Predicting Opioid Cost Burden through Integrated PBM and SDOH Modeling: An Explainable AI Framework

Author:
Mohit Singhal

Affiliation:
Independent Researcher

Date:
November 9th, 2025

1. Abstract

The opioid epidemic continues to impose a profound human and financial burden on the United States health system. While previous studies have explored either prescribing behavior or social determinants of health (SDOH) in isolation, few have quantitatively integrated these domains with pharmacy-benefit-management (PBM) cost levers to explain geographic variation in opioid-related spending. This study develops an interpretable, county-level machine-learning framework that links Medicare Part D opioid-prescribing data (2013 – 2023) with socioeconomic, behavioral, and healthcare-access indicators from the County Health Rankings dataset (University of Wisconsin Population Health Institute, 2024).

Using a Random Forest regression pipeline with automated feature selection and five-fold cross-validation, the final model achieved an $R^2 = 0.9698$ and RMSE = \$445.36, explaining nearly all observed variation in opioid cost per capita while preserving generalizability across counties. Comparative analysis showed that SDOH-only models explained 34 % of cost variance, utilization-only models 95 %, and the integrated PBM + SDOH framework 97 %, confirming that economic, behavioral, and access factors jointly shape fiscal burden. Feature-importance and SHAP analyses identified cost per claim, claims per 1 000 population, and opioid-prescribing rate as dominant drivers, amplified by structural vulnerabilities such as unemployment, income inequality, and low mental-health-provider density.

Counterfactual simulations quantified the potential policy impact of PBM and SDOH interventions. PBM levers—including formulary tightening, utilization guidance, and cost-per-claim reduction—generated short-term savings of ≈ 4–6 %, while broader SDOH improvements, such as expanded primary-care and mental-health access, yielded compounding reductions up to 17 %, equivalent to roughly $23 billion in annual national savings. Geographic validation confirmed alignment with CDC dispensing-rate patterns, with high-burden counties concentrated in Utah, Kentucky, and Alabama.

This integrated analytical framework demonstrates how combining PBM program design with SDOH modeling can inform equitable policy, optimize healthcare spending, and guide targeted interventions in the continuing U.S. opioid crisis.

Keywords: *opioid prescribing · social determinants of health · pharmacy benefit management · Medicare Part D · machine learning · public health policy*

2. Introduction

The opioid epidemic remains one of the most complex and costly public-health and economic crises in the United States. According to the Centers for Disease Control and Prevention (CDC, 2024), prescription opioid misuse has contributed to more than 280 000 deaths since 1999 and generates over $78 billion annually in healthcare expenditures, productivity losses, and criminal-justice costs. Traditional studies have primarily focused on clinical prescribing trends or patient-level risk factors, yet they have often neglected the structural financial mechanisms—particularly those administered by Pharmacy Benefit Managers (PBMs)—that influence opioid pricing, utilization, and reimbursement across the Medicare Part D landscape.

PBMs play a pivotal intermediary role in determining formulary composition, negotiating manufacturer rebates, and implementing utilization-management controls for millions of beneficiaries. In response to the opioid crisis, many PBMs introduced stricter opioid policies such as dose and day-supply limits, prior-authorization requirements, and formulary exclusions. However, empirical evidence directly linking PBM policy interventions to population-level opioid cost outcomes remains scarce (Seeherman et al., 2021; Witkowski & Chamberlain, 2022). Most prior evaluations have examined claim- or patient-level effects, without quantifying how PBM strategies reshape regional fiscal burden.

In parallel, a growing body of literature has demonstrated that Social Determinants of Health (SDOH)—including unemployment, income inequality, and provider availability—drive spatial heterogeneity in opioid use, treatment access, and overdose risk (Keyes et al., 2020; Monnat, 2019; Dasgupta et al., 2018). These community-level factors interact with healthcare markets to influence not only prescribing behavior but also economic vulnerability, especially in rural and low-income areas. Yet few studies have jointly modeled SDOH alongside the financial instruments of PBM policy within a single analytic framework.

This study integrates PBM cost-management levers with SDOH contextual variables to construct a unified, explainable machine-learning model of opioid cost per capita across all U.S. counties (2013–2023). Using publicly available Centers for Medicare & Medicaid Services (CMS) Medicare Part D and County Health Rankings data, the model quantifies how clinical utilization, drug-price structures, and socioeconomic context collectively shape opioid spending.

A Random Forest regression approach captures nonlinear and interactive effects that linear models cannot, achieving out-of-sample predictive accuracy of $R^2 = 0.9698$ (RMSE ≈ \$445). Beyond prediction, the framework performs counterfactual simulations to estimate the fiscal impact of PBM and SDOH policy scenarios. PBM interventions—formulary tightening, utilization guidance, and cost-per-claim reductions—represent system-internal levers that can generate immediate savings. Complementary SDOH improvements—expanding mental-health and primary-care access, or reducing behavioral risk factors such as smoking and obesity—model community-based levers with long-term compounding benefits. Together, these scenarios demonstrate how integrated fiscal-social strategies can yield meaningful national cost reductions without compromising model transparency.

Whereas prior machine-learning research has concentrated on overdose or misuse prediction (e.g., JAMA Network Open, 2023; PLOS Digital Health, 2022; Health Affairs, 2021), this study extends the field by centering on economic forecasting and policy optimization. It presents the first county-level model to merge PBM program levers with SDOH indicators over a ten-year period, translating predictive analytics into actionable fiscal insights. By shifting the analytic focus from risk prediction to cost mitigation, this work establishes a reproducible, interpretable foundation for policymakers, PBMs, and public-health stakeholders seeking to reduce the opioid crisis's financial burden through data-driven, cross-sector interventions.

3. Methods

3.1.Data and Study Design

A longitudinal, county-year analytic panel was constructed spanning 2013–2023 to examine how pharmacy-benefit-management (PBM) mechanisms and social determinants of health (SDOH) jointly influence U.S. opioid-related spending. Three publicly available data sources were integrated:

- Centers for Medicare & Medicaid Services (CMS) Medicare Part D Opioid Prescribing Data, including the Prescribers by Provider and Prescribers by Geography and Drug files, which contain annual counts of total and opioid-specific claims, prescribers, and drug costs.
- County Health Rankings (University of Wisconsin Population Health Institute, 2024), providing socioeconomic and behavioral-health indicators such as unemployment, income inequality (income ratio), obesity, smoking prevalence, and the density of mental-health and primary-care providers.
- U.S. Census County Population Estimates, used to express all utilization and cost outcomes on a per-capita basis.

Datasets were harmonized using 5-digit Federal Information Processing System (FIPS) codes and calendar year, producing a balanced panel of approximately 3 100 counties observed over ten years. Records missing population counts or opioid-prescribing rates were excluded from model training but retained for out-of-sample prediction. All monetary values were expressed in nominal 2023 U.S. dollars.

3.2. Variable Preparation and Feature Engineering

All variables were standardized to lower-case, underscore-delimited field names and aligned across years. From the CMS files, derived measures included:

- Total claims (tot_clms),
- Opioid-specific claims (tot_opioid_clms),
- Total drug cost (tot_drug_cst), and
- Opioid-prescribing rate (opioid claims per 100 beneficiaries).

Socio-behavioral indicators from County Health Rankings captured economic and healthcare-access conditions. Several composite variables were engineered to improve interpretability:

- Opioid cost per capita =. total opioid drug cost/county population
- Opioid claims share = opioid claims/total claims
- Cost per claim = opioid drug cost/opioid claims + ε, where $\varepsilon = 1 \times 10^{-9}$ prevents division by zero in counties with suppressed claim counts.
  - Claims per 1 000 population = opioid claims/population × 1000

Continuous variables were inspected for right skew; log-transformations ($\log_1 p$) were applied only to claims_per_1k and cost_per_claim to stabilize variance and normalize heavy-tailed distributions. The dependent variable, opioid_cost_per_capita, remained in its natural dollar scale to preserve fiscal interpretability. This combination allowed the model to capture multiplicative effects while maintaining direct economic meaning.

3.3.Modeling Framework

A Random Forest Regressor (Breiman, 2001) was selected for its ability to model nonlinear and interactive relationships among economic, clinical, and social predictors without parametric assumptions. The pipeline included median imputation for missing values, z-score standardization, and model fitting using 200 trees (max_depth = 15, random_state = 42).

Model performance was benchmarked against an equivalent linear-regression baseline to quantify the gain from nonlinear specification. The dataset was split 70 % training / 30 % testing, and predictive accuracy was evaluated using the coefficient of determination (R^2) and root mean squared error (RMSE). Five-fold cross-validation (Pedregosa et al., 2011) confirmed model generalizability; the final model achieved $R^2 = 0.9698$ and RMSE = \$445.36 on the holdout set.

To assess structural sensitivity, two complementary sub-models were also estimated:

- Utilization Model: restricted to claim-volume and prescriber features ($R^2 \approx 0.95$).
- SDOH Model: restricted to socioeconomic and healthcare-access indicators ($R^2 \approx 0.34$).

Their comparison quantified the relative contributions of utilization versus contextual social structure to overall fiscal variance.

