

Allen M. Smith<sup>1</sup>, Horacio Gomez-Acevedo<sup>2</sup>, Corey Hayes<sup>1</sup>, Melody Greer<sup>2</sup>, Bradley C. Martin<sup>1</sup>

<sup>1</sup> Division of Pharmaceutical Evaluation and Policy, Department of Pharmacy Practice, University of Arkansas for Medical Sciences, Little Rock, AR, USA; <sup>2</sup> Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

## BACKGROUND

- Currently, **24 states** have legalized cannabis for **adult recreational use** and **39 states** have legalized cannabis for **medical use**.<sup>1</sup>
- Among U.S. adults, the 2023 National Survey on Drug Use and Health (NSDUH) estimated that
  - ~**58.9 million (22.9%)** used cannabis in the past year.<sup>2</sup>
  - ~**17.9 million (7.0%)** met DSM-V criteria for CUD in the past year.<sup>2</sup>
- No **CUD risk prediction tool** has been developed among persons authorized to purchase **medical cannabis**
  - The underlying relationship between **CUD risk factors** and **CUD development** may look **different among cannabis users** relative to the **general population**<sup>3</sup>

**Objective:** To develop a **discrete-time-updating algorithm** to **predict CUD risk** within the **next 90 days** among **Arkansas medical marijuana (MMJ) cardholders** without a history of CUD.

