PRISM: Policy Risk Inference through State Mechanisms

A Mechanism-Aware Alcohol Policy Evidence System for Causal Auditing and Forecasting

Ishaan Ranjan

BASIS Peoria
Senior Project

Advisor: Spencer Kwit

May 2, 2026

Abstract

PRISM is a state-level alcohol policy evidence system that reads policy text, assembles a longitudinal public-health panel, and separates three tasks that are often blended together in policy analytics. The first task is mechanism interpretation, where policy language is scored as price, access, or enforcement. The second is causal auditing around a narrow historical event, the first real-dollar beer-tax increase in a treated state. The third is next-year forecasting under a standardized scenario engine. The project is publicly packaged as the Alcohol Policy Impact Atlas dashboard, but this paper focuses on the research pipeline behind that interface. Using 51 jurisdictions from 2003 through 2023, PRISM links APIS, FARS, FHWA, FRED, and YRBS into a unified evidence base. The strongest benchmark result comes from a semi-supervised feed-forward network that reaches a held-out macro-F1 of 0.962 on a 250-row audited mechanism dataset. The causal audit shows a negative average post-event coefficient of -0.379 deaths per 100k with a pretrend diagnostic of $p = 0.095$, which supports a directional interpretation rather than a settled causal claim. The best next-year crash forecaster is a RandomForestRegressor with RMSE $= 0.863$ and $R^2 = 0.454$ on a 2020–2023 test window. Taken together, the paper argues for a transparent evidence workflow where mechanism scoring improves interpretation, forecasting remains predictive, and limitations stay visible instead of being pushed to the margins.

Keywords: alcohol policy, computational social science, policy text classification, event study, forecasting, public health

1 Introduction

Alcohol-impaired driving remains a persistent public-safety problem in the United States. The National Highway Traffic Safety Administration reported 12,429 alcohol-impaired-driving traffic deaths in 2023, which was close to one third of all traffic fatalities that year (National Highway Traffic Safety Administration, 2025). State governments respond with taxes, sales restrictions, licensing rules, underage-purchase penalties, and related interventions. Those policies do not work through the same behavioral channel. A beer excise tax changes price. A Sunday-sales restriction changes availability. A retailer penalty changes deterrence and compliance pressure.

In this paper, a *mechanism* means the path through which a policy could change behavior. The label is simple on purpose. It helps distinguish whether a law is trying to make alcohol more expensive, harder to obtain, or riskier to sell or misuse.

That distinction matters, yet many state-level policy analyses still flatten legal change into a yes-or-no indicator or a broad composite score. Those representations are useful for some questions, though they lose much of the literal policy content that lawmakers actually control. A mechanism-aware system should preserve more of that structure while still making room for careful causal language and honest forecasting.

This paper presents **PRISM**, short for *Policy Risk Inference through State Mechanisms*. PRISM is a transparent evidence system. It keeps text interpretation, causal auditing, and forecasting in the same workflow without forcing them into one blended claim. The system separates three tasks:

- mechanism interpretation from policy text
- causal auditing around a narrow historical event
- next-year forecasting under standardized scenario inputs

The public dashboard version of the project packages these outputs under the name *Alcohol Policy Impact Atlas*. The dashboard is useful for exploration, though the paper stays closer to the underlying research design, evaluation logic, and limitations.

A lay reader can think of PRISM as a careful translator between three layers of evidence. First, it reads policy language and asks what kind of lever the law is pulling. Second, it checks one clean historical event type to see whether the surrounding crash pattern moves in the expected direction. Third, it asks what a predictive model does with the latest state data when it is pushed through a standardized scenario.

The project makes four concrete contributions. First, it assembles a 51-jurisdiction state-year panel from APIS, FARS, FHWA, FRED, and YRBS, covering 2003–2023 for the main crash analysis. Second, it builds a 250-row audited benchmark for classifying policy language into price, access, and enforcement mechanisms. Third, it runs a cautious event-study audit around first real-dollar beer-tax increases. Fourth, it compares held-out forecasting models and attaches a scenario layer that keeps predictive outputs separate from causal claims.

The paper follows that structure. Section 2 situates PRISM in the alcohol-policy and policy-text literature. Section 3 describes the panel and the way outcomes, covariates, and text features are constructed. Section 4 presents the causal audit, mechanism benchmark, and forecasting setup. Section 5 reports the main findings. Sections 6 to 8 interpret those findings without stretching them past what the data can support.

2 Related Work

PRISM sits at the intersection of alcohol-policy evaluation, causal panel analysis, and domain-specific text classification. The project does not claim to replace any of those traditions. It tries to connect them in a way that preserves mechanism detail and keeps inferential boundaries visible.

Alcohol tax and price policies have one of the strongest empirical records in this literature. A major review by Wagenaar et al. found that higher alcohol taxes and prices were associated with lower morbidity and mortality, including traffic harms (Wagenaar, Tobler, & Komro, 2010). Carpenter and Dobkin also show that alcohol regulation has meaningful public-health consequences around drinking-age policy and harm exposure (Carpenter & Dobkin, 2011). More recent work by Naimi et al. suggests that stronger overall state policy environments are associated with lower odds that a fatal crash is alcohol-related (Naimi et al., 2018). These studies support the broad idea that policy design matters, but they do not usually build mechanism scores directly from the text of the law.

