The final component of Step 1 (Step 1.C. in Figure 1) grouped the trajectories of all participants into potential *drinking pathways* via *time-series k-means clustering* (29). The number of potential pathways was set empirically based on the *KneeLocator* (48) algorithm (Supplemental Figure 2).

Robust pathways were those clusters with at least 20 participants that were identified across 100 different initializations of the deep learning approach. The 100 trained models had the identical architecture (i.e., hyperparameter settings) but different random seed initialization. Despite some variability between the maps resulting from different initializations, the overall structure of the pathways remains consistent as reflected in Supplemental Figure 3. The figure shows the same six clusters plotted on the 2D maps of five different exemplary initializations with the blue pathway, for example, starting at a distinct initial point across all maps.

## Identification of Pathway Drivers

In Step 2 (Figure 1), the analysis identified the most predictive of the 240 initial measurements via permutation testing. For each measurement and each model initialization, values were randomly permuted 100 times across subjects and visits, thereby perturbing the longitudinal trajectories at all time points and evaluating the contribution of each predictor to the model's ability to capture drinking behavior across the developmental period. For each permutation configuration, the prediction error between predicted and actual number of drinks was recomputed and compared to the error obtained using the original (non-

permuted) data for the corresponding model initialization. Statistical significance was then evaluated for each measurement across permutations and model initializations, and the resulting p-values were corrected with the False Discovery Rate (FDR) to account for multiple comparisons across all 240 measurements. Measurements with p-values below 0.05 were viewed as significant for the prediction process. Significance of predictors (without FDR correction) were confirmed by computing the average accuracy across the 100 initializations based on the original and each of the permuted settings.

## Results

Based on nested cross-validation, the longitudinal deep learning model achieved the following accuracy in predicting monthly alcohol consumption from 240 behavioral and other variables: a mean absolute error (MAE) of 0.123 ± 0.006, root mean squared error (RMSE) of 0.173 ± 0.009, and a coefficient of determination $R^2 = 0.436$ on held-out test fold (see Supplemental Figure 4).

Based on those 240 variables, the method also mapped each visit of a subject into a 2D map shown as a dot in Figure 2.A. The shade of the dot indicates the number of drinks consumed in the past month, i.e., the darker the shade the more the person drank. Supplemental Figure 5 colors the dots according to age, sex, and acquisition site revealing that while age and sex contribute to the placement on the maps, acquisition site did not. All visits (i.e., dots) of a participant are connected by a line to form an individual trajectory starting at age 15 years (left side of the map) as no-to-low drinkers before diverging into different drinking trajectories by age 20–21 years.

From those trajectories, the data-driven analysis identified six separate pathways (independent of drinking levels) based on reproducible cluster assignments across the 100 independently initialized deep learning models. Figure 2.A visualizes the pathway trend lines using color coding to highlight the progression pattern for each group into the three drinking levels according to Figure 2.B: no-to-low (green), moderate (turquoise, blue and purple), and heavy (orange and red). Each dot on a pathway represents a year (e.g., first dot is age 15 years) and the naming of pathways was done post-hoc based on the significant predictors identified via permutation testing (Figures 3 and 4).

The heavy drinking level included two pathways: an *extroverted heavy drinker* cluster (N=52) consuming an average of 36.7 monthly drinks by age 21 years and an *introverted heavy drinker* cluster (N=20) consuming an average of 21.2 monthly drinks peaking about 20 years. The moderate drinking level was characterized by steadily increasing drinking with age to approximately 14 drinks per month by age 21 years, comprised three pathways named for their salient features: parentally monitored (N=61), abstemious peers (N=25), and cannabis users (N=23). A *stable no-to-low drinker* (N=88) group consumed 6.07 drinks per month by age 21 years.

Permutation testing (49) identified nine measurements (Figure 3) that were significant ($p < 0.05$ after FDR correction) for predicting monthly alcohol consumption for one or several of the six pathways: cannabis use, expectancy of alcohol improving social behavior, parental

monitoring, extraversion, alcohol consumption of peers, number of intoxicated peers, sex, sensation seeking and alcohol relaxation expectancy. For each of the nine significant predictors, their linear trajectory was summarized by the normalized means at age 15 years (Figure 4 left) and slope (representing the rate of change between ages 15 and 21 years, Figure 4 right) and plotted for the six drinking pathways (x-axis). While the normalized mean encodes baseline differences prior to divergence of drinking trajectories, the slope captures longitudinal progression of predictors across adolescence.