## METHODS

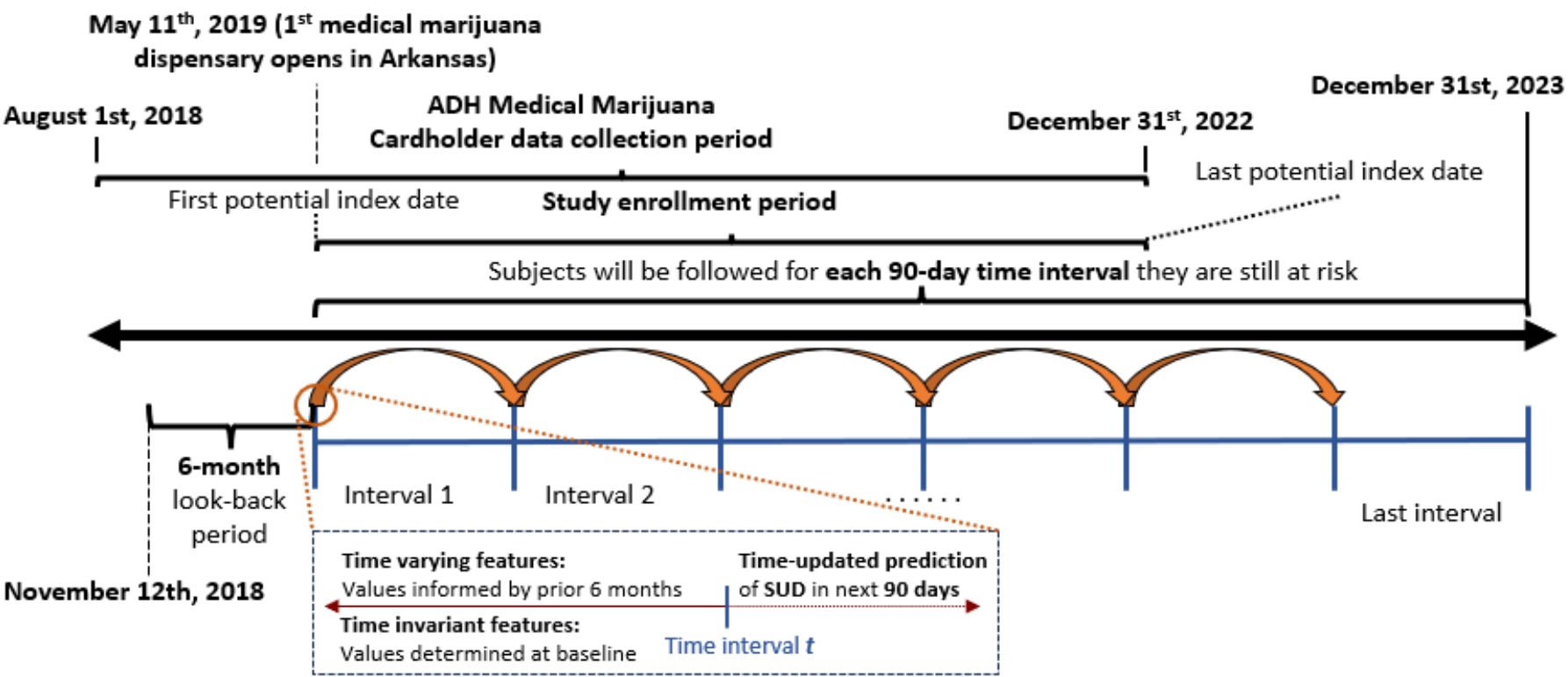
### Data Source

- This study utilized statewide health insurance claims data linked to Arkansas Department of Health (ADH) MMJ cardholder data between November 2018 – December 2023 from the **Arkansas All-Payer Claims Database (AR-APCD)**.<sup>4</sup>

### Study Sample

- Index Date:** May 11th, 2019 (Date 1<sup>st</sup> Arkansas MMJ dispensary opened) or receipt date of MMJ eligibility card, whichever came last
- Inclusion Criteria:**
  - MMJ cardholder **eligibility** between August 1<sup>st</sup>, 2018 and December 31<sup>st</sup>, 2022.
  - ≥ **18 years old** at the time of the index date.
  - ≥ **180 days of continuous medical and pharmacy benefits enrollment** prior to the index date
  - ≥ **180 days without a CUD diagnosis** prior to the index date.
- Follow-up:** Index date until whichever of the following first occurs: 1) New CUD diagnosis, 2) reach study end date (Dec. 31<sup>st</sup>, 2023), 3) health plan disenrollment, 4) death from any cause
- Data structure (Discrete-time-updating):** Person-period dataset (subject follow-up split into 90-day time intervals), where CUD prediction for each time interval informed by prior 6 months of features

### Study Design



### Feature Space [n=174]

- Labeled **prognostic** if evidenced by prior literature [**n=43**], labeled **agnostic** otherwise [**n=131**]
  - Demographics** (Age, sex, insurance payer type) [**n=3**]
  - Acute/chronic Comorbidities** (Categorized using Clinical Classifications Software Refined) [**n=114**]
  - Prescription Characteristics** (Categorized using First Databank therapeutic classes) [**n=52**]
  - Healthcare Utilization** (E.g. Cumulative healthcare costs, hospitalization count) [**n=5**]

### Model Training/Testing

- Train/test split:** Random 70:30 split at person level
- Data balancing:** 1:1, 1:3, and 1:10 random undersampling (RUS) of the majority class
- Hyperparameter tuning:** 90 iterations with 5-fold cross validation
- Feature selection:** Two-pronged recursive feature elimination approach:
  - Random Forest-derived feature importance scores
  - Cox proportional hazards-derived p-values

## METHODS

**Classifiers:** Random Survival Forest (RSF), Gradient Boosting (GB), Support Vector Machine (SVM), Cox Proportional Hazards Survival Model (CPH), Random Forest (RF), Logistic Regression (LR).

**Performance Metrics:** Cumulative sensitivity/dynamic specificity area under the receiver-operating characteristic (C/D AUC), Brier Score, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV)

## RESULTS

A total of **54,422 Arkansas MMJ cardholders** met eligibility, of which **1,479 (2.72%)** received a new CUD diagnosis during the follow-up period.

Model Performance for the Prediction of Cannabis Use Disorder within the next 90 days among Arkansas Medical Marijuana Cardholders						
Classifier	RUS 1:1		RUS 1:3		RUS 1:10	
	Mean C/D AUC	Mean Brier Score	Mean C/D AUC	Mean Brier Score	Mean C/D AUC	Mean Brier Score
Random Survival Forest <sup>a</sup>	0.7183	0.0237	0.7309	0.0082	0.7323	0.0054
Gradient Boosting <sup>a</sup>	0.7712	0.0250	0.7570	0.0079	0.7463	0.0054
Support Vector Machine <sup>a</sup>	<b>0.7776</b>	0.0242	<b>0.7905</b>	0.0075	0.7976	<b>0.0048</b>
Cox Proportional Hazards <sup>a</sup>	0.7683	0.0263	0.7802	0.0088	0.7888	0.0053
Random Forest <sup>b</sup>	0.6596	0.0102	0.6834	<b>0.0061</b>	0.6877	0.0053
Logistic Regression <sup>b</sup>	0.7045	0.0111	0.7721	0.0062	<b>0.8068</b>	0.0053
Random Forest <sup>c</sup>	0.6684	<b>0.0098</b>	0.6733	0.0061	0.6768	0.0053
Logistic Regression <sup>c</sup>	0.6790	0.0119	0.7118	0.0066	0.7233	0.0053