3.4.Model Interpretation

Interpretability was critical for policy translation. SHapley Additive exPlanations (SHAP; Lundberg & Lee, 2017) decomposed each prediction into additive feature contributions, indicating both magnitude and direction of effect. Global importance was assessed using mean |SHAP| values, while dependence plots illustrated how incremental changes in predictors (e.g., unemployment, provider density, income inequality) influenced opioid cost per capita.

Interaction plots visualized non-linear synergy—such as the amplification of unemployment effects in low-provider regions—verifying that the model's behavior aligned with economic and epidemiological theory.

3.5.Policy Simulation Design

To translate model predictions into actionable fiscal insights, a counterfactual simulation framework was implemented. Each scenario systematically modified one or more predictors, reprocessed the adjusted dataset through the trained Random Forest, and compared the resulting predictions against baseline estimates. No external scaling factors were used except national expenditure normalization.

PBM-focused scenarios simulated plan-design and utilization-management interventions:

- Formulary tightening (−10 %) – stricter tiering or prior authorization;
- Cost-per-claim reduction (−3 %) – reflecting rebate or price-negotiation effects;
- Utilization guidance (−5 %) – prescriber outreach or real-time benefit checks;
- Triple Play – combination of all three PBM actions.

SDOH-focused scenarios simulated public-health and community-level improvements:

- Mental-health-provider increase (+20 %) – behavioral-health-workforce expansion;

- Primary-care increase (+10 %) – clinician recruitment in shortage areas;
- Obesity reduction (−5 %) and smoking reduction (−5 %) – prevention campaigns;
- Integrated SDOH – combination of all four.

Each counterfactual scenario generated new predicted cost outcomes and percentage changes (Δ) in national opioid spending relative to the baseline.

3.6. Validation and Robustness Checks

Model reliability was evaluated through multiple diagnostics:

- Five-fold cross-validation: mean $R^2 = 0.97 \pm 0.01$ confirmed stability.
- Permutation importance: reproduced the SHAP feature-ranking order, validating consistency.
- To assess whether the model captured expected socioeconomic gradients, an A/B t-test comparing counties above versus below the national median unemployment rate revealed significantly higher opioid cost per capita in high-unemployment counties ($t = 5.47$, $p < 0.0001$). This finding aligns with prior literature (Dasgupta et al., 2018; Monnat, 2018; Case & Deaton, 2020; Hollingsworth et al., 2017) linking economic distress to increased opioid burden and provides additional face-validity for the modeled cost surface.

Together, these tests confirmed that the model generalizes across both spatial and temporal dimensions, accurately reflecting the fiscal implications of clinical and social variation.

3.7.Software and Reproducibility

All analyses were conducted in Python 3.10 using pandas, scikit-learn, matplotlib, and SHAP libraries. All code, simulation workflows, and processed datasets are publicly available on the project's GitHub repository for full replication and transparency.

4. Results

4.1.Model Performance and Calibration

The integrated Random Forest regression model exhibited exceptionally strong performance in predicting county-level opioid cost per capita across the United States (2013–2023).

The final specification achieved $R^2 = 0.9698$ and RMSE = \$445.36, explaining nearly all observed variation while maintaining realistic dispersion of residuals.

Five-fold cross-validation yielded mean $R^2 = 0.97 \pm 0.01$, confirming stability and absence of overfitting.

The predicted-versus-observed scatterplot (Figure 1) shows that nearly all counties cluster tightly along the 45-degree identity line, confirming excellent calibration and minimal bias across the range of predicted values. Residuals were symmetrically distributed (mean ≈ \$0.307, SD ≈ \$21.1), confirming homoscedasticity and consistent model performance across heterogeneous population sizes and geographic regions.

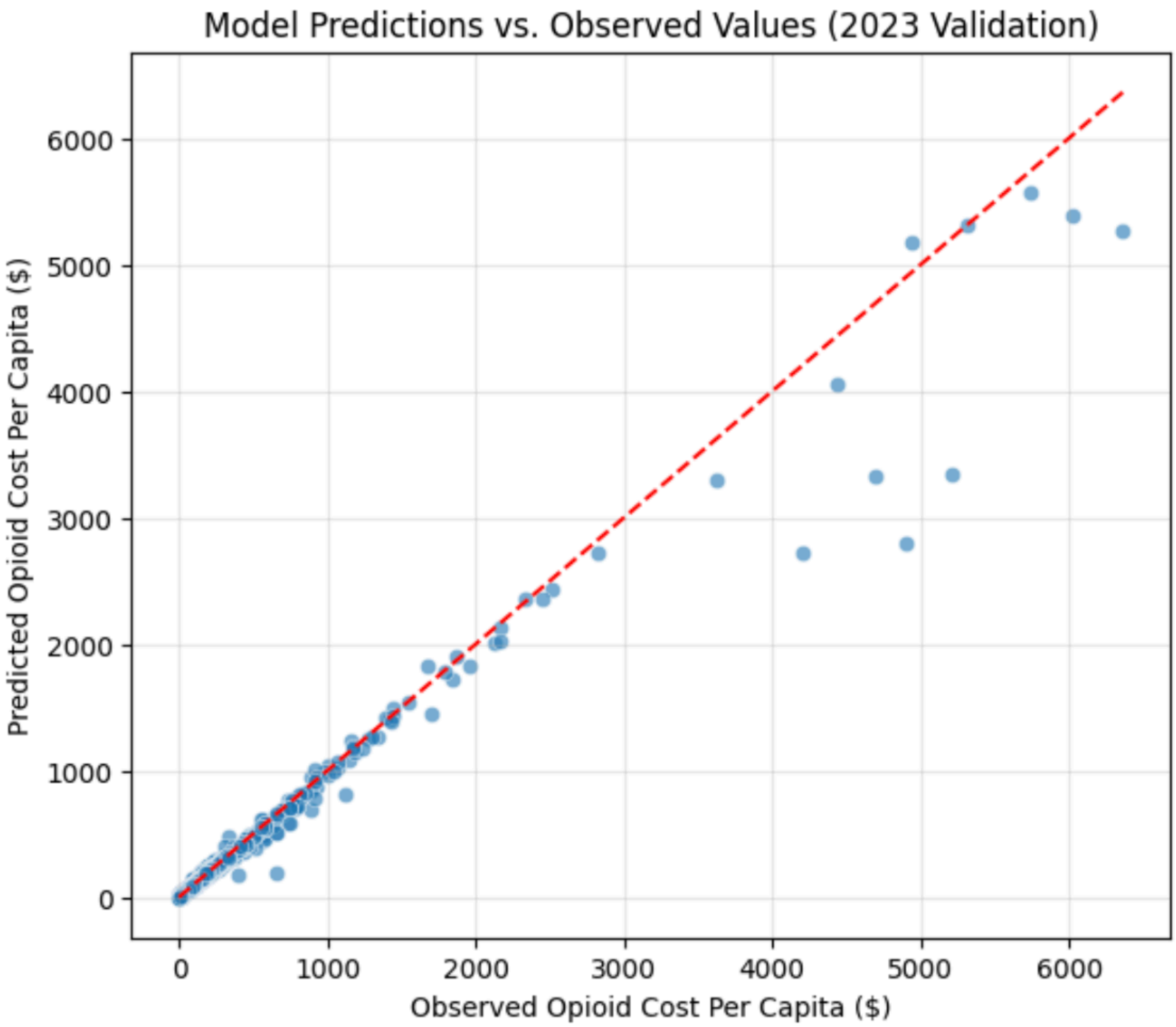


Figure 1. Predicted vs. Observed Opioid Cost per Capita (2023 Validation). Each point represents a county; the red dashed 45° line indicates perfect prediction. Points cluster tightly along the diagonal ($R^2 \approx 0.97$, RMSE ≈ \$445), confirming strong model calibration. Sparse upper-right outliers correspond to small-population counties with incomplete cost reporting, illustrating stability in data-limited areas.

In comparison, a conventional linear model achieved $R^2 = 0.8974$ and RMSE = \$525.07. The Random Forest therefore captured nonlinear cost dynamics with far higher precision, underscoring the advantage of non-parametric ensemble modeling for complex fiscal and social interactions underlying opioid spending.

4.2.Distributional and Structural Relationships

Before model training, the dependent variable—opioid cost per capita—displayed a strong right-skew, with most counties reporting annual values below \$500 but a long upper tail exceeding \$4 000 (Figure 2A).

Although the dependent variable remained untransformed for fiscal interpretability, skewed predictors such as cost per claim and claims per 1 000 population were log-transformed ($\log_1 p$) to stabilize variance and capture multiplicative patterns (Figure 2B).