Significant predictors of the *extroverted heavy drinker* pathway were high extraversion coupled with sensation seeking at age 15 years and peers drinking increasingly more alcohol in the subsequent years. This group also expressed belief that alcohol would help them relax and improve social interactions (AEQ Social Behavior Changes and Relaxation/ Tension Reduction scores) and exhibited a notable rise in cannabis use over the test interval. By contrast, the significant predictor of the *introverted heavy drinker* pathway was an increase between 15 and 21 years in expectations that alcohol would enhance social interactions. Sex emerged as a significant differentiating factor between pathways, with males overrepresented in both heavy drinking groups (nearly 70%).

The three moderate drinker pathways showed similar levels of alcohol consumption amongst each other (Figure 2.B). Distinguishing predictors between them were a high degree of parental monitoring at age 15 years, an increase in cannabis consumption until age 21 years similar to the *extroverted heavy drinkers,* and social circles abstaining from alcohol use at age 15 years.

Socializing with teens abstaining from alcohol use was also a predictor of *stable no-to-low drinkers,* as were low expectations towards the outcome from alcohol drinking and abstaining from cannabis use. A relative high ratio of females was a predictor shared with the other two moderate substance use pathways (i.e., parental monitoring and abstaining peers). Supplemental Figure 6 visualizes the distribution of measurement values for each of the six pathways.

## Discussion

Based on a data-driven analysis of 240 measurements recorded annually from 285 NCANDA participants between the ages of 15 to 21 years, peer drinking behavior and expectations of alcohol consumption emerged as key modifiable factors predicting the transition toward moderate to heavy alcohol use, whereas others predicted maintenance of low consumption. This holistic search across 7 dimensions (i.e., demographics, personality, cognition, mental health, substance use, environment, and brain imaging) was made possible by a novel deep learning algorithm reducing high-dimensional, longitudinal data into visit-level predictors of monthly drinking. This prediction model mapped yearly visits of each participant onto a 2D map so that the comprehensive information captured of each person was reduced to a simple trajectory. Clustering these trajectories revealed six dissociable pathways (Figure 2.A) built on specific features that predicted three levels of alcohol use (stable low drinkers, escalating users, and early heavy; Figure 2.B) that are consistent with prior developmental research (19). Unlike previous work, our analysis was not confined

to identifying trajectories of specific shapes (e.g., linear or spline) for a few selected measurements but instead relied on an entirely data-driven approach to derive distinct pathways from a comprehensive battery of annually recorded measures. The remainder of this discussion will focus on how the predictors of these pathways relate to the existing alcohol literature.

Two pathways followed mostly male youth who transitioned from no-to-low drinking at 15 years to an average of more than 20 drinks per month by age 21 years. Their drinking behavior was largely predicted by positive expectations toward alcohol consumption. While this has been reported on a group-level (7,10), the predictions herein were done on an individual level and for each visit, a critical step towards personalizing prevention program. At age 15 years, i.e., before moderate-to-heavy drinking was initiated, the *extroverted heavy drinker* group already held more positive expectations about the effects of alcohol than any other group, believing that alcohol would promote relaxation and enhance social interactions. While over the next seven years, positive expectations increased across all pathways, the rise was fastest in the two high-drinking groups. Importantly, expectations involving alcohol use outcomes are highly modifiable factors that can be used in targeted prevention programs such as distributing information on the adverse effects of alcohol use (50).

Both heavy drinking pathways were predominantly male. This finding aligns with epidemiological data indicating higher rates of heavy alcohol use among adolescent males (2,3). Sex may interact with peer norms and sensation-seeking tendencies, particularly when adolescents are embedded in peer environments that normalize alcohol consumption (13). Prevention strategies targeting adolescent drinking may benefit from considering sex-specific risk profiles, particularly the interaction between peer influence, alcohol expectancies, and sensation-seeking behavior.

As expected (7), personality traits, particularly extraversion, remained largely stable throughout the 7 years and marked the *extroverted heavy drinker* pathway. Nonetheless, factors (such as high extroversion) taken in isolation are not necessarily predicting heavy drinking, exemplified herein where extroversion at age 15 years was predictive of a heavy as well as a moderate drinking pathway (i.e., *moderate drinkers among abstemious peers*). Instead, characteristic of heavy drinking pathways were unique constellations of predictors that included associating with heavier-drinking peers even before participants had initiated drinking (at age 15 years), positive expectations surrounding alcohol drinking, and sensation-seeking tendencies. Note that sensation seeking might be part of the constellation predicting heavy substance use, as high levels were also reported for the cannabis using pathways.

That extroversion did not necessarily lead to heavy drinking also points to a constellation of protective predictors. Protective constellation could include avoiding social circles with heavy drinkers, which was a predictor shared between the pathways of the *no-low-drinking* and *moderate drinkers among abstemious peers*. Interestingly, the unique combination of risk and protective factors describing this moderate drinker pathway was captured in the 2D map (Figure 2) by having this group start at a different location from any other pathway.

