Received: 18 November 2021 | Revised: 13 March 2022 | Accepted: 10 April 2022

DOI: 10.1111/ajt.17059

ORIGINAL ARTICLE

Artificial intelligence to identify harmful alcohol use after early liver transplant for alcohol-associated hepatitis

Brian P. Lee[1] | Nitzan Roth[2] | Prathik Rao[1] | Gene Y. Im[3] | Alexander S. Vogel[4] | Johann Hasbun[5] | Yoel Roth[6] | Akhil Shenoy[7] | Antonios Arvelakis[3] | Laura Ford[3] | Inga Dawe[3] | Thomas D. Schiano[3] | Jordan P. Davis[1] | John P. Rice[8] | Sheila Eswaran[9] | Ethan Weinberg[10] | Hyosun Han[1] | Christine Hsu[11] | Oren K. Fix[12] | Haripriya Maddur[13] | R. Mark Ghobrial[14] | George Therapondos[15] | Bistra Dilkina[1] | Norah A. Terrault[1]

[1]University of Southern California Keck School of Medicine, Los Angeles, California, USA

[2]Northwell Health, New York City, New York, USA

[3]Mount Sinai Icahn School of Medicine, New York City, New York, USA

[4]Harvard Medical School, Boston, Massachusetts, USA

[5]New York University Grossman School of Medicine, New York City, New York, USA

[6]Twitter Inc, San Francisco, California, USA

[7]Columbia University Vagelos College of Physicians and Surgeons, New York City, New York, USA

[8]University of Wisconsin, Madison, Wisconsin, USA

[9]Rush Medical Center, Chicago, Illinois, USA

[10]University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

[11]Georgetown School of Medicine, Washington, District of Columbia, USA

[12]University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA

[13]Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[14]Houston Methodist Hospital, Houston, Texas, USA

[15]Ochsner Clinic, New Orleans, Louisiana, USA

Correspondence
Norah A. Terrault, 1450 San Pablo Ave, HC4 Room 3054, Los Angeles, CA 90033, USA.
Email: terrault@usc.edu

Bistra Dilkina, 941 Bloom Walk, SAL304, Los Angeles, CA 90089, USA.
Email: dilkina@usc.edu

Funding information
University of Southern California Center for Artificial Intelligence in Society

Early liver transplantation (LT) for alcohol-associated hepatitis (AH) is the fastest growing indication for LT, but prediction of harmful alcohol use post-LT remains limited. Among 10 ACCELERATE-AH centers, we examined psychosocial evaluations from consecutive LT recipients for AH from 2006 to 2017. A multidisciplinary panel used content analysis to develop a maximal list of psychosocial variables. We developed an artificial intelligence model to predict post-LT harmful alcohol use. The cohort included training (N = 91 among 8 centers) and external validation (N = 25 among 2

Abbreviations: ACCELERATE-AH, American Consortium of Early Liver Transplantation for Alcohol-Associated Hepatitis; AH, alcohol-associated hepatitis; AI, artificial intelligence; ALD, alcohol-associated liver disease; CI, confidence interval; LT, liver transplantation; MELD, Model of End-Stage Liver Disease.

Brian P. Lee, Nitzan Roth and Prathik Rao are co-first authors.

Bistra Dilkina and Norah A. Terrault are co-senior authors.

centers) sets, with median follow-up of 4.4 (IQR 3.0–6.0) years post-LT. In the training set, AUC was 0.930 (95%CI 0.862–0.998) with positive predictive value of 0.891 (95%CI 0.620–1.000), internally validated through fivefold cross-validation. In the external validation set, AUC was 0.692 (95%CI 0.666–0.718) with positive predictive value of 0.82 (95%CI 0.625–1.000). The model identified specific variables related to social support and substance use as highly important to predict post-LT harmful alcohol use. We retrospectively developed and validated a model that identified psychosocial profiles at LT predicting harmful alcohol use post-LT for AH. This preliminary model may inform selection and post-LT management for AH and warrants prospective evaluation in larger studies among all alcohol-associated liver disease being considered for early LT.