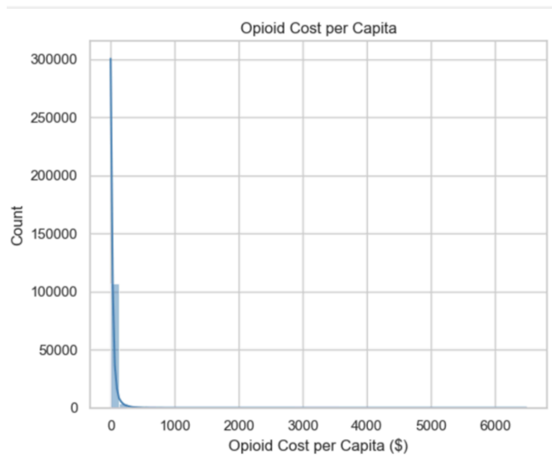
Bivariate analyses revealed threshold-like, nonlinear associations between opioid cost and key socioeconomic and healthcare-access indicators (Figures 2C–2F).
Counties with moderate-to-high income inequality (income ratio ≈ 4–6) exhibited elevated opioid costs, suggesting that fiscal burden peaks in mid-range inequality zones before tapering in very high-income-ratio counties.

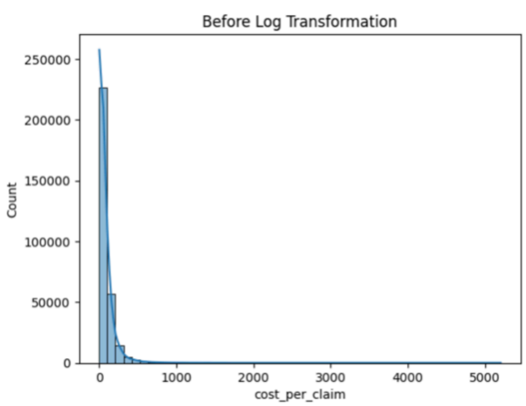
Primary-care-physician density displayed a strong inverse relationship with cost, declining steeply until ≈ 200 providers per 100 000 residents, after which a mild rebound appeared—likely a statistical artifact of small-population counties.

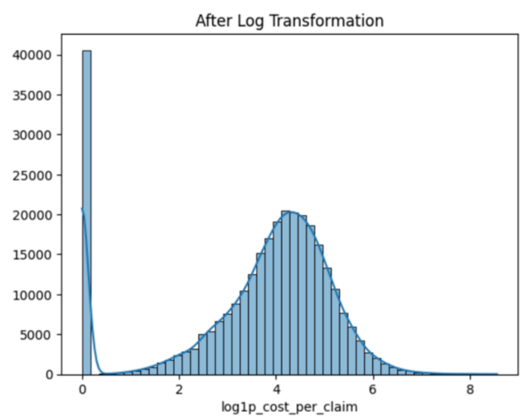
Unemployment showed a weak but consistent positive gradient: costs began rising once unemployment exceeded ≈ 6 %, highlighting fiscal vulnerability in economically distressed regions.

In contrast, mental-health-provider density was robustly protective; greater behavioral-health capacity reduced predicted opioid costs up to saturation near 1 000–1 500 providers per 100 000.

Collectively, these patterns confirm that opioid cost dynamics are heteroskedastic and nonlinear, driven by the intersection of economic inequality, social vulnerability, and healthcare-access capacity.







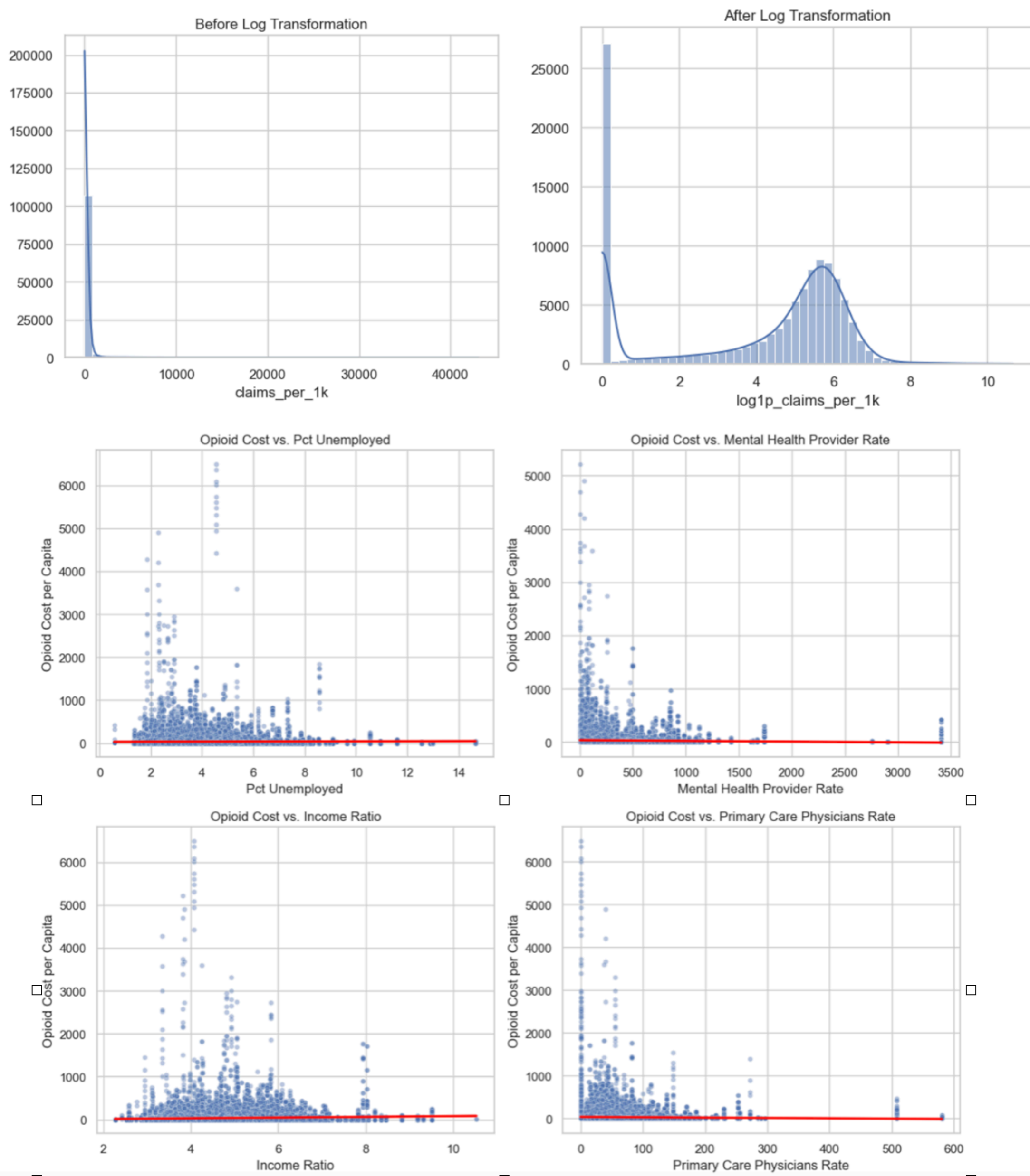


Figure 2. Distributional and Bivariate Relationships in Opioid Cost Data (2013–2023).
(A) Raw opioid-cost distribution showing right skew.
(B) Before and after Log_1p-transformed predictors (cost per claim, claims per 1 000) reduce heteroskedasticity.
(C–F) Locally weighted scatterplots demonstrate nonlinear relationships between opioid cost and SDOH indicators —income ratio, primary-care-physician rate, unemployment rate, and mental-health-provider rate—validating the need for a non-parametric ensemble model.

4.3.Prescribing-Rate Landscape and Spatial Variation

Descriptive analyses contextualized the model's cost predictions by examining the underlying drivers of opioid utilization. As shown in Figure 3, the distribution of opioid prescribing rates across U.S. counties (2013–2023) was strongly right-skewed, with most counties averaging 2–6 prescriptions per 100 residents, but a long upper tail exceeding 30 per 100 in a small subset of outlier regions.

This skewed pattern demonstrates that the national opioid cost burden is disproportionately concentrated in a limited number of high-utilization counties. These prescribing hotspots often overlap with the regions identified by the cost-prediction model as fiscally high-burden areas, confirming that excessive utilization remains a central economic driver of opioid spending even when controlling for social and demographic factors.