A factor that was neither a risk nor protective was *parental monitoring*, which measures the extent parents inquire about their children's activities and emerged as a characteristic for the pathway of the *parentally-monitored moderate drinkers*. Accordingly, neither its value at 15 years nor its slope showed a consistent increase or decrease in relation to higher levels of drinking (Figure 4). This inconsistency aligns with publications suggesting that while increased parental control can deter alcohol use, active monitoring may alternatively arise in response to perceived drinking risks already evident in the youth's behavior (9). Finally, none of the identified pathways was significantly influenced by brain imaging measurements, cognitive test performance, or mental health scores. This discovery challenges the emphasis on neural status predictors previously (13) and reinforces the critical role of behaviorally modifiable risk factors in efforts to avert hazardous drinking.

Our analysis has several limitations. First, it was restricted to 285 participants, resulting in half of the pathways comprising 25 or fewer individuals, so that the statistical power to identify significant predictors was low. A larger sample might reveal additional factors as influencing drinking behavior, particularly those with small effect sizes, such as brain imaging measures. Furthermore, key variables were self-reported, including alcohol consumption, personality traits, cannabis use, and peer drinking behaviors. Self-reported data are subject to biases, including recall bias and social desirability bias (51), which may lead to underreporting or overreporting of alcohol and substance use. In addition, missing data were imputed using linear interpolation, which assumes smooth temporal change and may not fully capture abrupt behavioral shifts. While we included a comprehensive set of 240 predictors based on previous research, certain factors influencing drinking behavior may have eluded our dataset. For example, this study did not account for changes in socio-economic status over the examination interval or environmental influences such as alcohol outlet density (52). Finally, although NCANDA was designed to approximate U.S. demographic distributions, it is intentionally enriched for youth at elevated risk for alcohol use disorder (e.g., family history of AUD). This enrichment may increase sensitivity for detecting divergent developmental pathways but may limit generalizability to lower-risk community samples. To replicate our findings on more representative samples, researchers can download our publicly available and cohort-agnostic code from https://github.com/camgbus/DivergentPathways.

In conclusion, our data-driven analysis tracked drinking trajectories of individuals and revealed distinct pathways driven by key predictors identified in adolescence that distinguished stable no-to-low drinkers from those transitioning to moderate or heavy drinking in early adulthood. From an initial set of 240 potential predictors, nine were found to have the greatest influence on drinking behavior. This set of discoveries emerged from a framework for transforming longitudinal measures into interpretable pathways on 2D maps that would be unattainable with conventional deep-learning models, which are inadequately suited for capturing nonlinear multi-year trajectories defining human developmental processes. Seven identified factors are modifiable shaped by environment, peer behaviors, and alcohol-related expectations. As similar drinking habits can result from widely different motivators, the heterogeneity of risk profiles should be recognized when designing prevention strategies, emphasizing personality traits in combination with socially

driven motivators. Such public health strategies could enhance efforts to mitigate heavy drinking in adolescence and avert the development of AUD in adulthood.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements:

This work was supported by the U.S. National Institute on Drug Abuse [DA057567 (KMP+SFT)], the U.S. National Institute on Alcohol Abuse and Alcoholism [AA010723 (EVS), AA028840 (QZ), AA021692 (SFT)], the Patricia A. Judd and Family Endowed Post-Doctoral Fellowship in ADHD (TNL), and the Stanford HAI Google Cloud Credits (KMP) and Hoffman-Yee Grant (EA). The U.S. National Institute on Alcohol Abuse and Alcoholism also supported the collection of NCANDA data [AA021697 (AP+KMP), AA021695 (SFT+SAB), AA021696, AA021681, AA021690, AA021691].