KEYWORDS
alcoholism and substance abuse, clinical research/practice, liver transplantation/hepatology, risk assessment/risk stratification

1 | INTRODUCTION

Alcohol-associated liver disease (ALD) accounts for 50% of global liver-related mortality.[1] Liver transplantation (LT) is the only definitive therapy for life-threatening ALD, but individuals with recent alcohol use are sometimes ineligible due to abstinence restrictions. Recent studies show that early (i.e., without mandated period of abstinence) LT can be life-saving in alcohol-associated hepatitis (AH).[2,3] Indeed, increasing acceptance of early LT in the broader ALD population has contributed to ALD recently becoming the most common indication for LT, and AH as the fastest growing indication for LT in the United States and Europe.[4,5] Despite the increasing application of early LT, there remains a paucity of data to inform risk assessment in LT candidates with AH.[3]

Return to harmful alcohol use post-LT is the strongest predictor of post-LT death.[6] While we previously identified factors to be associated with harmful alcohol use post-LT, these variables used standard regression techniques among a limited set of ~15 variables based on previous studies examining psychosocial predictors of alcohol relapse in individuals with alcohol use disorder.[2] Scoring systems studied in early LT for AH which combine multiple psychosocial variables, including the Sustained Alcohol Use after Liver Transplant (SALT), Hopkins Psychosocial Scale (HPSS), High-Risk Alcohol Relapse (HRAR), and Stanford Integrated Psychosocial Assessment for Transplant (SIPAT) scores each have area under curve (AUC) values ranging 0.60–0.75 for post-LT harmful alcohol use. These reflect high negative predictive value as these models have poor positive predictive value (0–0.25) limiting clinical utility.[7–11] Most other models have focused on the broader LT for ALD population, and have not been specifically evaluated in LT candidates with short duration of abstinence. With more LT candidates than potential grafts, identifying LT candidates at high-risk for alcohol use has important clinical implications (i.e., high positive predictive value can prevent graft loss by informing listing decisions and/or personalized post-LT interventions).[9] Models with higher positive predictive value were highlighted as a key research priority in the Dallas Consensus Conference on Liver Transplantation for Alcohol-Associated Hepatitis.[9]

We hypothesized that the detailed documentation from psychosocial evaluations undertaken by LT candidates could bridge these knowledge gaps. In this study, we used sociology methodology of content analysis to systematically re-evaluate all documentation from psychosocial evaluations within the multi-center American Consortium of Early Liver Transplantation for Alcohol-Associated Hepatitis (ACCELERATE-AH) to maximize identification of potential variables associated with post-LT harmful alcohol use. Using a widely expanded list of psychosocial variables, we subsequently applied artificial intelligence (AI) to obtain a model with higher positive predictive value and to identify novel predictors of post-LT harmful alcohol use.

2 | METHODS

2.1 | Study population

Ten LT centers provided detailed retrospective data on consecutive adults transplanted early (i.e., without a mandated period of abstinence, and less than 6 months of abstinence) with the indication of severe AH from 2006 to 2017, as previously described.[6] Other key inclusion criteria were clinically-diagnosed severe acute AH as the indication for LT and no prior diagnosis of chronic liver disease or episodes of AH. Liver biopsy was not required to confirm AH as this practice is atypical, and the majority of recipients met inclusion criteria[12] for AH as defined by the NIAAA Alcoholic Hepatitis Consortia using explant pathology to assess for histologic findings of steatohepatitis.

2.2 | Content analysis for variable generation from psychosocial evaluations

Variables of interest from the psychosocial evaluation were generated through content analysis,[13] rather than literature review. Content analysis[13] uses a rigorous approach to analyze and contextualize verbal narratives or documents through six steps: (1) defining the study design; (2) unitizing, which consists of development of the taxonomy or schema of units of analysis in the text data; (3) sampling, which consists of obtaining representative text data; (4) coding, which consists of human or computer categorization of the sampled data using the taxonomy; (5) analysis and interpretation; and (6) validation. LT centers provided deidentified psychosocial evaluation narratives as originally documented verbatim in medical records pre-LT by social workers, psychiatrists, and/or addiction specialists. LT recipients had at least 1 (e.g., social worker only) and maximum of three psychosocial evaluators (e.g., social worker, psychiatrist, addiction counselor). Narratives were transcribed at the data coordinating center (UCSF) to a standard free-text format to remove any information and/or formatting suggesting center origin, time period, and identifiers that may bias analyses. Then, as guided by a social scientist (Y.R., Twitter Inc.), a separate multidisciplinary panel of six experts at Mount Sinai Hospital (transplant surgeon, two hepatologists, transplant social worker, transplant psychiatrist, transplant coordinator) who were all blinded to center and outcomes used content analysis on a sample of 10 psychosocial narratives (1 randomly chosen per LT center) to develop a taxonomy of the narratives' organization, content, and quality, resulting in 219 variables within 18 domains (Tables S1 and S2). Next, two separate coders, also blinded, used the taxonomy to analyze and quantify the narratives. Interrater reliability was not measured as all discrepancies between coders were adjudicated in regular teleconference meetings led by a third coder until consensus was achieved. Only the final adjudicated variables were used in this analysis.