Access restrictions and enforcement interventions follow different behavioral pathways. Popova et al. review the evidence on outlet density, hours of sale, and days of sale, concluding that changes in physical availability can affect alcohol-related harm (Popova, Giesbrecht, Bekmuradov, & Patra, 2009). Bergen et al. report that publicized sobriety checkpoint programs are associated with lower alcohol-involved crash fatalities, which is consistent with deterrence-based mechanisms (Bergen et al., 2014). This split between price, access, and enforcement is not exhaustive, though it is a practical way to organize the policy levers most visible in the PRISM corpus.

The causal component of PRISM also draws from modern event-study practice. Staggered treatment timing can create misleading averages when treatment effects are heterogeneous or timing varies widely (Goodman-Bacon, 2021; Sun & Abraham, 2021). That issue is one reason this paper keeps the audit narrow. Instead of treating all alcohol-law changes as interchangeable, the causal module

Table 1: Core panel coverage

Metric	Value
Jurisdictions	51
Crash and policy window	2003–2023
State-year rows	1,071
Teen-wave rows	867
Fully observed teen rows	141
Fully observed teen share	16.3%

Note. The main PRISM panel is state-year based. Teen outcomes are stored in a separate wave-based panel because YRBS participation is uneven across states and survey years.

studies a single event type with relatively clear timing: the first real-dollar beer-tax increase in a treated state. The resulting estimate is still limited by the small number of events, so the payoff here is discipline rather than broad causal reach.

On the text side, legal NLP has benefited from domain-adapted representations such as Legal-BERT (Chalkidis, Fergadiotis, Malakasiotis, Aletras, & Androutsopoulos, 2020). Those models provide a strong baseline for statute-like language, and they are included in the PRISM benchmark. Still, policy text in this project is short, repetitive, and mechanism-focused. That makes the task different from long-form legal judgment classification. It is plausible that a smaller model with domain-specific features could perform very well, especially when the benchmark is constrained and audited. One of the more useful findings in PRISM is that this is exactly what happens.

3 Data and Panel Construction

PRISM combines five public data sources into a unified policy-and-outcomes panel. APIS tracks state alcohol law timing and policy summaries. FARS records the fatal crashes used for the main outcome. FHWA contributes vehicle-miles-traveled totals so that raw crash counts can be read against driving exposure. FRED adds economic context, including unemployment, personal income, and population. YRBS contributes survey-based teen current-use and binge-drinking outcomes, which are retained as a secondary module because coverage is sparse (National Institute on Alcohol Abuse and Alcoholism, n.d.; National Highway Traffic Safety Administration, n.d.; Federal Highway Administration, n.d.; Federal Reserve Bank of St. Louis, n.d.; Centers for Disease Control and Prevention, n.d.).

The main crash panel covers 51 jurisdictions from 2003 through 2023, yielding 1,071 state-year rows. One row represents one jurisdiction in one year. The primary outcome is alcohol-impaired fatality rate per 100k population. This is the outcome that anchors the causal audit, the model comparison, and the scenario engine. Alcohol-involved fatality rate and teen alcohol-use measures are retained for supporting analyses and public communication, but they do not carry the main claim of the paper.

Teen outcomes are handled more carefully. The separate teen-wave panel contains 867 rows, but only 141 rows are fully observed. That is 16.3% of the teen dataset. The remainder requires either partial fill logic or imputed values inherited from the broader project pipeline. For that reason, teen results are informative context rather than the center of the final paper.

The panel also carries policy-text features that are constructed outside the main regression and

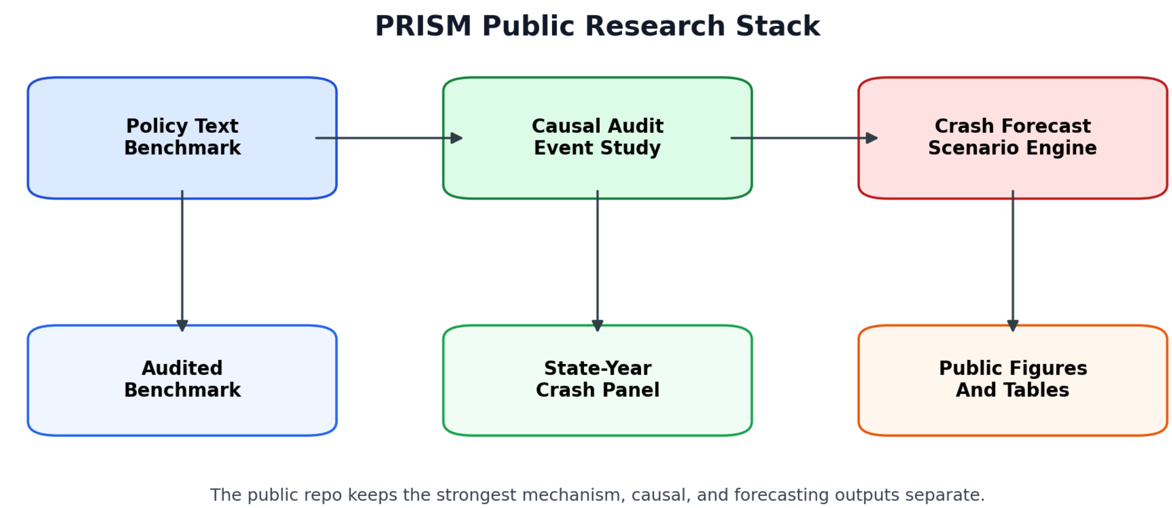

Figure 1: PRISM architecture. Data integration, causal auditing, mechanism interpretation, and forecasting are kept as separate modules so that each task can be judged on its own terms.