## References

1. White A, Hingson R. The burden of alcohol use: Excessive alcohol consumption and related consequences among college students. Alcohol Res Curr Rev. 2014;35(2):201.
2. Moure-Rodríguez L, Piñeiro M, Corral Varela M, Rodríguez-Holguín S, Cadaveira F, Caamaño-Isorna F. Identifying predictors and prevalence of alcohol consumption among university students: Nine years of follow-up. PLoS One. 2016;11(11):e0165514. [PubMed: 27812131]
3. Zucker RA. Alcohol Use and the Alcohol Use Disorders: A Developmental-Biopsychosocial Systems Formulation Covering the Life Course. In: Developmental Psychopathology [Internet]. John Wiley & Sons, Ltd; 2015 [cited 2024 Nov 14]. p. 620–56. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470939406.ch17 doi:10.1002/9780470939406.ch17
4. White AM. Gender Differences in the Epidemiology of Alcohol Use and Related Harms in the United States. Alcohol Res Curr Rev. 2020 Oct 29;40(2):01. doi:10.35946/arcr.v40.2.01
5. Childhood and adolescent predictors of heavy drinking and alcohol use disorders in early adulthood: a longitudinal developmental analysis - Englund - 2008 - Addiction - Wiley Online Library [Internet]. [cited 2024 Nov 13]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1360-0443.2008.02174.x
6. Moore A, Lewis B, Elton A, Squeglia LM, Nixon SJ. An investigation of multimodal predictors of adolescent alcohol initiation. Drug Alcohol Depend. 2024;265:112491. [PubMed: 39522301]
7. Schulenberg JE, Maggs JL. A developmental perspective on alcohol use and heavy drinking during adolescence and the transition to young adulthood. J Stud Alcohol Suppl. 2002 Mar;(s14):54–70. doi:10.15288/jsas.2002.s14.54 [PubMed: 12022730]
8. Grant BF. The Impact of a Family History of Alcoholism on the Relationship Between Age at Onset of Alcohol Use and DSM–IV Alcohol Dependence: Results From the National Longitudinal Alcohol Epidemiologic Survey. Alcohol Health Res World. 1998;22(2):144. [PubMed: 15706789]
9. Nguyen-Louie TT, Thompson WK, Sullivan EV, Pfefferbaum A, Gonzalez C, Eberson-Shumate SC, et al. Multi-dimensional predictors of first drinking initiation and regular drinking onset in adolescence: A prospective longitudinal study. Dev Cogn Neurosci. 2024 Oct;69:101424. doi:10.1016/j.dcn.2024.101424 [PubMed: 39089172]
10. Boyd SJ, Sceeles EM, Tapert SF, Brown SA, Nagel BJ. Reciprocal relations between positive alcohol expectancies and peer use on adolescent drinking: An accelerated autoregressive cross-lagged model using the NCANDA sample. Psychol Addict Behav. 2018;32(5):517. [PubMed: 29963874]
11. Patel H, Nooner KB, Reich JC, Woodley MMO, Cummins K, Brown SA. Trauma's Distinctive and Combined Effects on Subsequent Substance Use, Mental Health, and Neurocognitive Functioning with the NCANDA Sample. Dev Cogn Neurosci. 2024;101427. [PubMed: 39111118]
12. Conway KP, Green VR, Kasza KA, Silveira ML, Borek N, Kimmel HL, et al. Co-occurrence of tobacco product use, substance use, and mental health problems among adults: Findings from

Wave 1 (2013–2014) of the Population Assessment of Tobacco and Health (PATH) Study. Drug Alcohol Depend. 2017;177:104–11. [PubMed: 28582698]