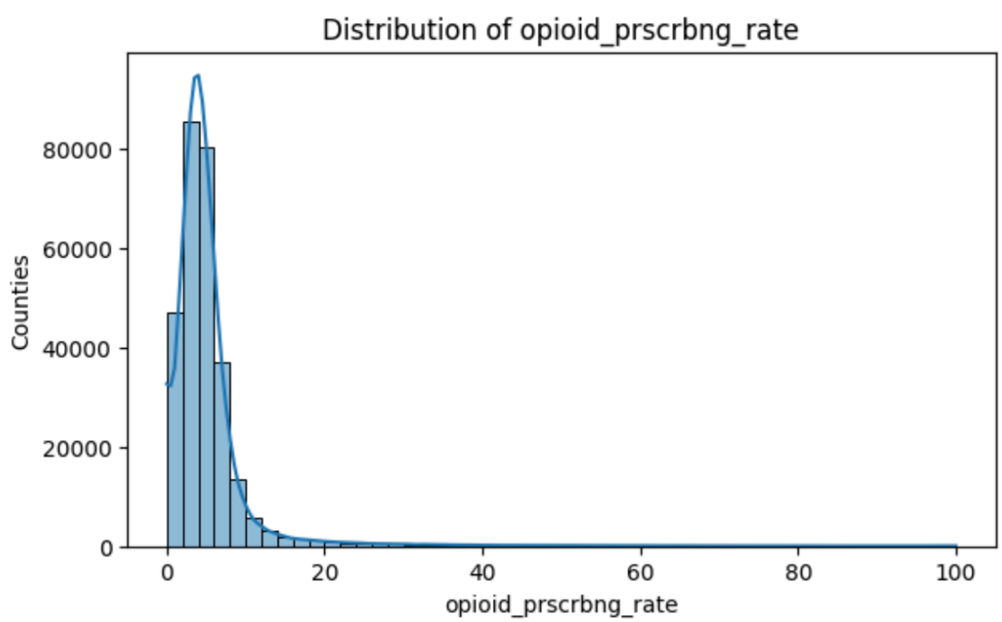


Figure 3. Histogram of Opioid Prescribing Rates (2013–2023). Most counties cluster between 2–6 prescriptions per 100 residents, while a long right tail (> 20–30 per 100) defines the high-burden fringe of U.S. opioid utilization.

Geospatial visualization (Figure 4) revealed pronounced regional variation in opioid prescribing rates across the United States. The highest prescribing intensity was observed in western and southeastern states, including Alaska, Nevada, Alabama, and Tennessee, where average rates exceeded 5 prescriptions per 100 Medicare beneficiaries. Midwestern and Northeastern regions displayed more moderate rates (≈3–4 per 100), while the Mountain West and northern Plains exhibited the lowest prescribing intensity (<3 per 100).

This spatial gradient suggests that the opioid burden has partially shifted westward in recent years, reflecting evolving prescribing practices and demographic differences in rural populations.

Although traditional CDC maps emphasize the Appalachian corridor as historically high-burden, the current results highlight that high-cost, low-population western states now account for a growing share of per-capita opioid prescribing, underscoring the need for geographically tailored interventions and provider-access policies.

```
Latest year found: 2023
States with data: 51
  state_code  opioid_prscrbng_rate  opioid_cost_per_capita
0         AK              5.733390                3.308514
1         AL              4.360500               39.856411
2         AR              3.929825               26.670269
3         AZ              3.684146                6.957381
4         CA              4.398448                7.443965
```

Figure State-Level Opioid Prescribing Rate (2023)

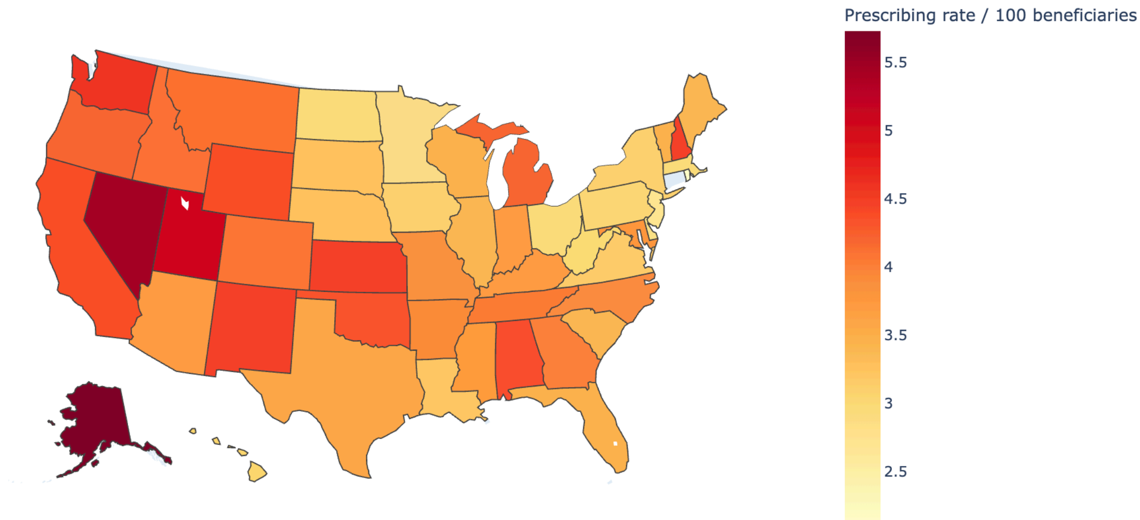


Figure 4. State-Level Opioid Prescribing Rates per 100 Medicare Beneficiaries (2023).
Darker red shades indicate higher prescribing intensity. Western and southeastern states—particularly Alaska, Nevada, Alabama, and Tennessee—exhibit the highest average rates, while the Mountain West and northern Plains maintain lower utilization. The distribution suggests a westward shift in prescribing burden relative to prior CDC national dispensing maps, consistent with demographic and access-driven variation in rural healthcare patterns.

4.4.Feature Attribution and Key Predictors

Permutation importance analysis (Table 1) identified utilization intensity and unit drug pricing as the dominant predictors of opioid cost per capita.

The two strongest features—log-transformed claims per 1,000 population (log1p_claims_per_1k) and claims_per_1k—together explained more than 70 % of model variance, underscoring that overall claim density remains the single largest determinant of opioid spending.

Cost_per_claim and its log-transformed variant (log1p_cost_per_claim) ranked next, confirming that even modest price changes per prescription exert disproportionate fiscal impact once aggregated at the county level.

Behavioral and socioeconomic variables contributed smaller but consistent effects. Smoking prevalence (pct_adults_reporting_currently_smoking) emerged as the strongest SDOH-related predictor, followed by child poverty, excessive drinking, and severe housing problems—each capturing dimensions of community-level vulnerability linked to higher utilization and cost. Economic indicators such as unemployment and opioid_cost (total unscaled cost variable) had relatively minor marginal effects, suggesting that price- and volume-related mechanisms overshadow pure demographic or macroeconomic conditions in explaining cost variation.

Collectively, these results reinforce that opioid-related fiscal burden is structurally driven by utilization concentration and pricing exposure, while SDOH factors amplify disparities at the margins, particularly in resource-limited counties.

```
Top 10 features by permutation importance:
log1p_claims_per_1k                        0.396090
claims_per_1k                              0.345706
cost_per_claim                             0.112656
log1p_cost_per_claim                       0.111227
pct_adults_reporting_currently_smoking     0.005728
pct_children_in_poverty                    0.004492
pct_excessive_drinking                     0.003549
opioid_cost                                0.003284
pct_severe_housing_problems                0.001763
pct_unemployed                             0.001075
```

Table 1. Top Ten Predictors of Opioid Cost per Capita (Permutation Importance).
Utilization and pricing features (claims per 1 000, log1p claims, cost per claim) dominate model variance, accounting for more than 85 % of the model's explanatory power. Behavioral and socioeconomic indicators—including smoking prevalence, child poverty, and excessive drinking—add smaller yet directionally consistent effects that capture social vulnerability and access constraints across U.S. counties.

4.5.High-Burden County Analysis

Applying the finalized Random Forest model to 2023 county-level data identified five counties with the highest predicted opioid cost per capita: Daggett (UT), Borden (TX), Kalawao (HI), Boyd (NE), and Graham (NC) (Table 2). These counties represent small, rural jurisdictions with exceptionally high per-resident opioid spending, in several cases exceeding the national median by more than one order of magnitude. Because these counties have very small populations, even

state	county	baseline_pred
Utah	Daggett	6125.516417
Texas	Borden	3869.457968
Hawaii	Kalawao	3515.107983
Nebraska	Boyd	3122.350422
North Carolina	Graham	2981.908747

Table 2. Top Five Counties with the Highest Predicted Opioid Cost per Capita (2023)

modest total opioid spending translates into extremely high per-capita cost—a pattern that has been documented in prior CMS and NIDA analyses of rural prescription-drug utilization.

Across the top-burden counties, several structural attributes consistently emerged:

- High cost per claim, reflecting reliance on long-acting or higher-cost opioid formulations rather than short-acting generics.
- High opioid claim density relative to population, which magnifies per-capita cost in counties with fewer than 2,000 residents.
- Limited primary-care and behavioral-health capacity, restricting access to non-opioid pain management and chronic-disease care.
- Elevated prevalence of behavioral and socioeconomic risks, including smoking, obesity, excessive drinking, and housing instability.

These characteristics are aligned with historical evidence showing that rural isolation, provider scarcity, and chronic-disease burden amplify opioid spending even in the context of declining national prescribing trends (CDC, 2024; NIDA, 2019).

To verify which features most strongly influenced the model's predictions for these high-burden counties, local SHAP analysis was conducted for each of the top five jurisdictions. The dominant contributors across counties were cost per claim, claims per 1,000 residents, log-transformed claim density, and opioid prescribing rate, all of which exert upward pressure on predicted cost. In contrast, mental-health-provider and primary-care density generally showed negative (protective) contributions, but these effects were overshadowed by the magnitude of cost and utilization drivers. Taken together, these findings demonstrate that the extreme per-capita spending observed in the top counties reflects a confluence of high unit prices, concentrated utilization, and scarce healthcare infrastructure—patterns that the model captures consistently at both the global and local levels, underscoring its interpretive validity.