C/D AUC = Cumulative sensitivity/dynamic specificity area under the receiver-operating characteristic, NPV = Negative Predictive Value, PPV = Positive Predictive Value, RUS = Random Undersampling

<sup>a</sup>Time interval included for calculation of time-dependent survival probabilities

<sup>b</sup>Time interval included as feature

<sup>c</sup>Time interval excluded from model training

Top 10 Predictors of the Best Performing Models				
Logistic Regression (RUS 1:10)			Support Vector Machine (RUS 1:10)	
Rank	Feature <sup>a</sup>	$\chi^2$ Value	Feature <sup>a</sup>	Coefficient Absolute Value
1	Time Interval	0.085	Age	0.130
2	Sex	0.064	Exposure/screening with infectious disease <sup>c</sup>	0.097
3	Age	0.059	Medical examination/evaluation <sup>c</sup>	0.086
4	Depression <sup>c</sup>	0.026	Back Pain <sup>c</sup>	0.083
5	Tobacco Use Disorder <sup>c</sup>	0.024	Other perinatal conditions <sup>b</sup>	0.081
6	Bipolar and Related Disorders <sup>c</sup>	0.020	Count of PCP visits <sup>d</sup>	0.075
7	Malaise and Fatigue <sup>b</sup>	0.012	Limb/Extremity/Joint Pain <sup>c</sup>	0.074
8	Anxiety <sup>c</sup>	0.011	Sleep Disorders <sup>c</sup>	0.066
9	Ischemic Heart Disease <sup>c</sup>	0.009	Abdominal and Bowel Pain <sup>c</sup>	0.065
10	Lifestyle/life management factors <sup>c</sup>	0.009	Chronic Eye disorders <sup>c</sup>	0.057

CCSR = Clinical Classifications Software Refined, PCP = Primary Care Provider visits, RUS = Random Undersampling

<sup>a</sup>Prognostic features are outlined in **BOLD**

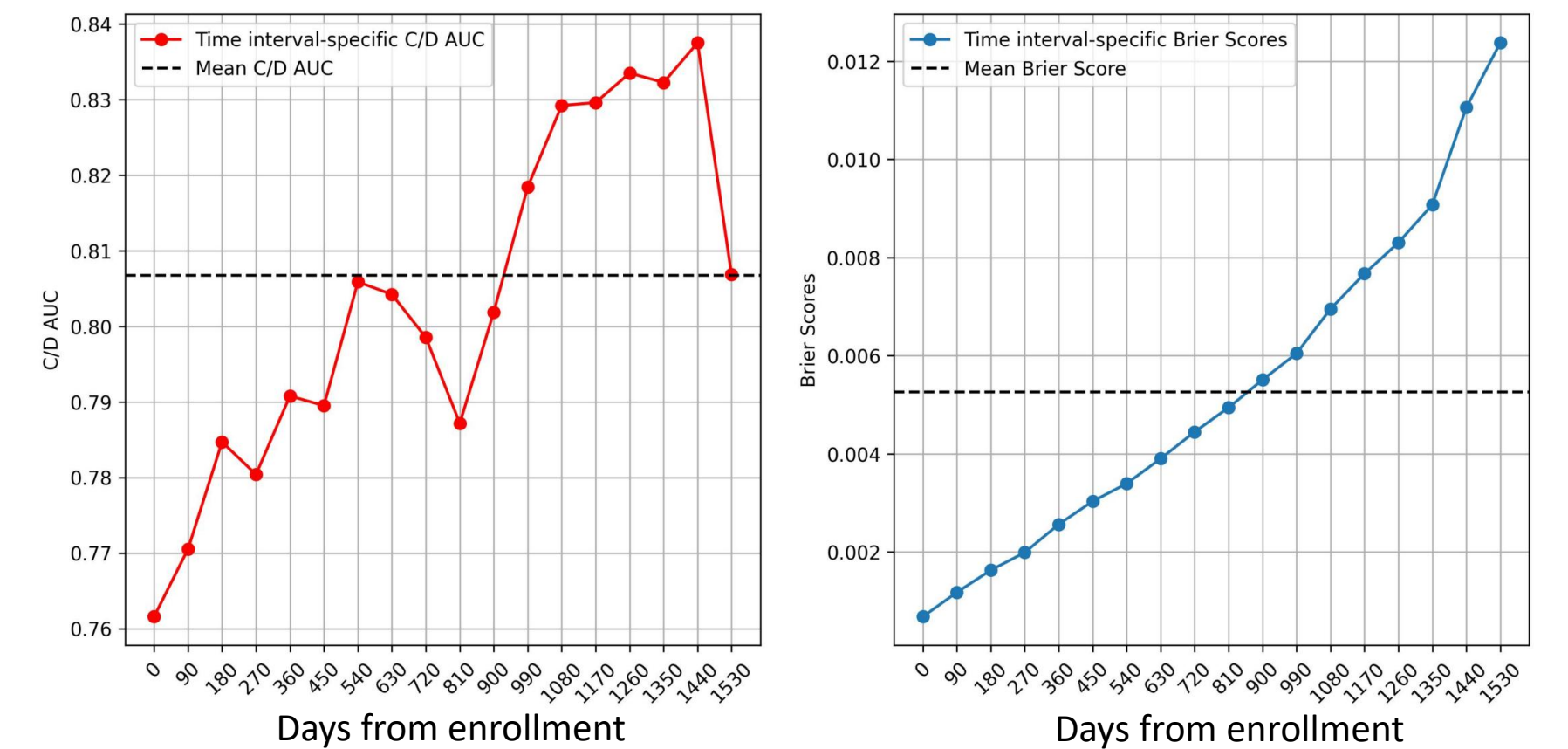
<sup>b</sup>Acute conditions recorded in the 6 months prior to each 90-day time interval