13. Miller AP, Baranger DA, Paul SE, Garavan H, Mackey S, Tapert SF, et al. Neuroanatomical variability and substance use initiation in late childhood and early adolescence. JAMA Netw Open. 2024;7(12):e2452027–e2452027. [PubMed: 39786408]
14. Pfefferbaum A, Rohlfing T, Pohl KM, Lane B, Chu W, Kwon D, et al. Adolescent development of cortical and white matter structure in the NCANDA sample: role of sex, ethnicity, puberty, and alcohol drinking. Cereb Cortex. 2016;26(10):4101–21. [PubMed: 26408800]
15. Pfefferbaum A, Kwon D, Brumback T, Thompson WK, Cummins K, Tapert SF, et al. Altered brain developmental trajectories in adolescents after initiating drinking. Am J Psychiatry. 2018;175(4):370–80. [PubMed: 29084454]
16. Zhao Q, Sullivan EV, Honnorat N, Adeli E, Podhajsky S, De Bellis MD, et al. Association of heavy drinking with deviant fiber tract development in frontal brain systems in adolescents. JAMA Psychiatry. 2021;78(4):407–15. [PubMed: 33377940]
17. Jackson KM, Marceau K, Colby SM, Barnett NP, Rogers ML, Hayes KL. Trajectories of early alcohol use milestones: Interrelations among initiation and progression. Alcohol Clin Exp Res. 2021;45(11):2294–308. doi:10.1111/acer.14723 [PubMed: 34585748]
18. McKetta S, Espinoza P, Keyes K, Jager J. Maturing out or in? Demographic determinants of young adult drinking trajectories and midlife alcohol use disorder risks. Alcohol Clin Exp Res. 2026;50(2):e70226. doi:10.1111/acer.70226
19. Chassin L, Pitts SC, Prost J. Binge drinking trajectories from adolescence to emerging adulthood in a high-risk sample: Predictors and substance abuse outcomes. J Consult Clin Psychol. 2002;70(1):67–78. doi:10.1037/0022-006X.70.1.67 [PubMed: 11860058]
20. Shamblen SR, Ringwalt CL, Clark HK, Hanley SM. Alcohol Use Growth Trajectories in Young Adolescence: Pathways and Predictors. J Child Adolesc Subst Abuse. 2014 Jan 1;23(1):9–18. doi:10.1080/1067828X.2012.747906
21. Moid MZI, Bray JW, Van Hasselt M, Barbosa C. Estimating Lifetime Drinking Trajectories for Alcohol Use from Adolescence to Older Adulthood in the United States: A Three-Step Approach. MDM Policy Pract. 2026 Jan 1;11(1):23814683261421957. doi:10.1177/23814683261421957 [PubMed: 41773270]
22. Stephenson M, Barr P, Thomas N, Cooke M, Latvala A, Rose RJ, et al. Patterns and predictors of alcohol misuse trajectories from adolescence through early midlife. Dev Psychopathol. 2025 May;37(2):734–50. doi:10.1017/S0954579424000543 [PubMed: 38465371]
23. Wallace GT, Whichard C, Augustyn M, Henry KL. Heavy episodic drinking in adolescence and alcohol-related problems in adulthood: A developmental approach to alcohol use across the life course. Dev Psychopathol. 2024 Feb;36(1):349–65. doi:10.1017/S0954579422001249 [PubMed: 36503558]
24. Zhao Q, Paschali M, Dehoney J, Baker FC, de Zambotti M, De Bellis MD, et al. Identifying high school risk factors that forecast heavy drinking onset in understudied young adults. Dev Cogn Neurosci. 2024 Aug 1;68:101413. doi:10.1016/j.dcn.2024.101413 [PubMed: 38943839]
25. Berg N, Piirtola M, Marttunen M, Latvala A, Kiviruusu O. Trajectories of Concurrent Psychological Distress, Heavy Episodic Drinking and Daily Cigarette Smoking From Adolescence to Midlife: Patterns and Their Sociodemographic Correlates. Nord Stud Alcohol Drugs. 2026 Feb 1;43(1):20–44. doi:10.1177/14550725251408211
26. Bailey JA, Le VT, McMorris BJ, Merrin GJ, Heerde JA, Batmaz EA, et al. Longitudinal associations between adult-supervised drinking during adolescence and alcohol misuse from ages 25–31 years: A comparison of Australia and the United States. Addict Behav. 2024 Jun 1;153:107984. doi:10.1016/j.addbeh.2024.107984 [PubMed: 38401424]
27. Brown SA, Brumback T, Tomlinson K, Cummins K, Thompson WK, Nagel BJ, et al. The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): a multisite study of adolescent development and substance use. J Stud Alcohol Drugs. 2015;76(6):895–908. [PubMed: 26562597]
28. Graves A Long Short-Term Memory. In: Supervised Sequence Labelling with Recurrent Neural Networks [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012 [cited 2024 Dec

24]. p. 37–45. (Studies in Computational Intelligence). Available from: http://link.springer.com/10.1007/978-3-642-24797-2_4 doi:10.1007/978-3-642-24797-2_4

29. Huang X, Ye Y, Xiong L, Lau RY, Jiang N, Wang S. Time series k-means: A new k-means type smooth subspace clustering for time series data. Inf Sci. 2016;367:1–13.
30. Cahalan D, Cisin IH. Drinking behavior and drinking problems in the United States. In: Social aspects of alcoholism. Springer; 1976. p. 77–115.
31. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. J Youth Adolesc. 1988;17(2):117–33. [PubMed: 24277579]
32. Gosling SD, Rentfrow PJ, Swann WB Jr. A very brief measure of the Big-Five personality domains. J Res Personal. 2003;37(6):504–28.
33. Lynam DR, Smith GT, Whiteside SP, Cyders MA. The UPPS-P: Assessing five personality pathways to impulsive behavior. West Lafayette Purdue Univ. 2006;10.
34. Cyders MA, Smith GT, Spillane NS, Fischer S, Annus AM, Peterson C. Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. Psychol Assess. 2007;19(1):107. [PubMed: 17371126]
35. Brown SA, Christiansen BA, Goldman MS. The Alcohol Expectancy Questionnaire: an instrument for the assessment of adolescent and adult alcohol expectancies. J Stud Alcohol. 1987 Sep;48(5):483–91. doi:10.15288/jsa.1987.48.483 [PubMed: 3669677]
36. Spilsbury JC, Drotar D, Rosen CL, Redline S. The Cleveland Adolescent Sleepiness Questionnaire: A New Measure to Assess Excessive Daytime Sleepiness in Adolescents. J Clin Sleep Med. 2007 Oct 15;03(06):603–12. doi:10.5664/jcsm.26971
37. Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, et al. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. J Neurosci Methods. 2010;187(2):254–62. [PubMed: 19945485]
38. Tobler AL, Komro KA, Maldonado-Molina MM. Relationship between neighborhood context, family management practices and alcohol use among urban, multi-ethnic, young adolescents. Prev Sci. 2009;10:313–24. [PubMed: 19381808]
39. Komro KA, Maldonado-Molina MM, Tobler AL, Bonds JR, Muller KE. Effects of home access and availability of alcohol on young adolescents' alcohol use. Addiction. 2007 Oct;102(10):1597–608. doi:10.1111/j.1360-0443.2007.01941.x [PubMed: 17854336]
40. Achenbach TM, Rescorla LA. Manual for the ASEBA Adult Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2003.
41. Zhao Q, Kwon D, Müller-Oehring EM, Le Berre AP, Pfefferbaum A, Sullivan EV, et al. Longitudinally consistent estimates of intrinsic functional networks. Hum Brain Mapp. 2019;40(8):2511–28. [PubMed: 30806009]
42. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. [PubMed: 9377276]
43. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: International Conference on Learning Representations (ICLR) [Internet]. 2015. Available from: https://arxiv.org/abs/1412.6980
44. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32 [Internet]. Curran Associates, Inc.; 2019. p. 8024–35. Available from: https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf
45. Iglesias G, Talavera E, González-Prieto Á, Mozo A, Gómez-Canaval S. Data Augmentation techniques in time series domain: a survey and taxonomy. Neural Comput Appl. 2023 May 1;35(14):10123–45. doi:10.1007/s00521-023-08459-3
46. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9(11).
47. Hui A, Gao BJ. When is Nearest Neighbor Meaningful: Sequential Data. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management [Internet]. Virtual Event Queensland Australia: ACM; 2021 [cited 2024 Dec 19]. p. 3103–6. Available from: https://dl.acm.org/doi/10.1145/3459637.3482219 doi:10.1145/3459637.3482219