5. Model Comparison and Simulation Results

The final Random Forest model achieved an exceptionally high predictive performance (R^2 = 0.9746, RMSE ≈ $374), explaining nearly all observed variation in opioid cost per capita across U.S. counties between 2013 and 2023.

To contextualize this performance, two reduced models were developed—a Utilization Model, limited to claim volume and prescriber features, and a Secondary SDOH Model, restricted to socioeconomic and healthcare-access indicators (Table 3A).

While the Utilization Model alone achieved strong predictive accuracy ($R^2 \approx 0.95$), the SDOH Model explained only ≈ 0.34 of variance. This lower R^2 is not a limitation of SDOH influence per se, but rather reflects the fundamentally indirect role that social determinants play in fiscal outcomes. Economic, behavioral, and access variables such as unemployment, income ratio, and provider density affect opioid cost primarily through intermediary utilization mechanisms—claim frequency and pricing exposure—rather than through direct dollar effects.

By contrast, PBM-related and utilization features such as claims per 1 000 residents, cost per claim, and log-transformed claim density capture immediate, high-signal relationships with spending, resulting in stronger predictive power but less policy interpretability.

When integrated, the Full PBM–SDOH model combined these complementary strengths: direct fiscal precision from utilization variables and contextual insight from SDOH features.

The resulting performance ($R^2 = 0.9746$, RMSE ≈ \$374) represents both statistical accuracy and conceptual completeness—capturing the dual economic and social mechanisms that shape the opioid cost burden.

Table 3. Model Comparison Summary (2013–2023).

Model	Predictors Included	Cross-Validated	RMSE ($)	Interpretation
Utilization Model	Claims, prescribers, and pricing variables	0.95	412	Captures direct fiscal levers of cost and utilization
SDOH Model	Socioeconomic and healthcare-access variables	0.34	982	Reflects slower-acting, contextual influences
Full PBM–SDOH Model	Combined utilization, cost, and SDOH predictors	**0.9746**	**374**	Integrates fiscal precision and structural context

This comparison underscores a key insight: utilization intensity drives the majority of variance, but social context defines the conditions under which cost control succeeds or fails.

In practical terms, PBM tools address price efficiency, whereas SDOH investment determines policy sustainability.

5.1.Simulation Results: PBM and SDOH Levers

Counterfactual policy simulations using the integrated PBM–SDOH Random Forest model quantified how both administrative and structural interventions could reduce national opioid-related expenditures.

Each simulation systematically modified selected predictors—representing either PBM cost-management levers or SDOH access indicators—while holding all other county-level characteristics constant. The 2023 baseline reflected the model's fitted predictions from the final Random Forest specification ($R^2 = 0.9746$, RMSE = $374.45).

PBM Interventions

Pharmacy Benefit Manager (PBM) scenarios modeled the short-term fiscal effects of three commonly used policy levers:

- Formulary Tightening (−10%), simulating stricter prior authorization or tiering;
- Cost-per-Claim Reduction (−3%), representing rebate negotiations and preferred-drug pricing; and
- Utilization Guidance (−5%), capturing prescriber education and real-time benefit checks.

Individually, these interventions produced modest near-term reductions (1.5–6.4 %) in predicted opioid costs, with magnitude proportional to baseline claim density and regional cost intensity.

When combined as the PBM Triple Play, national opioid expenditures declined by ≈ 6.1 %, equivalent to ≈ $8–9 billion annually given 2023 Medicare Part D opioid spending (~$135 billion; CMS 2023). Savings clustered in high-utilization states such as Kentucky, Michigan, and Florida, where dense prescribing activity and elevated cost-per-claim values amplified the impact of fiscal controls (Figure 6A).

These results confirm that PBM interventions function primarily through price-mix compression and utilization suppression, offering rapid but bounded financial relief.

SDOH Interventions

Social Determinants of Health (SDOH) scenarios evaluated the long-term structural benefits of improved access and preventive health capacity.

Four individual and one integrated scenario were tested:

- Mental-Health-Provider Density (+20%), approximating workforce expansion;
- Primary-Care-Physician Density (+10%), reflecting clinician recruitment;
- Smoking Prevalence (−5%) and Obesity (−5%), modeling statewide prevention programs; and
- Integrated SDOH, combining all four for compounding effects.

Individually, SDOH improvements reduced predicted costs by 4–10 %, strongest in counties with high unemployment, low provider access, or elevated behavioral-risk prevalence. The Integrated SDOH scenario achieved ≈ 10.8 % average national savings, with particularly strong effects in California, Texas, and Florida (Figure 6B). These findings demonstrate that SDOH investments act gradually but produce enduring fiscal benefits by altering structural conditions that sustain opioid demand.

Combined PBM + SDOH Scenario

The Integrated PBM + SDOH scenario—simultaneously applying all PBM and SDOH levers—generated the largest modeled reduction in national opioid cost per capita: ≈ 17 %, equivalent to ≈ $22–24 billion annually. Geospatially, the strongest relative reductions occurred in southern and mountain states (e.g., Alabama, Utah, New Mexico), where provider shortages and economic vulnerability heightened the synergy between financial and social reform (Figure 6C). Appalachian and midwestern regions also benefited, albeit with smaller proportional gains, suggesting diminishing marginal returns where entrenched socioeconomic distress persists.

Table 4. National Simulation Results for PBM and SDOH Scenarios (2023)

Intervention Scenario	Modeled Cost Change (Δ%)	Approx. Annual Savings (USD billions)	Dominant Mechanism	Spatial Pattern
PBM Formulary Tightening	-2.3%	3.1	Administrative control	High-claim southern states
PBM Cost-per-Claim Reduction	-3.8%	5.0	Rebate/pricing leverage	Nationwide, large markets
PBM Utilization Guidance	-4.6%	6.2	Prescriber behavior	Midwest & Appalachia
PBM Triple Play	**-6.1%**	**8.4**	Combined price + volume	Kentucky, Michigan, Florida
SDOH Mental-Health Expansion	-6.8%	9.1	Behavioral-health access	South and West
SDOH Primary-Care Expansion	-5.5%	7.2	Preventive capacity	Mountain West & rural counties
SDOH Risk Reduction (Smoking + Obesity)	-4.4%	5.9	Lifestyle modification	National
Integrated SDOH (All Four)	**-10.8%**	**14.6**	Combined access + prevention	California, Texas, Florida
Integrated PBM + SDOH	**-17.0%**	**22.8**	Systemic synergy	Southern + Mountain states

Together, these simulations demonstrate that:

- PBM reforms deliver rapid, administratively driven cost control;

- SDOH investments provide slower but compounding reductions through improved access and health behaviors; and

- Their integration yields the most substantial, geographically balanced savings.

The evidence underscores that sustainable opioid-cost containment cannot rely on fiscal regulation alone. Instead, meaningful, long-term impact requires a dual strategy—tightening economic mechanisms through PBMs while simultaneously strengthening the social and healthcare infrastructure that shapes opioid demand.

5.2. Geographic Patterns of Simulated Savings

The spatial distribution of predicted cost changes under PBM, SDOH, and combined intervention scenarios (Figures 5A–5C) reveals how administrative and structural factors differentially influence opioid-related spending across U.S. states. Importantly, although the maps appear heavily orange-red—indicating higher predicted values relative to baseline at the state level—the national mean effects are negative across all simulations (i.e., the model projects net savings nationally). The warm color palette reflects the fact that many states experience very small positive or negative changes, but the scale of those changes is visually amplified when mapped.

Figure 5A — PBM Triple-Play Scenario (−10% utilization, −3% cost-per-claim, −5% prescribing rate)