2.3 | Primary outcome

The primary outcome was harmful alcohol use post-LT, defined as any evidence of binge (≥5 drinks in men, ≥4 drinks in women in one setting) or frequent (≥4 days in the week) drinking, consistent with federal definitions.[14] Harmful alcohol use represents a standardized outcome advocated by federal organizations.[14] Binge and/or frequent patterns of drinking are highly associated with post-LT death in ALD, whereas other patterns have not shown this association.[6,15–17] One drink was defined as a US standard drink (i.e., 14 g of alcohol).

LT recipients who survived to home discharge post-LT ($n = 116$) were interviewed and questioned regarding presence, quantity, and frequency of alcohol use at every post-LT visit (typically at least every 3 months in the first year post-LT, and every 6 months thereafter) and responses were documented at all centers. In addition, as previously reported, almost all patients had monitoring of post-LT alcohol use with biomarkers of alcohol metabolites via urine ethyl glucuronide or blood phosphatidylethanol testing, and that there is 5% missing longitudinal alcohol use data.[6,14,15]

2.4 | Artificial intelligence model to predict post-LT harmful alcohol use

Multiple machine learning algorithms (e.g., logistic regression, Random Forest, XGBoost, with or without mode imputation of missing data) were evaluated to gauge which would best predict harmful alcohol use post-LT. Among these models, XGBoost[18] without imputation of missing data was identified as the top performing model based on positive predictive value and area under the curve c-statistic (Table S3). XGBoost[18] relies on the nature of decision trees to identify non-linear relationships between predictors. Additionally, XGBoost's implementation supports missing data which was of key interest in this study as interviewers do not necessarily ask the same questions to patients during LT evaluation, and we hypothesized that having the question "missed" may actually be predictive of the outcome. For example, we hypothesized that a patient with overt encephalopathy may have much of the specifics in their alcohol history missing, or a psychosocial evaluator who does not conduct a thorough evaluation may have more missing values, could be scenarios associated with higher risk (i.e., point estimate) or higher uncertainty (i.e., wider confidence interval) of harmful alcohol use post-LT. Thus, we sought to retain missingness within each variable as a potential predictor. Training leverages "gradient boosting," which is a gradient descent algorithm to minimize the loss when adding new models.

2.5 | Artificial intelligence pipeline

For our pipeline, several steps were taken to ensure our model would be reliable, robust, and generalizable to new data from external centers, detailed in the Supplemental Methods. Briefly, we a priori split the study population into a training ($N = 91$ among 8 centers) and validation ($N = 25$ among 2 centers) sets. Fivefold cross-validated forward feature selection (FFS)[19] was applied to the data to reduce the overall variable set. Once we reduced our variable set, a single fivefold cross-validation was performed on our training data to generate five models—each training on 80% of the training data and using 20% as internal validation. Once training was complete, each of the five models made predictions on the external validation set. Finally, each of the five models was examined to identify which of the variables were most important to each model's prediction. XGBoost offers a built-in metric for variable importance: Mean Decrease in Impurity (MDI). The associated importance of variables across the five models were averaged and a total of 13 variables were identified as representative of an ensemble model predictive of post-LT harmful alcohol use (optimized for AUC). We also calculated other model performance metrics, including positive predictive value, negative predictive value, specificity, and sensitivity to provide objective performance metrics in training and validation sets.

The threshold for predicted probability was 0.5 for these reported performance metrics. AI analyses were performed using scikit-learn version 0.21 (https://scikit-learn.org/) and xgboost version 1.5 (https://xgboost.readthedocs.io/).

2.6 | Statistical analysis

Patient characteristics were described using means (SDs), medians (interquartile ranges [IQRs]), and proportions as appropriate. Categorical variables were compared using chi-square test. The sample size was fixed, representing all patients among ACCELERATE-AH sites participating in this substudy; thus power was not calculated—confidence intervals are provided to guide interpretation of results.

This study was approved by the Institutional Review Board at each participating site. This study used the EQUATOR TRIPOD checklist for Prediction Model Development and Validation.

3 | RESULTS

3.1 | Study population (*N* = 116)

Among 116 LT recipients surviving to post-LT home discharge, with median 4.4 year (IQR 2.8–6.0) follow-up post-LT, 34 (29%) had evidence of post-LT harmful alcohol use. Median age was 42 years (IQR 36–50), 72% male, and interval between last drink and LT listing of 54 days (IQR 36–94) (Table 1). We a priori split the study population into a training (*N* = 91 among 8 centers) and validation (*N* = 25 among 2 centers) sets. Baseline characteristics stratified by number of psychosocial evaluators are summarized in Table S4. The proportion of patients with 2 or more psychosocial evaluations was similar among training versus external validation sets (52% vs. 65%, $p = 0.37$).