48. Satopaa V, Albrecht J, Irwin D, Raghavan B. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In: 2011 31st international conference on distributed computing systems workshops. IEEE; 2011. p. 166–71.

49. Ojala M, Garriga GC. Permutation tests for studying classifier performance. J Mach Learn Res. 2010;11(6).

50. Perry CL, Grant M, Ernberg G, Florenzano RU, Langdon MC, Myeni AD, et al. WHO Collaborative Study on Alcohol Education and Young People: Outcomes of a Four-Country Pilot Study. Int J Addict. 1989 Jan 1. Located at: world. doi:10.3109/10826088909048710

51. Althubaiti A Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016 May;211. doi:10.2147/JMDH.S104807

52. Martín-Turrero I, Sureda X, Escobar F, Bilal U, Berasaluce M, Valiente R. How Can We Measure Alcohol Outlet Density Around Schools? A Comparison Between Two Buffer-Based Methods. J Urban Health. 2023 Jun;100(3):627–37. doi:10.1007/s11524-023-00740-z [PubMed: 37351726]

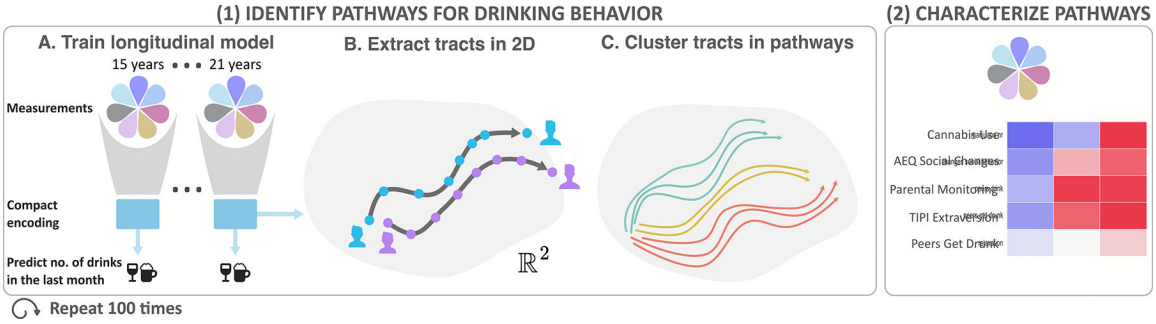


**Figure 1:**

Overview of the analysis. Step 1 identifies pathways for drinking behavior by (A) training a longitudinal deep learning model that predicted the monthly consumption of drinks containing alcohol based on 240 annually acquired measurements between the ages of 15 and 21 years, (B) extracting tracts in a 2-dimensional map, and (C) clustering tracts into drinking pathways by identifying the most robust clusters across 100 runs of this analysis. Step 2 characterized the pathways by identifying salient predictors.