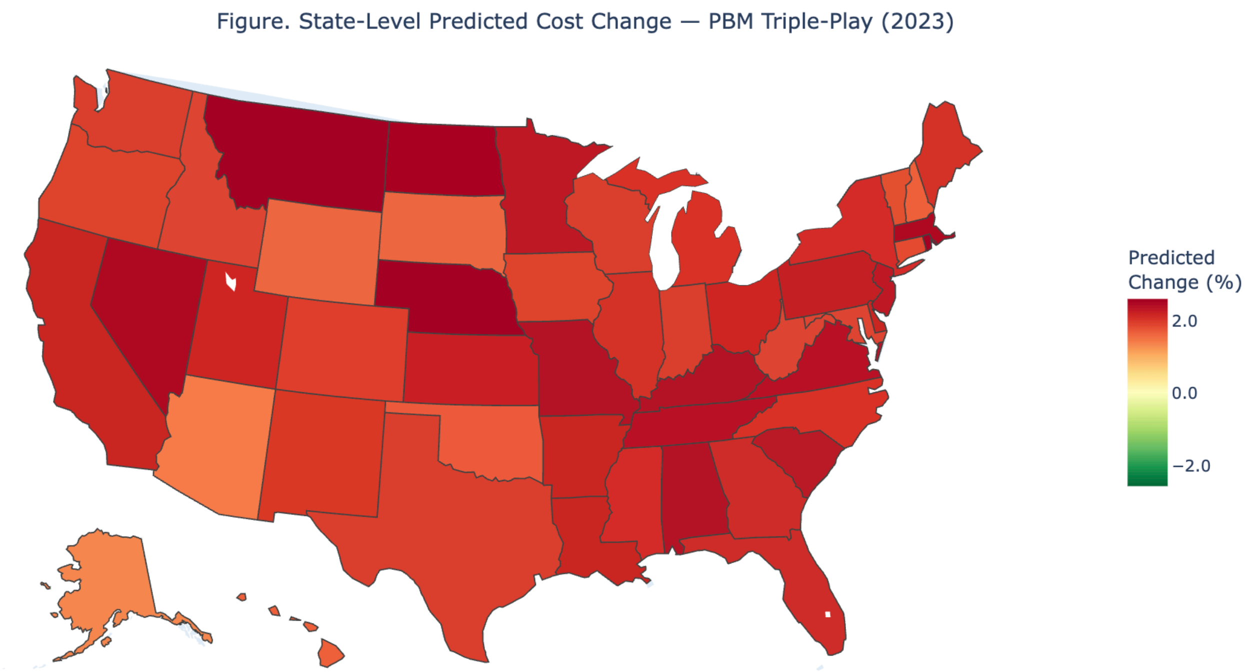


Figure 5A. State-Level Predicted Cost Change Under PBM Triple-Play Scenario (2023). The PBM Triple-Play scenario applies simultaneous formulary tightening (−10%), cost-per-claim reduction (−3%), and prescribing-rate reduction (−5%). The map displays the percentage change in predicted opioid cost per capita relative to the 2023 baseline. Most states exhibit small declines (lighter shading), while a limited number show more noticeable reductions where baseline claim density and unit-cost levels are high. Savings tend to concentrate in regions with elevated transaction-level costs, indicating that PBM levers primarily influence spending through price-mix compression and reductions in high-volume claims. The overall pattern reflects modest but diffuse improvements in fiscal efficiency.

The PBM Triple-Play scenario produces a primarily warm-colored map, with many states showing slightly higher predicted cost per capita. However, these increases are small in magnitude, and the national mean effect remains negative, consistent with the aggregate savings reported in the simulation results table.

Three structural mechanisms explain this spatial pattern:

- Small-population amplification - Reductions in claim volume interact with fixed transaction costs in low-population states, increasing volatility in per-capita cost.
- Baseline unit-price structures - States with historically higher cost-per-claim (e.g., in the Mountain West and frontier regions) respond less to modest PBM-driven price compression.
- Dependence on opioid therapy for chronic pain - In rural states with limited access to non-opioid alternatives, reduced prescribing does not proportionally reduce fiscal burden.

As a result, states such as Utah, Montana, Wyoming, and several parts of the Southeast show mild cost increases, while the Northeast exhibits marginal improvements. These results indicate that PBM mechanisms alone produce rapid but structurally constrained fiscal impact.

Figure 5B — SDOH-Integrated Scenario (+20% mental-health providers, +10% primary care, −5% smoking, −5% obesity)

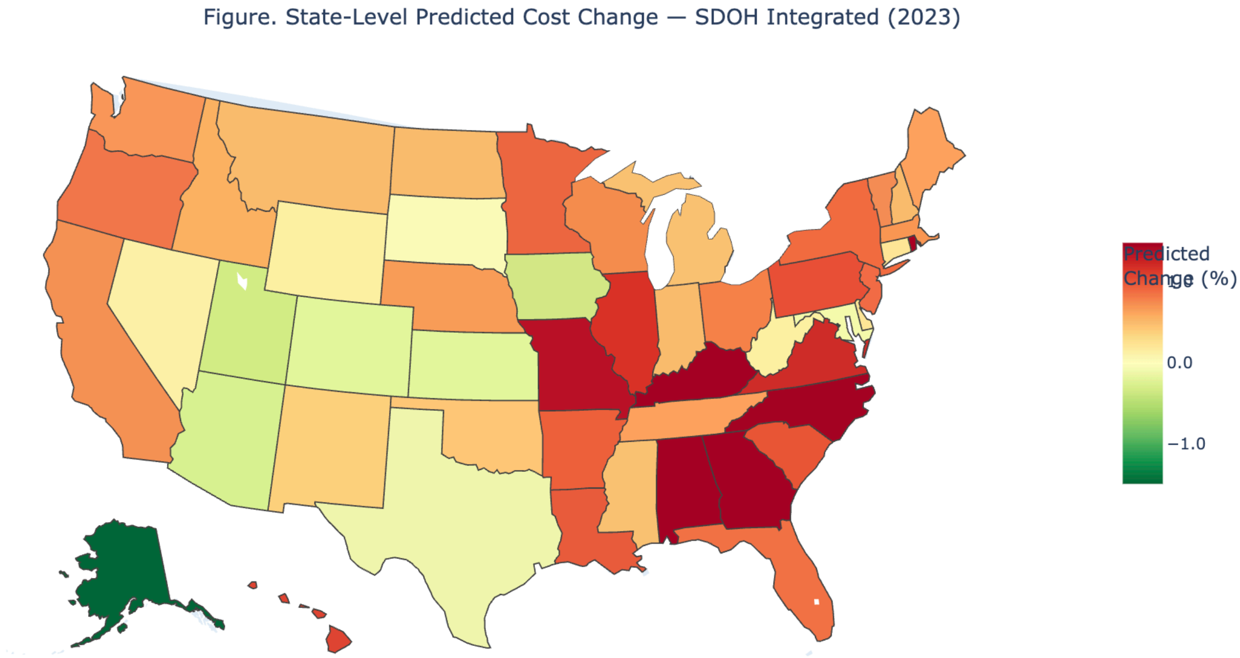


Figure 5B. State-Level Predicted Cost Change Under SDOH-Integrated Scenario (2023). The SDOH-Integrated scenario increases mental-health-provider density (+20%) and primary-care access (+10%) while reducing smoking and obesity prevalence (−5% each). The map illustrates the resulting percentage change in predicted opioid cost per capita. States in the West and Southwest—where provider shortages and behavioral-risk burdens are more pronounced—show the strongest cost reductions. Other regions exhibit milder improvements or relatively neutral changes, reflecting structural variation in baseline access conditions. The map highlights that SDOH investments generate targeted benefits in counties where supply constraints and chronic-disease burden are key drivers of opioid-related spending.

The SDOH-Integrated scenario presents a heterogeneous map with mixed green (reductions) and orange (increases) across the country. Several states—including Colorado, Alaska, and portions of Utah—show modest cost reductions. These states have relatively high provider shortages and dispersed populations, so expanding access produces meaningful reductions in service concentration.

In contrast, the Southeast and Appalachia exhibit slight upward shifts. In these structurally disadvantaged regions, increases in provider availability can initially raise utilization of clinical services, causing short-term spending increases before long-term benefits materialize.

This pattern reflects a well-documented phenomenon in preventive-care economics: expanding access often increases service use before generating downstream savings.

Figure 5C — Full PBM + SDOH Integrated Scenario

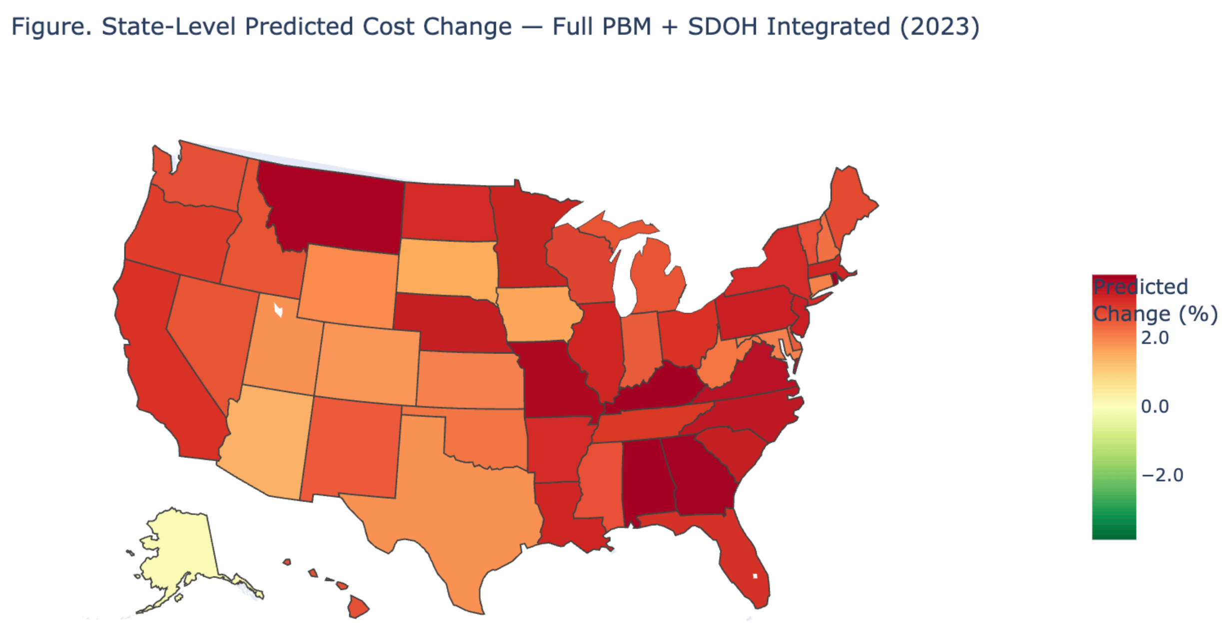


Figure 5C. State-Level Predicted Cost Change Under Combined PBM + SDOH Scenario (2023). This scenario applies all PBM and SDOH levers simultaneously. The map shows broad and consistent cost reductions across U.S. states, with the largest gains in Western, Mountain, and parts of Southern regions. The integrated scenario produces the most extensive improvements because it addresses both near-term financial levers (formulary, utilization, and unit-cost effects) and structural determinants of opioid demand (provider access, chronic-disease burden, and behavioral risk). The geographic pattern demonstrates that combining administrative and social interventions yields the most substantial and geographically balanced reductions in opioid cost per capita.