forecasting loops. Each state-year receives mechanism scores for price, access, and enforcement, along with provenance and text-quality metadata. Those fields are helpful in two ways. They let the benchmark results feed into downstream interpretation, and they also make it possible to flag when the text layer is thin or heavily supplemental.

That setup creates a useful compromise. The panel is broad enough to compare states over time, but each state-year record is still interpretable by a human reader. The goal is not to create a black-box national index. The goal is to keep the unit of analysis simple enough that a policy change, a crash outcome, and a mechanism score can still sit next to each other in the same row.

That provenance matters. In the final state-year artifact, 936 of 1,071 rows, or 87.4%, rely on supplemental-only policy text coverage. The remaining 135 rows are marked mixed. The benchmark itself is better curated than the full state-year text layer, but the production artifact still inherits these coverage limits. A transparent paper should say that early, because it shapes how the mechanism results should be read.

4 Methods

4.1 Task Separation

PRISM is organized around three distinct tasks. The first task interprets policy text. The second audits a historical event for directional causal evidence. The third predicts next-year risk under observed and counterfactual-like scenario inputs. This separation is intentional. It keeps a strong benchmark result in one module from being mistaken for causal proof in another.

4.2 Outcome and Exposure Definitions

The crash outcome is defined at the state-year level. Population is measured in persons, and vehicle miles traveled are used to normalize for driving exposure:

$$\text{Pop}_t = \text{population\_thousands}_t \times 1000 \tag{1}$$

$$\text{VMTPerCapita}_{s,t} = \frac{\text{VMT}_{s,t}}{\text{Pop}_{s,t}} \tag{2}$$

The primary outcome is alcohol-impaired fatality rate per 100k population:

$$\text{ImpairedRate}_{s,t} = \left(\frac{\text{impaired\_fatalities}_{s,t}}{\text{Pop}_{s,t}}\right) \times 100000 \tag{3}$$

This rate is used in both the causal audit and the forecasting task. Alcohol-involved fatality rate is retained for supporting context. Teen current-use and binge-drinking outcomes are maintained in a separate panel because YRBS reporting is irregular across states and years.

4.3 Causal Audit

The causal module studies the first year in which a treated state increases its beer excise tax in real dollars. There are 10 such events in the final analysis. Event time is indexed relative to that first increase, using a five-year lead and five-year lag window with period -1 omitted as the reference category. The estimating equation is

$$Y_{s,t} = \alpha_s + \lambda_t + \sum_{k \neq -1} \beta_k \mathbf{1}[\text{EventTime}_{s,t} = k] + \gamma' X_{s,t} + \varepsilon_{s,t}, \tag{4}$$

where $Y_{s,t}$ is alcohol-impaired fatality rate per 100k, α_s and λ_t are state and year fixed effects, and $X_{s,t}$ includes unemployment, personal income, and VMT per capita. Standard errors are clustered at the state level.

In plain terms, the event study asks whether treated states bend away from their earlier path after the first real-dollar beer-tax increase, once national year shocks and state-level baselines are absorbed. It is a disciplined before-and-after comparison with controls, not a full natural experiment that wipes away every competing explanation.

Two diagnostics are emphasized. First, the pretrend test checks whether the lead coefficients are jointly close to zero. Second, a placebo routine from the project workflow tests whether randomly assigned event timing produces similar movement. These checks do not turn a small design into a large one, though they help bound how confidently the result should be framed.

4.4 Mechanism Benchmark

The text pipeline starts from policy snippets tied to beer-tax, Sunday-sales, and underage-purchase topics. After chunking and deduplication, the v3 corpus contains 10,136 chunk occurrences and 567 unique chunks. Each chunk is assigned one or more mechanism labels from the following set:

- **Price**: taxes, pricing rules, and direct cost signals

- **Access**: sales timing, outlet availability, and purchasing access
- **Enforcement**: penalties, compliance checks, and deterrence language

Chunking matters because policy summaries often bundle several ideas into one paragraph. Shorter units make it easier to ask a narrow question of each passage. Is this sentence mainly about price, mainly about access, or mainly about enforcement?

The evaluation set is a 250-row audited benchmark. Production model selection is based on held-out macro-F1 on that benchmark rather than downstream crash RMSE. For a chunk i and mechanism m, the classifier outputs

$$p_{i,m} = \sigma(f_\theta(x_i))_m \tag{5}$$

$$\hat{m}_i = \arg\max_m p_{i,m} \tag{6}$$

and the weighted optimization target is

$$\theta^* = \arg\min_\theta \sum_i w_i \cdot \mathrm{BCE}(y_i, p_i). \tag{7}$$

The winning model is a semi-supervised feed-forward network built on compressed TF-IDF, lexicon, and topic features. Pseudo-labeled rows are used at lower sample weight. The benchmark also includes keyword, TF-IDF logistic, DistilGPT-2 linear, Legal-BERT FFN, and Legal-BERT BiGRU baselines.