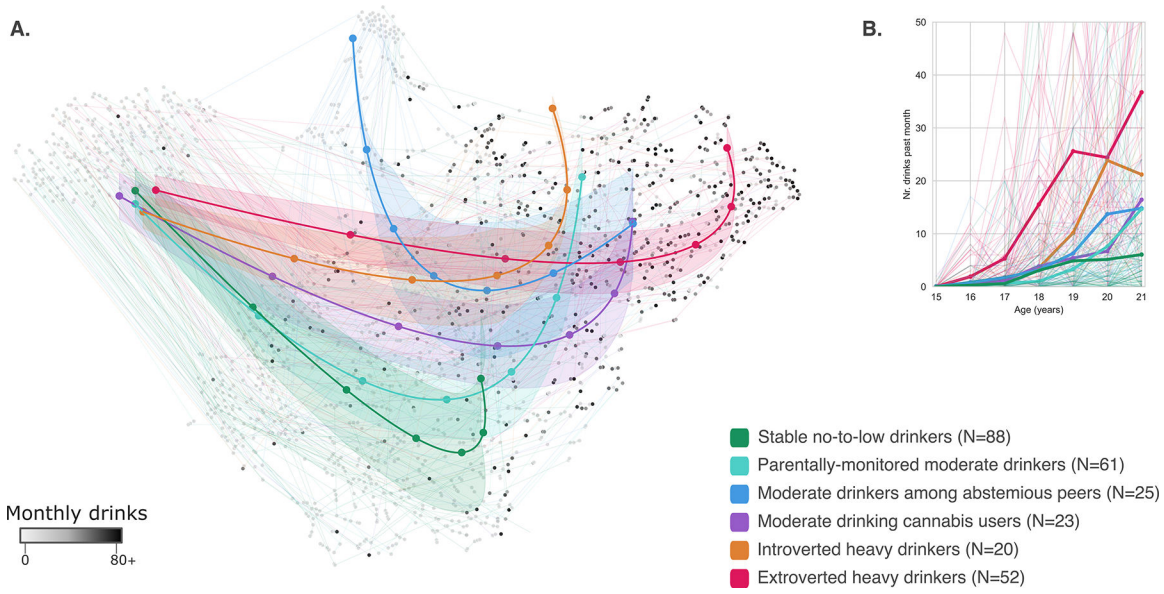


**Figure 2:**

(A) Individual visits (gray dots), trajectories of subjects (gray lines) and pathways of drinking behavior (color trend lines with each dot representing a specific age) from ages 15 to 21 years visualized in a 2D map. While five of the six pathways start at a similar section on the left of the map at 15 years, all pathways diverge so that stable no-to-low drinkers end at the bottom right and heavy drinkers at the top right of the map. (B) Per-year average monthly drinking differentiates the pathways into no-to-low, moderate, and heavy drinking groups by young adulthood. Pathway labels are descriptive and were assigned post-hoc based on the dominant predictors identified through permutation testing.