<sup>c</sup>Non-acute conditions recorded anytime prior to the current 90-day time interval

<sup>d</sup>Collected in the 6 months prior to each 90-day time interval

## Prediction of Cannabis Use Disorder within the next 90 days: Time-dependent AUC and Brier scores for the top performing model

### Logistic Regression-trained model with Random Undersampling 1:10



C/D AUC = Cumulative sensitivity/dynamic specificity area under the receiver-operating characteristic

## Prediction of Cannabis Use Disorder within the next 90 days:

### Confusion Matrix for the top performing model

#### Logistic Regression-trained model with Random Undersampling 1:10

		Actual			
		New CUD Diagnosis <sup>a</sup>	No CUD Diagnosis <sup>b</sup>	Total	
Predicted	New CUD Diagnosis	345	8,437	8,782	PPV: 3.93%
	No CUD Diagnosis	109	7,436	7,545	NPV: 98.56%
Total		454	15,873	16,327	

Sensitivity: 75.99%  
Specificity: 46.85%

CUD = Cannabis Use Disorder, NPV = Negative Predictive Value, PPV = Positive Predictive Value

<sup>a</sup>Subjects with an actual new CUD diagnosis during follow-up were labelled as a true positive if they were predicted positive in the time interval the actual new CUD diagnosis was received. Otherwise, actual positive subjects were labelled as false negative.

<sup>b</sup>Subjects with no actual new CUD diagnosis during follow-up were labelled as a false positive if they were predicted positive in ≥1 time interval(s) they remained at risk. Otherwise, actual negative subjects were labelled as true negative.

## CONCLUSION

✓ A discrete-time-updating prediction tool achieved high levels of discrimination for predicting a new CUD diagnosis among Arkansas MMJ cardholders.

✓ The logistic regression-trained model with 1:10 random undersampling achieved the highest discrimination in predicting CUD in the next 90 days (mean C/D AUC = 0.8068) while the support vector machine-trained model achieved the highest calibration (mean Brier Score = 0.0048)

✓ This tool could be used to identify persons authorized to purchase MMJ that are at risk of developing CUD to develop targeted screening and treatment programs.

## References

- National Conference of State Legislatures. State Medical Cannabis Laws. <https://www.ncsl.org/health/state-medical-cannabis-laws>. Published online 2025.
- Center for Behavioral Health Statistics and Quality. 2023 National Survey on Drug Use and Health: Detailed tables. <https://www.samhsa.gov/data/report/2023-nsduh-detailed-tables>. Published online 2023.
- National Academies of Sciences E and M. The Health Effects of Cannabis and Cannabinoids: The Current State of Evidence and Recommendations for Research. National Academies Press; 2017. doi:10.17226/24625
- Arkansas All-Payer Claims Database. Welcome to the Arkansas All-Payer Claims Database (APCD). <https://www.arkansasapcd.net/Home/>.