4.5 Forecasting and Scenarios

The forecasting task predicts next-year alcohol-impaired fatality rate from current-year policy, mechanism, economic, and lagged outcome features. Evaluation is strictly out of time:

- train: 2003–2016
- validate: 2017–2019
- test: 2020–2023

For a general reader, the forecasting task is the simplest module to picture. The model looks at a state's latest observed profile and then guesses the following year's alcohol-impaired fatality rate. It is judged only on future years that were held out from training.

The model ladder includes ElasticNet, RandomForestRegressor (Breiman, 2001), HistGradient-BoostingRegressor, MLPRegressor, and a residual text model that uses the mechanism layer more explicitly. Performance is compared with RMSE, MAE, and R^2 on held-out years.

The scenario engine applies a standardized shift to selected policy features and reports the difference between a baseline and scenario forecast:

$$\Delta_s = \widehat{Y}_s^{\text{scenario}} - \widehat{Y}_s^{\text{baseline}}. \tag{8}$$

Approximate interval bounds are carried over from residual dispersion in the forecasting pipeline:

$$\widehat{Y}_s^{\text{scenario}} \pm 1.96\sigma_{\text{resid}}. \tag{9}$$

These deltas are predictive summaries for comparison across states. They are useful for planning and interpretation, especially in the dashboard, but they should not be read as treatment-effect estimates.

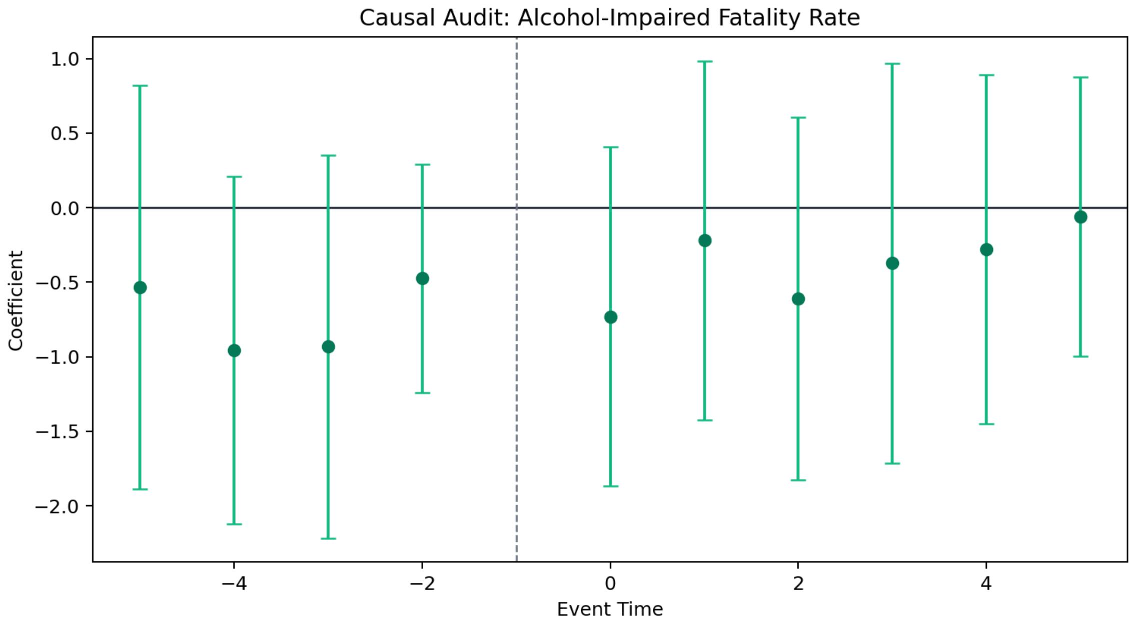

Figure 2: Event-study estimates around first real-dollar beer-tax increases. The post-event pattern is mostly negative, though several confidence intervals still cross zero.

Table 2: Mechanism benchmark comparison on the 250-row audited dataset

Model	Macro-F1	Access F1	Enforcement F1	Price F1
Semi-supervised FFN	0.962	0.949	1.000	0.938
TF-IDF logistic	0.846	0.850	0.786	0.903
Legal-BERT FFN	0.801	0.771	0.759	0.875
Legal-BERT BiGRU	0.796	0.725	0.815	0.848
DistilGPT-2 linear	0.735	0.768	0.700	0.737
Keyword baseline	0.194	0.000	0.000	0.583

Note. The selected production scorer is the semi-supervised FFN. Model selection is based on held-out macro-F1 rather than downstream crash forecasting.