3.2 | Training set model performance (*N* = 91)

The training set included 91 LT recipients from 8 centers, of which 27 (30%) had post-LT harmful alcohol use; baseline characteristics of LT recipients in the training set (*N* = 91) vs. external validation set (*N* = 25) are summarized in Table 1. Training validation performance metrics were AUC 0.930 (95%CI 0.862–0.998), positive predictive value 0.891 (95%CI 0.620–1.000), negative predictive value 0.851 (95%CI 0.730–0.972), specificity 0.908 (95%CI 0.678–1.000), and sensitivity 57.3% (95%CI 23.1–91.5%).

3.3 | External validation (*N* = 25)

The external validation set included 25 LT recipients from 2 centers, of which 7 (28%) had post-LT harmful alcohol use. Performance metrics were AUC 0.692 (95%CI 0.666–0.718), positive predictive value 0.820 (95%CI 0.625–1.000), negative predictive value 0.811 (95%CI

TABLE 1 Patient characteristics in training and validation sets

Characteristic at listing	Training set (*n* = 91)	Validation set (*n* = 25)
Age–year–median (IQR)	42 (36–50)	44 (37–48)
Male, *n* (%)	66 (73)	17 (68)
Race/Ethnicity, *n* (%)		
Caucasian	76 (84)	21 (84)
African American	4 (4)	2 (8)
Hispanic	7 (8)	0 (0)
Asian	1 (1)	2 (8)
Other	3 (3)	0 (0)
Employed, *n* (%)	50 (55)	13 (52)
Medical insurance, *n* (%)		
Private	60 (66)	20 (80)
Medicare	11 (12)	1 (4)
Medicaid	20 (22)	4 (16)
Married/Stable companion, *n* (%)	60 (66)	14 (56)
History of co-morbid psychiatric disease, *n* (%)	35 (38)	10 (40)
Substance abuse history, *n* (%)		
Active smoker	20 (22)	4 (17)
Marijuana	9 (10)	5 (21)
Non-Marijuana illicit substance	11 (12)	1 (4)
History of failed rehabilitation[a] attempt, *n* (%)		
No prior attempt	59 (65)	24 (96)
1 Prior attempt	21 (23)	1 (4)
≥2 Prior attempts	11 (12)	0 (0)
Family history of alcohol use disorder[b], *n* (%)		
First degree relative	4 (4)	6 (25)
Second degree relative only	24 (27)	4 (17)
History of alcohol-related legal issues, *n* (%)		
1 Prior episode	17 (19)	2 (8)
≥2 Prior episodes	9 (10)	1 (4)
Alcohol consumption immediately prior to hospitalization–units/day–median (IQR)	10 (6–15)	9 (5–16)
Years of heavy drinking–median (IQR)	13 (8–20)	23 (10–30)
Sodium–mg/dl–median (IQR)	135 (133–139)	136 (132–139)
INR–median (IQR)	2.2 (1.8–2.5)	2.2 (1.8–3.0)
Total Bilirubin–mg/dl–median (IQR)	25.7 (19.8–36.0)	23.7 (16.8–29.5)
Creatinine–mg/dl–median (IQR)	2.6 (1.7–3.9)	2.6 (1.4–4.5)
Renal replacement therapy, *n* (%)	42 (46)	14 (56)
Mechanical ventilation, *n* (%)	14 (16)	6 (24)
Encephalopathy west-haven grade, *n* (%)		
None	30 (34)	3 (12)

(Continues)

TABLE 1 (Continued)

Characteristic at listing	Training set ($n = 91$)	Validation set ($n = 25$)
Grade 1	20 (22)	2 (8)
Grade 2	19 (21)	9 (36)
Grade 3	5 (6)	5 (20)
Grade 4	15 (17)	6 (24)
MELD-Na score–median (IQR)	38 (35–40)	40 (38–41)
Time between last drink and LT–days–median (IQR)	53 (36–101)	59 (42–85)
Follow-up Time–years–median (IQR)	4.1 (2.7–5.8)	5.3 (4.6–6.6)

[a] Rehabilitation program defined as formal intensive outpatient or inpatient treatment program dedicated to alcohol addiction.

[b] Family history among biologic relatives only.

0.803–0.819), specificity 0.956 (95%CI 0.726–1.000), and sensitivity 42.9% (95%CI 42.9–42.9%).

3.4 | Novel predictors of harmful alcohol use post-LT and example patients

The strongest predictors of post-LT harmful alcohol use were primary support person for post-LT care not yet being identified at the time of the psychosocial evaluation, presence of children/grandchildren living with the patient, and whether the patient was recently a home caregiver for relatives. Two of the 13 variables in the final model were related to history of opioid use disorder. The primary model is summarized in Table 2. Variable intercorrelation among the final 13 variables was assessed, summarized in Figure S1. Psychosocial profiles of example patients predicted risk of post-LT harmful alcohol use by AI are presented in Table 3.