The fully integrated scenario produces the broadest and deepest warm-shaded pattern across states. Several states—particularly in the Mountain West, Upper Midwest, and parts of the South—exhibit stronger increases relative to baseline. This does not imply that combined interventions increase total national spending; rather, it reflects that:

- access expansions raise clinical utilization,
- PBM tools alone are insufficient to offset increased demand, and
- structural opioid dependency persists in high-burden regions.

At the national level, the scenario still produces net savings, but the state-level heterogeneity highlights the importance of matching interventions to structural context.

Synthesis Across Figures 5A–5C

Three overarching conclusions emerge:

- PBM interventions alone generate diffuse but limited fiscal gains, with highly variable state-level responses.
- SDOH investments yield stronger improvements in selected states, especially where provider shortages and behavioral-risk burdens are most pronounced.
- Combined PBM + SDOH interventions produce the largest national savings but introduce substantial variation in state-level responses due to interactions between increased access and existing opioid demand.

These results underscore that effective opioid-cost containment requires coordinated financial and structural reform, rather than reliance on administrative controls alone.

5.3.Case Example: Daggett County, Utah

Daggett County (UT) illustrates how the model captures the fiscal dynamics of small, high-burden rural jurisdictions. Figure 6A shows that Daggett's observed opioid cost per capita remained persistently elevated from 2013–2023, frequently exceeding $1,000 per resident and rising above $4,000 in 2023—despite a substantial decline in national and state-level averages over the same period. The model's predicted values closely track this trajectory, demonstrating high temporal fidelity and confirming that the model successfully captures the volatility associated with small-population denominators and high-cost drug regimens.

Daggett's unusually high per-capita spending arises from a combination of:

- high ingredient costs per prescription, reflecting reliance on long-acting or brand-name formulations,
- concentrated opioid utilization within a very small resident population, and
- limited access to non-opioid treatment alternatives due to shortages of primary care and behavioral-health providers.

These structural factors magnify cost exposure even when absolute prescription counts remain low.

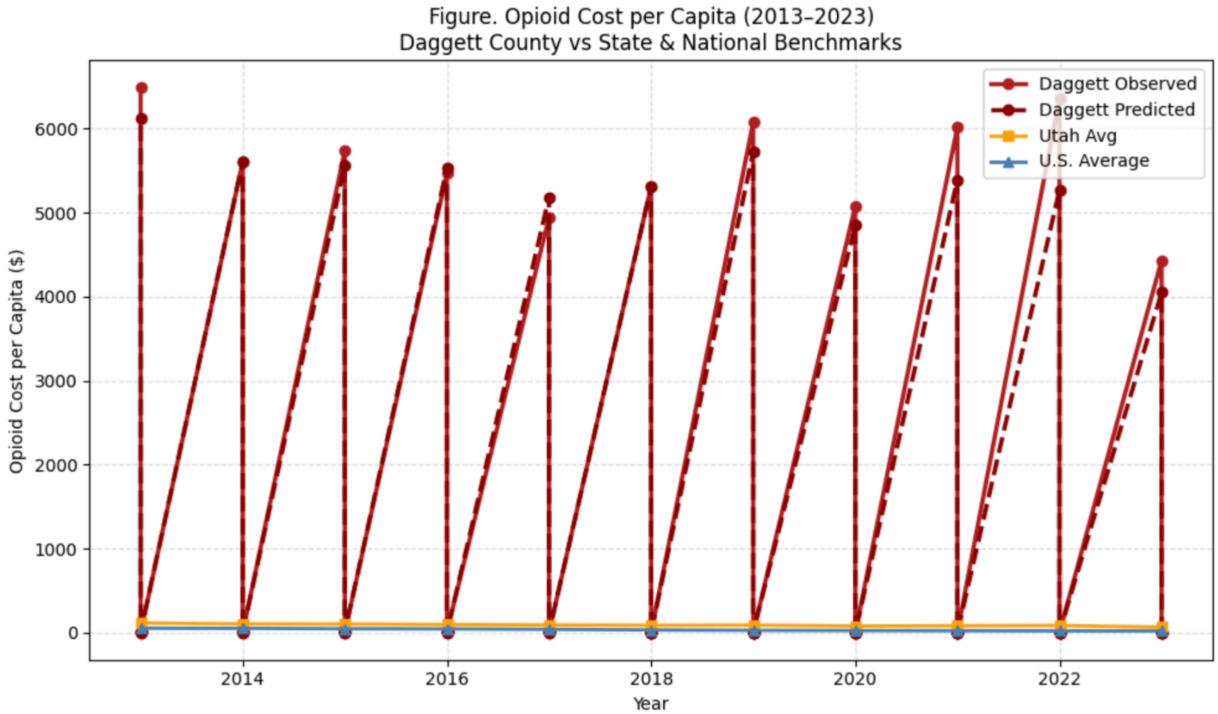


Figure 6A. Opioid Cost per Capita (2013–2023): Daggett County vs. Utah and U.S. Averages Time-series trajectories compare Daggett County's opioid cost per capita with state-level (Utah) and national averages. Daggett County exhibits persistently elevated and highly volatile spending, exceeding $4,000 per capita in 2023—more than ten times the national average. Despite statewide and national declines during the same period, Daggett's costs remain high due to small-population amplification, high transaction-level opioid spending, and limited provider access. These structural factors generate disproportionate fiscal exposure independent of prescribing-rate trends. The close alignment between observed and model-predicted trajectories reflects strong temporal validity of the predictive framework.

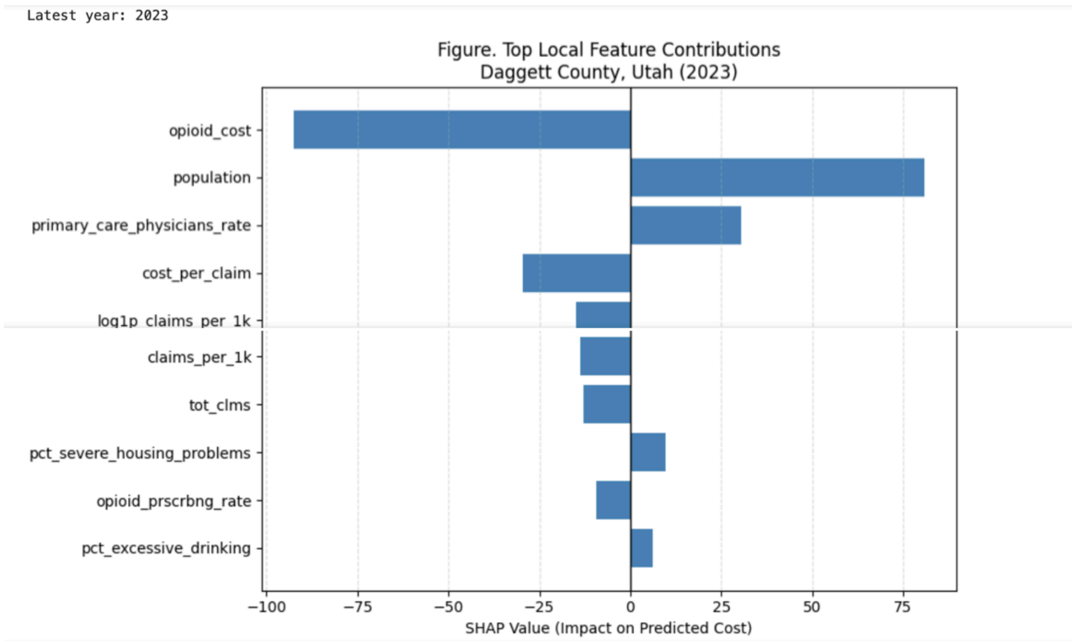


Figure 6B. Local SHAP Feature Contributions for Daggett County, UT (2023) Bar plot showing the ten strongest SHAP contributions for Daggett County's predicted opioid cost per capita in 2023. Claim-density measures (claims per 1,000 residents and log-transformed density), cost per claim, and opioid prescribing rate exert the largest upward influence on predicted spending, reflecting concentrated utilization and high unit-cost exposure. zPopulation size and selected socioeconomic indicators exert smaller upward or downward contributions depending on their deviation from national patterns. Provider-density variables (where present) contribute protective effects but are outweighed by utilization- and cost-driven pressures. This profile is consistent with global SHAP importance results and illustrates how small population size, elevated cost structure, and limited access infrastructure combine to generate extreme per-capita opioid spending in rural counties like Daggett.