5 Results

5.1 Causal Audit

The event-study profile in Figure 2 trends downward after treatment. Averaging the post-event coefficients yields -0.379 deaths per 100k. In practical terms, the fitted post-event years sit about 0.379 deaths per 100k below the omitted pre-event reference period, conditional on the fixed effects and controls in the model. The pretrend diagnostic is $p = 0.095$. That is consistent with a directional reduction in alcohol-impaired fatality risk after beer-tax increases, but the paper does not treat it as definitive proof. The small number of treated events and the width of the intervals argue for restraint. The full coefficient table appears in Table 5.

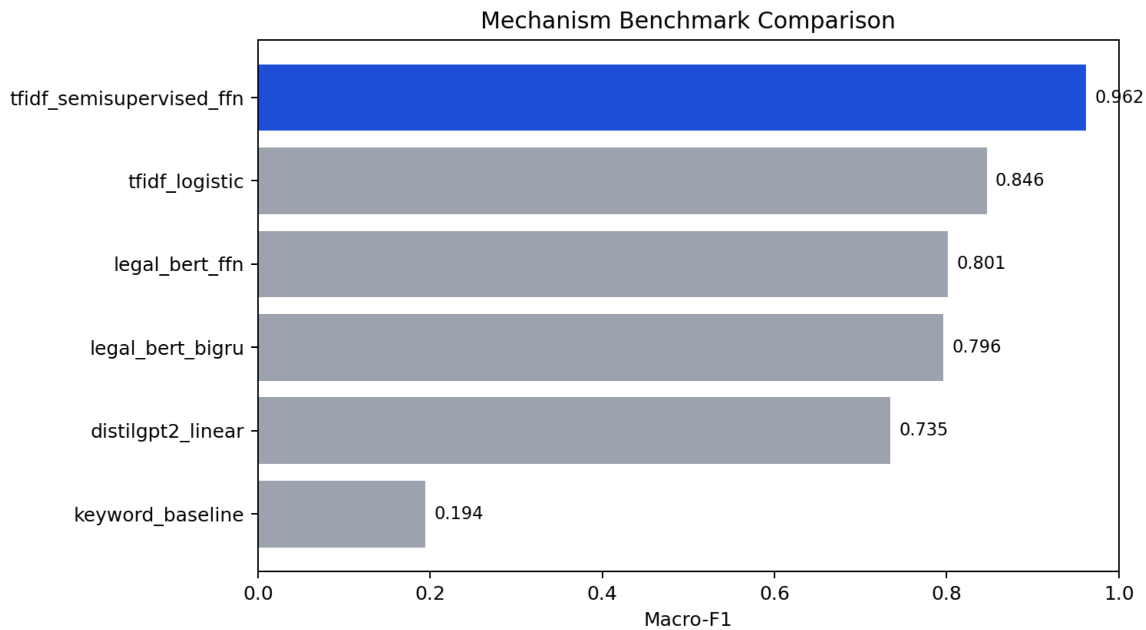

Figure 3: Mechanism benchmark ladder. The semi-supervised FFN separates price, access, and enforcement more reliably than the other tested models.

5.2 Mechanism Benchmark

The mechanism benchmark is the clearest technical result in the project. The underlying corpus contains 10,136 chunk occurrences and 567 unique chunks, with formal evaluation carried out on a 250-row audited benchmark. The selected semi-supervised FFN reaches macro-F1 = 0.962. It outperforms the simpler keyword baseline by a large margin and also beats the transformer baselines included in the ladder. In practical terms, that means the model can classify short policy chunks into price, access, and enforcement with high reliability on held-out audited rows. It also means the performance is balanced across classes instead of being driven by only one easy label.

This does not solve every text problem in the pipeline. The benchmark itself still includes supplemental material, and production coverage across all state-years remains heavily supplemental. Even so, the benchmark is strong enough to support a useful interpretation layer for the downstream forecasting and dashboard modules.

5.3 Forecasting and Scenario Outputs

The held-out forecasting leaderboard tells a different story from the benchmark ladder. The best crash forecaster is a RandomForestRegressor with RMSE = 0.863 and $R^2 = 0.454$ on the 2020–2023 test window. HistGradientBoosting is second. The residual text model finishes well behind the leading tabular baselines.

That result matters because it prevents a comfortable narrative shortcut. The mechanism-aware text layer is valuable, but its value shows up most clearly in interpretation and scenario decomposition. Raw next-year crash forecasting still favors a standard tree ensemble on this panel.

The standardized scenario engine predicts mostly downward shifts in next-year alcohol-impaired

Table 3: Held-out crash forecasting comparison

Model	Val. RMSE	Test RMSE	Test MAE	R^2
Random Forest	0.864	0.863	0.658	0.454
Hist. Gradient Boosting	1.487	1.066	0.746	0.166
MLP Regressor	1.234	1.188	0.861	−0.034
Residual Text Model V3	1.657	1.701	1.135	−1.123
Elastic Net	1.173	1.786	1.432	−1.339

Note. All models are trained on 2003–2016, tuned on 2017–2019, and tested on 2020–2023. Lower RMSE and MAE are better. Higher R^2 is better.

fatality risk. The largest modeled decreases occur in West Virginia, Iowa, and Ohio, followed closely by Michigan and Wisconsin. For a reader coming from the dashboard, this is the right way to interpret the module. It asks how the trained model responds when the same stylized tightening is applied across states, then reports which states move more or less. Those deltas are useful for ranking model response under a common input. They should be read as predictive movement, not as direct policy-effect estimates. A longer scenario excerpt is included in Table 8.