The final model is available at https://halt-ai-v2.herokuapp.com/.

4 | DISCUSSION

In this multicenter cohort of early LT for AH, we describe the application of AI to predict post-LT harmful alcohol use and provide proof of concept for AI-based prediction of post-LT outcomes. We leveraged mixed methodology, first applying content analysis from social science to identify hundreds of potentially important variables, followed by AI to distill these variables into a model containing 13 variables optimized to predict harmful alcohol use post-LT. While modest in size for prediction, this study represents the largest US dataset in this patient population to provide important preliminary insights to direct future studies. Our results suggest promise in AI to augment positive predictive value and to identify novel predictors of post-LT harmful alcohol use, which can serve as an adjunct tool for transplant providers to tailor interventions for alcohol use disorder based on a predicted risk of alcohol relapse. The model is

TABLE 2 Psychosocial variables in final model to predict harmful alcohol use post-LT

#	Psychosocial variable	Coef[a] (±SD)
1	Patient's primary support person for peri- and post-LT care has not yet been identified at time of this evaluation	16.3 ± 4.2
2	Are there any pediatric children or grandchildren (<18 years old) who live with the patient?	10.6 ± 1.4
3	Was the patient recently a home caregiver for children or elderly relatives?	10.2 ± 0.7
4	Has the patient ever abused opioid pills?	10.0 ± 7.0
5	Is the patient observant in religion and/or attend services regularly?	9.5 ± 2.5
6	If applicable, does the patient currently have a healthy/strong relationship with his/her siblings?	7.8 ± 3.0
7	Did the patient ever complete a rehabilitation program?	7.3 ± 1.5
8	During the interview, did the patient make eye contact with the writer?	6.8 ± 3.3
9	Is the writer's background in social work?	6.2 ± 4.2
10	Has the patient ever been treated with methadone for opioid addiction?	6.2 ± 2.2
11	Medicaid/Medicare (vs. Private/Other) insurance?	5.2 ± 4.9
12	Did the writer discuss potential living donors?	3.0 ± 2.5
13	Patient's primary support person for peri- and post-LT is non-spouse/significant other (vs. spouse or significant other)	0.9 ± 1.2

[a]Coefficient is the Gini coefficient from XGBoost, to be interpreted as relative importance of the variable in predicting harmful alcohol use post-LT, calculated as the mean importance with standard deviation (SD) across the fivefold internal cross-validation of the training set. The coefficient does not have a fixed "direction" (positive or negative) in XGBoost models. The XGBoost model is a "tree" of variables rather than individual variables. Higher coefficients indicate variables that are higher in the tree. An answer (yes or no) to any of these 13 variables can infer positive risk with one combination of other variables, but negative risk with other variables, as the tree needs to be interpreted as a unique combination of all 13 variables.

TABLE 3 Example psychosocial profiles with corresponding probability of harmful alcohol use post-LT by artificial intelligence model

#	Psychosocial variable	Patient 1 (low risk)	Patient 2 (encephalopathy)	Patient 3 (high risk)
1	Has the patient's primary support person for peri- and post-LT care been identified yet at time of this evaluation?	Yes	Yes	No
2	Are there any pediatric children or grandchildren (<18 years old) who live with the patient?	No	Yes	Yes
3	Was the patient recently a home caregiver for children or elderly relatives?	Yes	No	No
4	Has the patient ever abused opioid pills?	Never	Not collected	Former
5	Is the patient observant in religion and/or attend services regularly?	Yes	Yes	No
6	If applicable, does the patient currently have a healthy/strong relationship with his/her siblings?	Yes	Yes	No
7	Did the patient ever complete a rehabilitation program?	No	Not Collected	Yes
8	During the interview, did the patient make eye contact with the writer?	Yes	No	No
9	Is the writer's background in social work?	Yes	No	Yes
10	Has the patient ever been treated with methadone for opioid addiction?	Never	Not Collected	Current
11	Medicaid/Medicare (vs. Private/Other) insurance?	No	Yes	Yes
12	Did the writer discuss potential living donors?	No	No	No
13	Patient's spouse or significant other (vs. non-spouse/significant other) has been identified as primary support person for peri and post-LT care	Yes	Yes	Not Collected
Probability of Harmful Alcohol Use Post-LT (95% CI)		8.3% (0–20.3%)	45.4% (8.7–82.1%)	93.4% (88.6–98.2%)

not intended to be used to singlehandedly deny transplant or replace transplant committee processes for decision-making, as this could lead to inequity or potential disparities. These results can serve as the basis to design larger prospective studies, which can refine selection practices and direct targeted post-LT interventions in the most rapidly growing indication for LT in the United States and Europe.