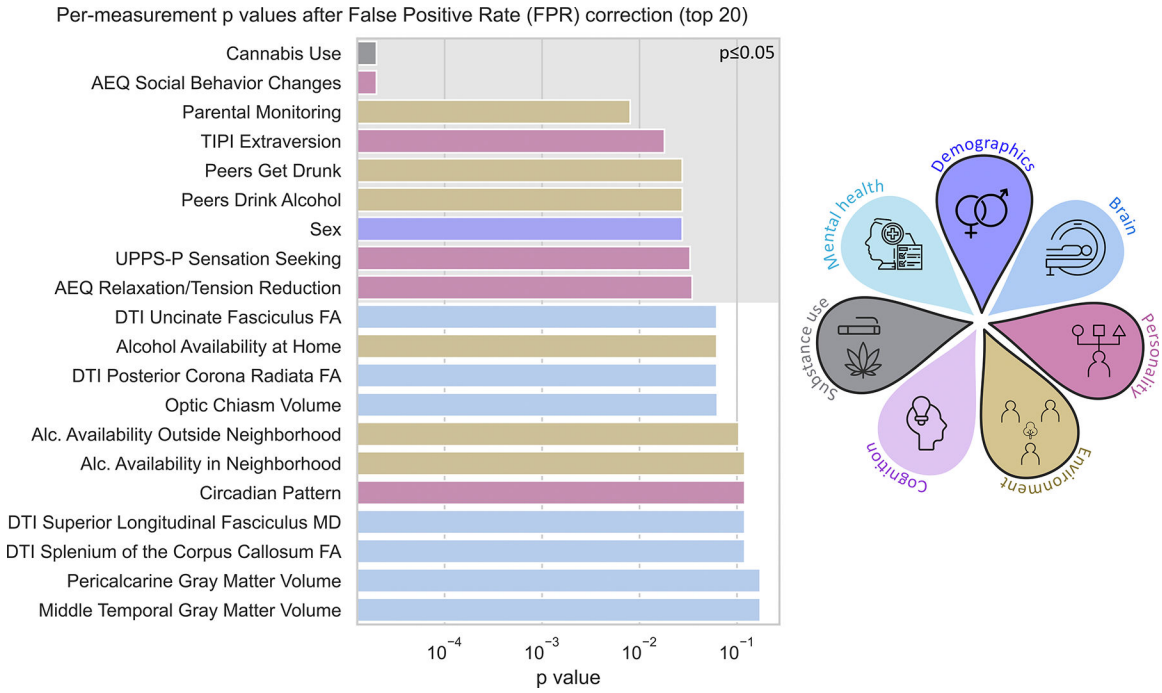


**Figure 3:**

(Left) Significant measurements (top 9 factors) characterizing developmental pathways identified via permutation testing. Measurements were evaluated for their statistical significance with a p-value threshold of 0.05 applied after False Positive Rate (FPR) correction. (Right) The figure highlights categories for which at least one measurement was significant. A total of nine measurements across four categories were relevant, with cannabis use and expectations that alcohol would improve social interactions the most predictive of alcohol use.

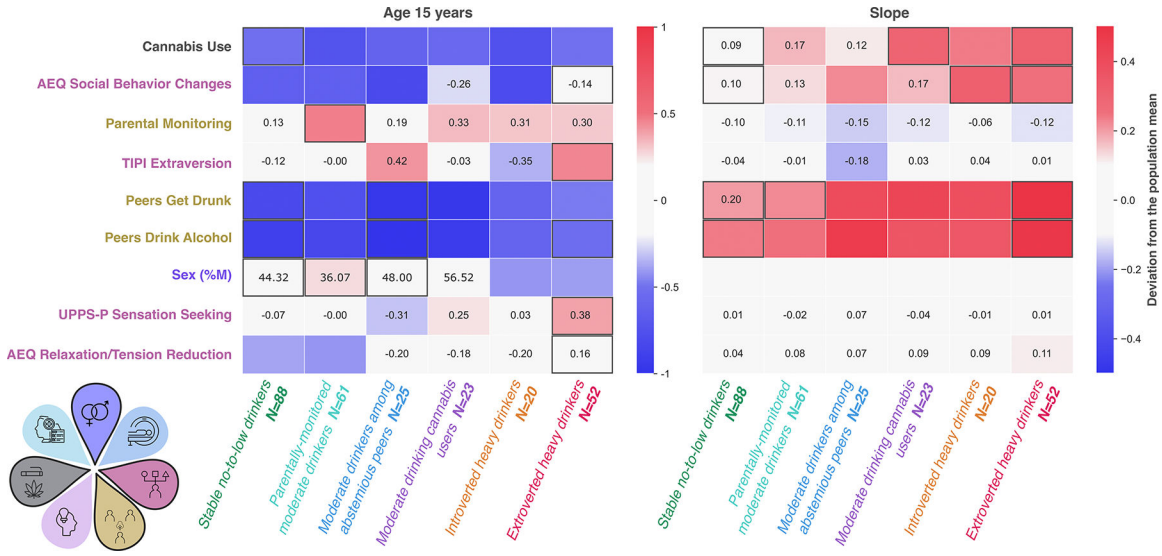


**Figure 4:**

Average standardized (standard deviations from the mean) values at 15 years and slope for the participants in each drinking pathway. Depicted are the measurements identified via permutation testing. Highlighted with a black frame are the values that are significantly ($p < 0.05$, Mann-Whitney U test) different from the population. These emphasize the importance of social and familial influences in shaping adolescent drinking behaviors over time.

**Table 1:**

Demographic information of the NCANDA participants included in this study.

| | | |
|---|---|---|
| **N (male/female)** | 285 | (144/141) |
| **Number of visits between ages 15 and 21 years (mean ± std)** | 6.26 | ± 0.89 |
| **Age (mean ± std years)** | | |
| *At the first visit* | 15.47 | ± 0.28 |
| *At the last visit* | 21.35 | ± 0.42 |
| **Number of drinks in the past month (mean ± std)** | | |
| At 15 years | 0.02 | ± 0.19 |
| At the last visit | 15.62 | ± 24.76 |
| **Site (nr. of participants, %)** | | |
| *University of California San Diego* | 76 | (26.26%) |
| *Oregon Health & Science University* | 72 | (25.26%) |
| *Duke University* | 57 | (20.00%) |
| *University of Pittsburgh* | 55 | (19.30%) |
| *SRI International* | 25 | (8.77%) |
| **Socioeconomic status** (parental highest baseline education level) | 17.18 | ± 2.39 |
| **Ethnicity (nr. of participants, %)** | | |
| *Caucasian/White* | 175 | (61.40%) |
| *African-American/Black* | 39 | (13.68%) |
| *Hispanic* | 35 | (12.28%) |
| *Asian* | 19 | (6.67%) |
| *Other* | 17 | (5.96%) |