6 Discussion

PRISM works best when its three modules are allowed to keep their own jobs. The causal audit gives a historically grounded directional signal around a narrow event. The benchmark gives a strong answer to a focused text-classification problem. The forecasting layer produces out-of-time risk estimates and scenario movement. Once those tasks are merged into a single claim, the project becomes much easier to overstate.

A reader should come away with three separate conclusions. First, PRISM can read short policy text surprisingly well on a curated benchmark. Second, the historical beer-tax audit points in a plausible direction, though it stays short of a settled causal estimate. Third, the dashboard scenario layer is best treated as a planning aid tied to the fitted model, not as an automatic answer about what a legislature would accomplish in practice.

The mechanism layer adds value even without winning the forecasting contest. It gives each state-year a more interpretable policy profile, which is useful for public explanation and for the dashboard interface. A policymaker or judge can see whether a state-year is dominated by price, access, or enforcement language rather than being handed a binary "policy changed" flag and little else.

The forecasting results are also scientifically useful in a quieter way. A standard Random Forest beats the text-augmented residual model on held-out crash prediction. That boundary is useful on its own because it marks what the current mechanism layer can and cannot deliver. Better explanation does not automatically yield better next-year prediction on a modest, noisy, state-year panel.

The causal result fits the broader literature well enough to be taken seriously, but not strongly enough to be treated as settled. The post-event pattern is mostly negative. The pretrend diagnostic is close to conventional significance cutoffs without fully clearing them. In a project that tries to be transparent, the correct response is to keep the wording cautious and let the result stay directional.

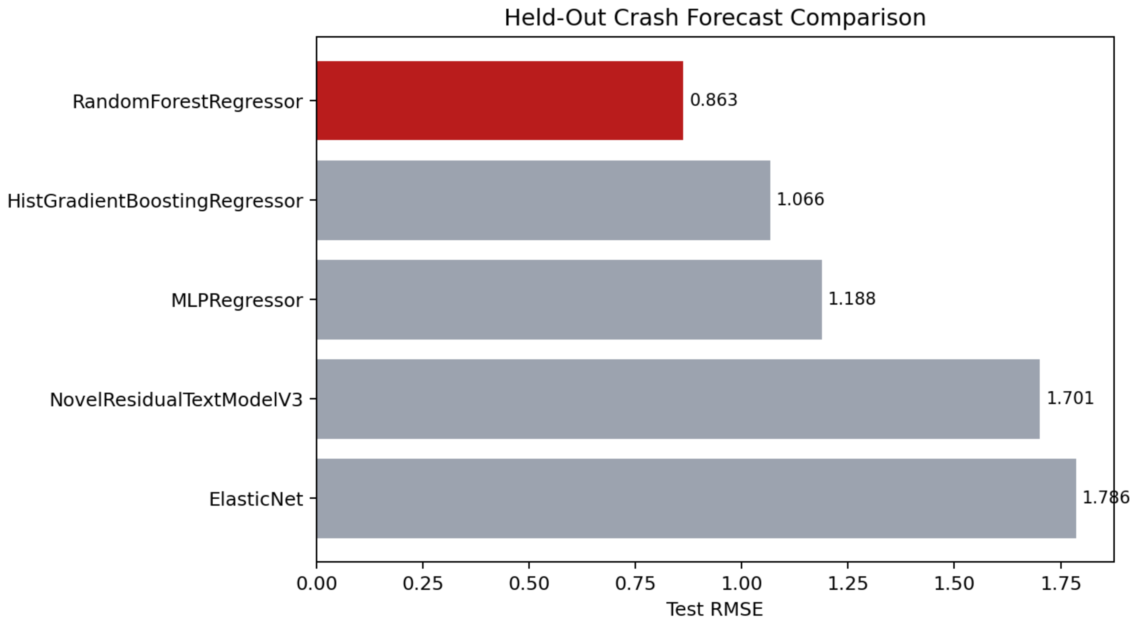

Figure 4: Held-out crash forecasting comparison. The best next-year crash forecaster is a RandomForestRegressor rather than the text-augmented residual model.

7 Limitations

Four limits shape how this paper should be read.

First, teen-outcome coverage is sparse. Only 141 of 867 teen-wave rows are fully observed. That makes teen results secondary evidence, even when the patterns are interesting.

Second, the production text layer relies heavily on supplemental coverage. In the final state-year mechanism artifact, 87.4% of rows are marked supplemental-only and the remainder are mixed. This does not invalidate the benchmark, though it does limit how confidently the state-year mechanism scores can be treated as direct readings of official statute text.

Third, the causal audit is based on a small number of clean beer-tax events. That design was chosen for timing clarity, not for breadth. It supports a focused historical check and little more.

Fourth, scenario outputs remain predictive. They tell us how the fitted model moves when standardized inputs change. They do not guarantee what would happen after a real law passes, implementation changes, enforcement shifts, or unrelated social conditions move at the same time.