Our AI model appears to have superior positive predictive value (82% in external validation) versus existing scoring systems (0%–25%), and could eventually replace these other existing scoring systems if the enhanced performance metrics are validated prospectively.[7–11] Many ALD prediction studies focus on AUC, but our study highlights that this may not be the most appropriate metric, particularly in ALD where alcohol relapse is a relatively infrequent, but important outcome. Models to predict infrequent events inherently allow high negative predictive value, which can skew AUC higher and mask a low positive predictive value. Rates of post-LT alcohol use in ACCELERATE-AH also appear to be relatively infrequent, similar to historic cohorts in LT for ALD. In our model, the positive predictive value is high, but the sensitivity is relatively low. This likely reflects challenges to predict future alcohol use; future alcohol use can be precipitated by incident and unpredictable stressors which cannot be known at time of LT, and other predictors yet to be analyzed or discovered. Whether AI can provide a more individualized approach to help predict patient behavior in response to future stressors and which interventions are best suited to prevent and treat alcohol relapse are promising future research directions in this field. Application of this model could help selection committees exclude high-risk patients, and allow transplant providers to combine early LT with targeted post-LT treatments for alcohol use disorder. For example, an early LT recipient with predicted probability of post-LT harmful alcohol use exceeding the rate already observed by the center's LT for ALD cohort could preferentially be targeted for more intensive post-LT treatments for alcohol use disorder. The low sensitivity emphasizes the need for close monitoring for post-LT alcohol use, aggressive application of post-LT management of alcohol use disorder, and future research with larger cohorts and longer follow-up.

As opposed to strict cut-offs for "high-risk" versus "low-risk", our AI model differs from other scoring tools by providing a personalized percentage probability with 95%CI of harmful alcohol use for any permutation of psychosocial factors. XGBoost can model missingness as a unique value for any given variable, rather than regression techniques which either exclude or impute missing values. This analytic strength was particularly useful for our patient population: as exemplified by the example patient with hepatic encephalopathy, the amount of missing data can significantly increase the uncertainty of risk, which can be more easily appreciated by LT providers by wide 95%CI. We also hypothesize that the amount of missing data may be related to less thorough or less experienced psychosocial evaluators, contributing to uncertainty of risk, which could be modified with appropriate training, and by setting standards for best practices

in a comprehensive psychosocial evaluation. These are examples of the potential applications of a validated AI model, to move toward a more personalized medicine approach in early LT for ALD.

Our model is distinct from regression techniques which provide a "direction" (positive or negative) of risk (e.g., odds or hazard ratio) for variables individually. Having a fixed direction of a risk factor can conceal the possibility that the variable may confer positive or negative risk depending on potentially complex combinations with other variables. To address this limitation, our AI model identifies risk factors without assigning fixed "directions" of risk, and is better-positioned to identify complex interactions between large combinations (i.e., trees) of risk factors. Our AI model is a "tree" of 13 variables rather than 13 individual variables, and is interpreted as a unique combination of all 13 variables, which may better leverage the value of a comprehensive psychosocial evaluation.

The three variables with highest coefficients were all related to social support, which has been a known protective factor for alcohol relapse.[20,21] The requirement of "strong social support" has often been criticized as subjective, and our study provides clear questions which may help providers to gauge social support in this patient population. Additionally, we identified a number of novel protective variables, including whether the patient maintained eye contact and whether the patient was recently a caregiver for relatives, which we hypothesize may be related to hepatic encephalopathy, impairing the ability to conduct an accurate psychosocial evaluation and motivation to return to good health, respectively. The presence of young children has been a recurring predictor of alcohol relapse in prior LT for ALD cohorts,[22,23] hypothesized to represent a stressor within the households of LT recipients, which may not be apparent to many LT providers. These findings reflect the value in a comprehensive psychosocial evaluation, as opposed to excessive prioritization of single or few risk factors (e.g., duration of abstinence pre-LT) to justify LT eligibility. While risk factors may not be necessarily modifiable, they may still be clinically meaningful and should encourage providers and researchers alike to further probe the meaning of risk factors and how they contribute to risk of alcohol relapse—for example, young children should clearly not be a reason to deny LT, but if there are young children in the household, LT providers may investigate the relationships of children to the patient and primary LT support person, and plans for childcare post-LT. Indeed, clinicians may find the list of variables in our primary model (Table 2) as a minimum foundation for their own comprehensive psychosocial evaluation in early LT for AH.