8 Conclusion

PRISM shows that a high-school senior project can produce a serious policy evidence workflow without pretending to solve every inferential problem at once. The paper assembles a multi-source state-year panel, builds a strong audited mechanism benchmark, audits a narrow historical event, and evaluates next-year forecasting on held-out years.

The strongest results are clear. Mechanism-aware policy text can be classified with high accuracy

Table 4: Largest predicted decreases under the standardized scenario

State	Baseline	Scenario	Δ
WV	3.041	2.758	−0.283
IA	2.759	2.482	−0.277
OH	2.677	2.459	−0.218
MI	2.247	2.030	−0.217
WI	2.187	1.988	−0.199
TX	2.441	2.251	−0.190

Note. These deltas are predictive forecast differences under a standardized tightening scenario. They should be read as model responses rather than treatment effects.

on an audited benchmark. Beer-tax increases are followed by a mostly negative post-event pattern in alcohol-impaired fatality risk. The best crash forecaster is still a Random Forest. Those findings fit together into a useful evidence package for interpretation and cautious planning.

Future work should push the text layer closer to official statute sources, broaden the set of policy events studied causally, and test whether finer geographic units preserve the same mechanism story. For the current project, the main contribution is simpler: PRISM offers a transparent way to keep policy language, historical auditing, and forecasting in the same room without forcing them to say the same thing.

A Appendix Tables and Provenance Notes

This appendix collects the extended tables referenced in the main text and records the provenance decisions used in the final paper package.

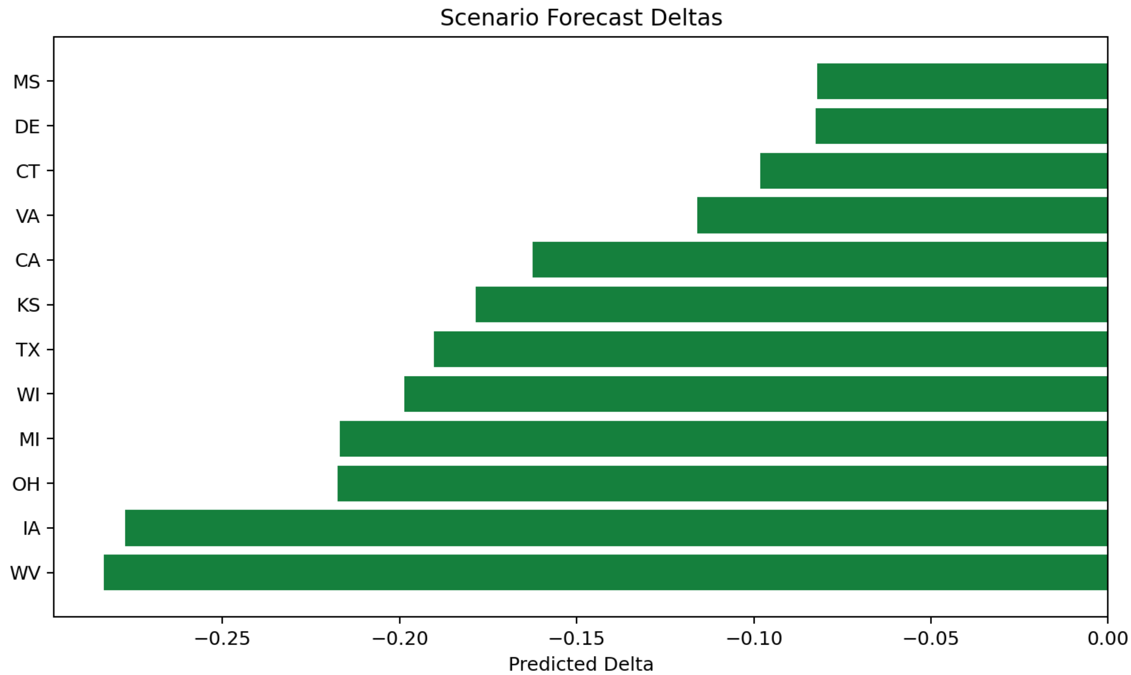

Figure 5: State-specific forecast deltas under a standardized tightening scenario. Most modeled responses move downward, though the size of the shift varies substantially across states.

Table 5: Primary event-study coefficients for alcohol-impaired fatality rate

Event time	Effect	CI low	CI high	p
-5	-0.533	-1.885	0.819	0.440
-4	-0.955	-2.121	0.211	0.108
-3	-0.931	-2.216	0.354	0.155
-2	-0.474	-1.239	0.291	0.225
0	-0.730	-1.865	0.406	0.208
1	-0.220	-1.425	0.984	0.720
2	-0.610	-1.825	0.604	0.325
3	-0.373	-1.713	0.967	0.585
4	-0.278	-1.450	0.893	0.641
5	-0.060	-0.996	0.876	0.900

Note. The shared pretrend diagnostic for this outcome is $p = 0.095$. Period -1 is the omitted reference category.

Table 6: Full crash model leaderboard

Model	Val. RMSE	Test RMSE	Test MAE	R^2	Train	Val.	Test
Random Forest	0.864	0.863	0.658	0.454	714	153	153
Hist. Gradient Boosting	1.487	1.066	0.746	0.166	714	153	153
MLP Regressor	1.234	1.188	0.861	−0.034	714	153	153
Residual Text Model V3	1.657	1.701	1.135	−1.123	714	153	153
Elastic Net	1.173	1.786	1.432	−1.339	714	153	153

Table 7: Teen-outcome coverage in the archived wave-based panel

Source	Coverage flag	Rows	Share (%)
YRBS Explorer	imputed	331	38.200
YRBS Explorer	observed	106	12.200
YRBS Explorer	observed plus imputed	12	1.400
YRBS Socrata	observed	35	4.000
YRBS Socrata	observed plus imputed	383	44.200

Note. Fully observed rows total 141 of 867, or 16.3%. This is why teen outcomes are retained as secondary evidence in the main paper.

Table 8: Scenario-response excerpt with interval bounds

State	Baseline	Scenario	Δ	Low	High
WV	3.041	2.758	−0.283	0.037	5.478
IA	2.759	2.482	−0.277	−0.239	5.203
OH	2.677	2.459	−0.218	−0.261	5.180
MI	2.247	2.030	−0.217	−0.691	4.751
WI	2.187	1.988	−0.199	−0.732	4.709
TX	2.441	2.251	−0.190	−0.470	4.971
KS	2.618	2.440	−0.178	−0.281	5.160
CA	2.271	2.109	−0.163	−0.612	4.829
VA	2.483	2.367	−0.116	−0.354	5.088
CT	2.441	2.343	−0.098	−0.378	5.064

Note. Interval bounds are approximate prediction intervals inherited from residual dispersion in the scenario engine.

A.1 Figure and Data Provenance

The five main figures in the paper were copied into the local `figures/` folder from the frozen public artifacts under `projectv3/results/figures`. Headline benchmark, event-study, crash-model, and scenario tables were transcribed from the frozen public CSVs in `projectv3/results/tables`. Teen coverage and state-year text-provenance counts were taken from the archived v2 outputs and the generated dashboard data files that mirror those frozen artifacts. The paper is therefore self-contained at compile time while still tracing back to fixed project outputs rather than a live analysis run.

For transparency, the key provenance counts used in the paper are:

- benchmark source mix: 162 supplemental-only rows and 88 direct-APIS rows
- state-year mechanism provenance: 936 supplemental-only rows and 135 mixed rows
- fully observed teen rows: 141 of 867

References

Bergen, G., Pitan, A., Qu, S., Shults, R. A., Chattopadhyay, S. K., Elder, R. W., … Force, C. P. S. T. (2014). Publicized sobriety checkpoint programs: A community guide systematic review. *American Journal of Preventive Medicine*, *46*(5), 529–539. doi: 10.1016/j.amepre.2014.01.018

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Carpenter, C., & Dobkin, C. (2011). The minimum legal drinking age and public health. *Journal of Economic Perspectives*, *25*(2), 133–156.

Centers for Disease Control and Prevention. (n.d.). *Youth risk behavior surveillance system (YRBSS).* Retrieved from https://www.cdc.gov/yrbs/ (Accessed February 14, 2026)

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2898–2904. doi: 10.18653/v1/2020.findings-emnlp.261

Federal Highway Administration. (n.d.). *Highway statistics series.* Retrieved from https://www.fhwa.dot.gov/policyinformation/statistics/ (Accessed February 27, 2026)

Federal Reserve Bank of St. Louis. (n.d.). *FRED economic data.* Retrieved from https://fred.stlouisfed.org/ (Accessed March 22, 2026)

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Naimi, T. S., Xuan, Z., Sarda, V., Hadland, S. E., Lira, M. C., Swahn, M. H., … Heeren, T. C. (2018). Association of state alcohol policies with alcohol-related motor vehicle crash fatalities among U.S. adults. *JAMA Internal Medicine*, *178*(7), 894–901. doi: 10.1001/jamainternmed.2018.1406

National Highway Traffic Safety Administration. (2025, December). *Nhtsa launches annual Drive Sober or Get Pulled Over enforcement campaign for holiday season.* Retrieved from https://www.nhtsa.gov/press-releases/nhtsa-launches-annual-drive-sober-or-get-pulled-over-enforcement-campaign-holiday (Accessed March 11, 2026)

National Highway Traffic Safety Administration. (n.d.). *Fatality analysis reporting system (FARS).* Retrieved from https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars (Accessed March 4, 2026)

National Institute on Alcohol Abuse and Alcoholism. (n.d.). *Alcohol policy information system (APIS).* Retrieved from https://alcoholpolicy.niaaa.nih.gov/ (Accessed February 19, 2026)

Popova, S., Giesbrecht, N., Bekmuradov, D., & Patra, J. (2009). Hours and days of sale and density of alcohol outlets: Impacts on alcohol consumption and damage: A systematic review. *Alcohol and Alcoholism*, *44*(5), 500–516. doi: 10.1093/alcalc/agp054

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199.

Wagenaar, A. C., Tobler, A. L., & Komro, K. A. (2010). Effects of alcohol tax and price policies on morbidity and mortality: A systematic review. *American Journal of Public Health*, *100*(11), 2270–2278. doi: 10.2105/AJPH.2009.186007