While transplant center was not a key variable in the model, there were variables (e.g., whether the psychosocial evaluator discussed live donors, background training of psychosocial evaluator) which seem to suggest potential center effect (i.e., most ACCELERATE-AH centers do not perform living donor LT, and not all have access to transplant psychiatrists). Whether these findings are related to differences in patient populations, selection practices, provider experience, or post-LT management resources remain unclear, and will be an important area of future prospective study. The model does not include peri- or post-operative events, or post-LT management resources to mitigate risk of post-LT alcohol use, which may vary by center. However, while there was a drop-off in AUC, the model did perform reasonably well in a validation set separated a priori by center, which suggests generalizability and consistent importance of identified variables across centers.

There were limitations. First, the sample size is relatively modest for AI, and the data are retrospective and subject to biases, missingness, and potential underreporting of alcohol use. The model would benefit from validation in prospective and more diverse cohorts to enhance generalizability. However, ACCELERATE-AH provides the largest US dataset and only US consortium in this field, which is in urgent need of data to inform selection process, given the rapidly expanding uptake of early LT for AH. Additionally, the relatively narrow confidence intervals for AUC in external validation suggests a sufficient sample size. As highlighted in a recent *Gastroenterology* review[24] and *Hepatology* review[25] regarding the landscape of AI in gastroenterology and hepatology, ALD was consistently and specifically identified as an area of high unmet need, and the promising results from our study should encourage future prospective studies to investigate novel ways to predict post-LT alcohol use. Second, this cohort represents a heterogeneous group of patients, including patients without biopsy-proven AH, and among which selection practices were not standardized across centers and could have changed over time. Patients may also have been exposed to different peri- or post-operative management, and post-LT treatments for alcohol use disorder, the contribution of which were outside the scope of this study, which could have introduced differences between training and validations groups, and also impact outcomes. However, even though there was drop-off in model performance in external validation, which may suggest some overfitting, the model did still perform reasonably well by AUC compared to other scoring systems for post-LT alcohol use, which range 0.60–0.75.[7–11] Additionally, we provide both internal and external validation, which is a key strength. Continued refinement of this artificial intelligence model as the cohort grows and with longer follow-up is desirable. Third, the exact significance of the variables remains unclear, and some (e.g., eye contact) may be subjective. However, identification of these variables in this study should be foundational to advance research to better understand and identify novel predictors of harmful alcohol use. Likewise, this study cohort represents a carefully selected group of LT recipients; patients are unlikely to be LT recipients with uncontrolled psychiatric disease or inability to commit to treatment for alcohol use disorder. Thus, variables not selected in the final prediction model (e.g., assessment of severity of alcohol use disorder, insight into alcohol use disorder, co-morbid psychiatric disease, and ability to engage meaningfully in treatment for alcohol use disorder), which are standard components of a typical comprehensive psychosocial evaluation, should not be ignored. These components should still be carefully evaluated and considered with the results of any risk prediction tool—for example, having these components align with the predicted risk from the artificial intelligence model could help reassure selection decisions, whereas discordance could help

prompt further reflection or investigation to make a final selection decision. Finally, clinical variables were not assessed as the purpose of this study was to inform the psychosocial evaluation.

In conclusion, AI may help predict post-LT harmful alcohol use among carefully selected patients with life-threatening AH without prolonged abstinence. Our results may be especially promising to augment positive predictive value and to identify novel predictors of post-LT harmful alcohol use. If confirmed in prospective studies, these findings may help to refine selection practices and direct targeted post-LT interventions in early LT for ALD.

DISCLOSURE

The authors of this manuscript have no conflicts of interest to disclose as described by the *American Journal of Transplantation*.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Brian P. Lee https://orcid.org/0000-0003-2108-1287
Gene Y. Im https://orcid.org/0000-0003-0009-8418
Akhil Shenoy https://orcid.org/0000-0001-8639-2751
Haripriya Maddur https://orcid.org/0000-0001-7192-9501
Norah A. Terrault https://orcid.org/0000-0003-4143-1950

REFERENCES

1. Singal AK, Bataller R, Ahn J, Kamath PS, Shah VH. ACG clinical guideline: alcoholic liver disease. *Am J Gastroenterol*. 2018;113:175-194.
2. Lee BP, Mehta N, Platt L, et al. Outcomes of early liver transplantation for patients with severe alcoholic hepatitis. *Gastroenterology*. 2018;155(2):422–430.e1.
3. Lee BP, Samur S, Dalgic OO, et al. Model to calculate harms and benefits of early vs delayed liver transplantation for patients with alcohol-associated hepatitis. *Gastroenterology*. 2019;157(2):472–480.e5.
4. Lee BP, Vittinghoff E, Dodge JL, Cullaro G, Terrault NA. National trends and long-term outcomes of liver transplant for alcohol-associated liver disease in the United States. *JAMA Intern Med*. 2019;179:340-348.
5. Cotter TG, Sandıkçı B, Paul S, et al. Liver transplantation for alcoholic hepatitis in the United States: excellent outcomes with profound temporal and geographic variation in frequency. *Am J Transplant*. 2021;21(3):1039-1055.
6. Lee BP, Im GY, Rice JP, et al. Patterns of alcohol use after early liver transplantation for alcoholic hepatitis. *Clin Gastroenterol Hepatol*. 2022;20(2):409-418.e5.
7. Lee BP, Vittinghoff E, Hsu C, et al. Predicting low risk for sustained alcohol use after early liver transplant for acute alcoholic hepatitis: the sustained alcohol use post-liver transplant score. *Hepatology*. 2019;69:1477-1487.
8. Lee BP, Chen P-H, Haugen C, et al. Three-year results of a pilot program in early liver transplantation for severe alcoholic hepatitis. *Ann Surg*. 2017;265:20-29.
9. Asrani SK, Trotter J, Lake J, et al. Meeting report: the Dallas consensus conference on liver transplantation for alcohol associated hepatitis. *Liver Transpl*. 2020;26(1):127–140
10. Deutsch-Link S, Weinrieb RM, Jones LS, Solga SF, Weinberg EM, Serper M. Prior relapse, ongoing alcohol consumption, and failure to engage in treatment predict alcohol relapse after liver transplantation. *Dig Dis Sci*. 2020;65:2089-2103.
11. Zhou M, Wagner LM, Diflo T, Naegle M. Implementation of the high-risk alcoholism relapse scale in a liver transplant clinic. *Gastroenterol Nurs*. 2015;38:447-454.
12. Crabb DW, Bataller R, Chalasani NP, et al. Standard definitions and common data elements for clinical trials in patients with alcoholic hepatitis: recommendation from the NIAAA alcoholic hepatitis consortia. *Gastroenterology*. 2016;150:785-790.
13. Krippendorff K. *Content Analysis: An Introduction to its Methodology*, 4th ed. SAGE Publications; 2019.
14. Mellinger J, Winder GS, Fernandez AC. Measuring the alcohol in alcohol-related liver disease: choices and challenges for clinical research. *Hepatology*. 2021;73(3):1207-1212.
15. Rice JP, Eickhoff J, Agni R, Ghufran A, Brahmbhatt R, Lucey MR. Abusive drinking after liver transplantation is associated with allograft loss and advanced allograft fibrosis. *Liver Transpl*. 2013;19:1377-1386.
16. Dumortier J, Dharancy S, Cannesson A, et al. Recurrent alcoholic cirrhosis in severe alcoholic relapse after liver transplantation: a frequent and serious complication. *Am J Gastroenterol*. 2015;110(8):1160–1166. quiz 7.
17. DiMartini A, Dew MA, Day N, et al. Trajectories of alcohol consumption following liver transplantation. *Am J Transplant*. 2010;10:2305-2312.
18. XGBoost at https://xgboost.readthedocs.io/
19. Ververidis D, Kotropoulos C. Sequential forward feature selection with low computational cost. 13th European Signal Processing Conference. 2005;1-4.
20. DiMartini A, Day N, Dew MA, et al. Alcohol consumption patterns and predictors of use following liver transplantation for alcoholic liver disease. *Liver Transpl*. 2006;12:813-820.
21. De Gottardi A, Spahr L, Gelez P, et al. A simple score for predicting alcohol relapse after liver transplantation. *Arch Intern Med*. 2007;167(11):1183.
22. Pfitzmann R, Schwenzer J, Rayes N, Seehofer D, Neuhaus R, Nussler NC. Long-term survival and predictors of relapse after orthotopic liver transplantation for alcoholic liver disease. *Liver Transpl*. 2007;13:197-205.
23. Foster P, Fabrega F, Karademir S, Sankary H, Mital D, Williams J. Prediction of abstinence from ethanol in alcoholic recipients following liver transplantation. *Hepatology*. 1997;25:1469-1477.
24. Le Berre C, Sandborn WJ, Aridhi S, et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology*. 2020;158(1):76–94.e2.
25. Ahn JC, Connell A, Simonetto DA, Hughes C, Shah VH. Application of artificial intelligence for the diagnosis and treatment of liver diseases. *Hepatology*. 2021;73:2546-2563.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lee BP, Roth N, Rao P, et al. Artificial intelligence to identify harmful alcohol use after early liver transplant for alcohol-associated hepatitis. *Am J Transplant*. 2022;22:1834–1841. doi:10.1111/ajt.17059