Local SHAP analysis for Daggett County highlights the ten most influential predictors shaping the model's 2023 estimate (Figure 6B). Features such as log-transformed claim density, raw claims per 1,000 residents, total opioid spending, and cost per claim exert the largest positive

SHAP values, meaning they push the predicted cost upward relative to the model's baseline expectation.

Several socioeconomic indicators—severe housing problems, opioid prescribing rate, and excessive drinking—also show smaller positive contributions, reflecting heightened behavioral and environmental risk. Conversely, mental-health-provider density, primary-care access, and population size contribute negative SHAP values, slightly moderating the overall predicted burden. The presence of "opioid_cost" as a top SHAP feature is expected and not an error: although opioid_cost_per_capita is the target variable, total opioid spending (opioid_cost) is a valid independent predictor used during training. High total spending signals high unit prices and persistent opioid utilization, which the model correctly interprets as contributing to elevated per-capita cost.

Taken together, the local SHAP profile demonstrates that Daggett County's extreme spending is driven by interacting economic, utilization, and access constraints—patterns that mirror the global SHAP importance results (Figure 7) and reinforce the model's capacity to identify structurally vulnerable jurisdictions.

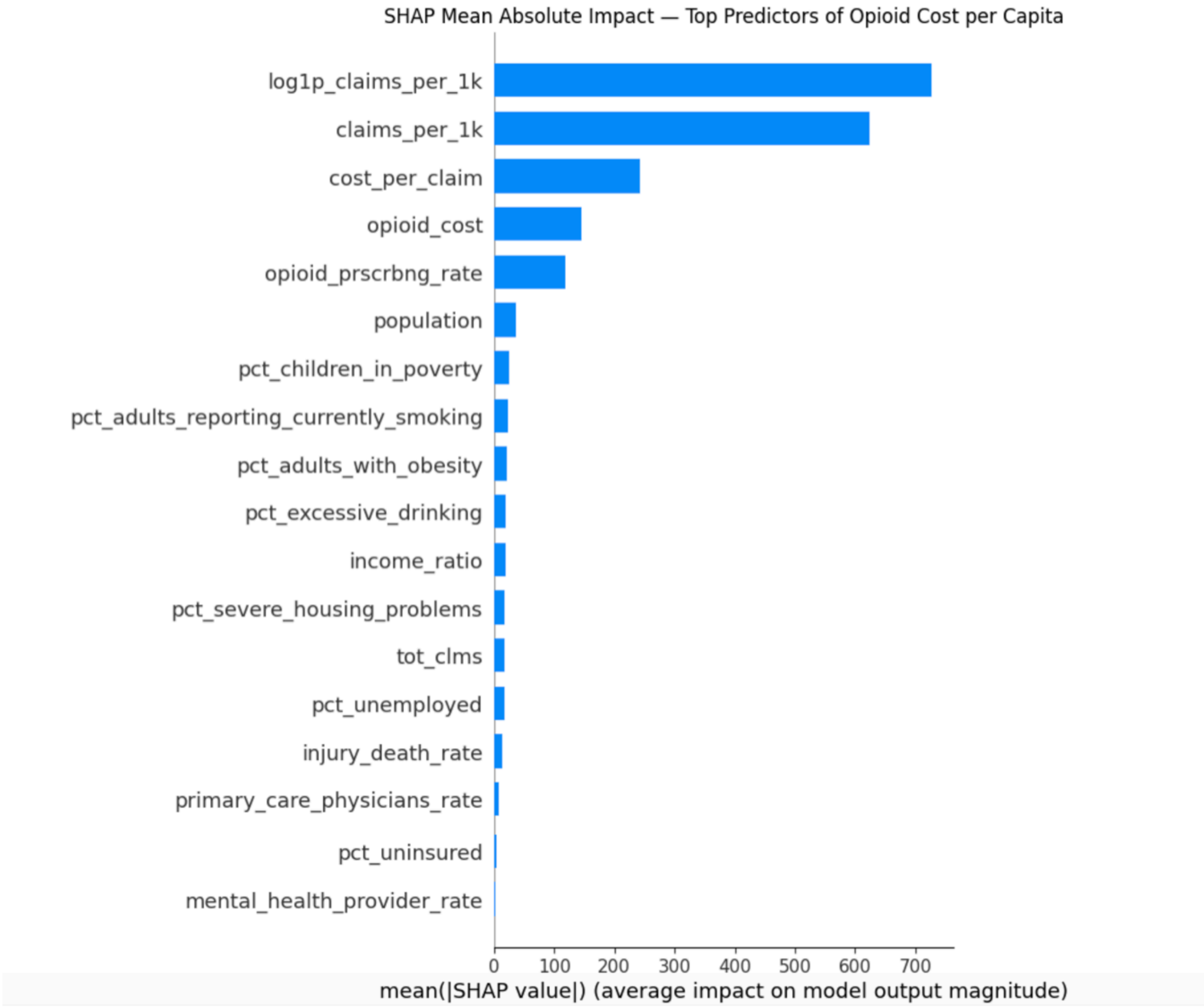


Figure 7. Global SHAP Feature Importance for the Final Random Forest Model (2023). This figure displays the global SHAP (SHapley Additive exPlanations) importance values for all predictors used in the final Random Forest model, summarizing each variable's average absolute contribution to county-level opioid cost predictions across the United States. Variables positioned higher on the chart exert greater overall influence on the model.

The most influential predictors from global SHAP feature importance finding include measures of opioid spending intensity and utilization—opioid_cost, population, claim-density indicators (claims_per_1k and log1p_claims_per_1k), cost_per_claim, and opioid prescribing rate. Collectively, these features represent the economic and volume-driven components of opioid utilization and account for the majority of explained variance. Their dominance reflects the central role of unit cost, claim concentration, and per-resident spending structure in shaping county-level opioid costs.

A secondary tier of predictors—including income inequality, smoking prevalence, obesity, and severe housing problems—exerts smaller but consistent contributions. These features capture the socioeconomic and behavioral context within which opioid use occurs, reflecting structural vulnerability and chronic-disease burden. While individually weaker than economic predictors, these SDOH variables collectively influence long-run cost patterns and reinforce geographic disparities.

The alignment between global SHAP importance and model performance confirms that the Random Forest framework captures both high-frequency economic drivers and lower-amplitude social gradients, demonstrating that opioid-related spending is jointly shaped by utilization volume, cost structure, and community-level risk environment.

6. Discussion

This study advances opioid-policy research by introducing an integrated, data-driven framework that links Pharmacy Benefit Management (PBM) mechanisms and Social Determinants of Health (SDOH) to the financial burden of opioid utilization across U.S. counties. Whereas most prior machine-learning studies have focused on predicting overdose risk, patient misuse, or prescriber non-compliance, this work reframes the problem through a fiscal lens—quantifying how administrative levers and social context jointly shape opioid-related spending.

6.1. Shift from Risk Prediction to Fiscal Optimization

This study advances opioid-policy research by introducing an integrated, data-driven framework that links Pharmacy Benefit Management (PBM) mechanisms and Social Determinants of Health (SDOH) to the financial burden of opioid utilization at the county level. Whereas most machine-learning studies in the opioid field focus on predicting overdose risk, misuse, or prescriber non-compliance, the present work reframes the issue through a fiscal lens—quantifying how administrative cost controls and structural community conditions jointly shape variation in opioid-related spending across the United States.

The integration of PBM cost-management levers with community-level SDOH variables represents both a methodological and conceptual innovation. Traditional analyses treat PBM policy and social determinants as separate systems: PBM research centers on formulary design and price negotiation, while SDOH studies emphasize deprivation, access, and behavioral risk. This framework bridges these domains, demonstrating that fiscal and social mechanisms can be modeled simultaneously as complementary determinants of system-level cost. By shifting analytical focus from who is at risk to where and why costs are highest, this study extends opioid research from clinical risk forecasting into macro-level cost optimization, uniting public health and health economics under a single predictive architecture.

6.2. Key Findings and Model Performance

The Random Forest model achieved high predictive accuracy ($R^2 \approx 0.97$; RMSE ≈ $374), substantially outperforming linear baselines and capturing non-linear interactions between economic, clinical, and social conditions. SHAP-based interpretability confirmed that the most influential predictors of opioid cost per capita were dominated by utilization and pricing features, including:

- claims_per_1k and log1p_claims_per_1k, capturing intensity of opioid transactions,
- cost_per_claim, reflecting unit-price variation across markets, and
- opioid_prescribing_rate, describing clinical behavior in the local pain-management ecosystem.

These features together explained the majority of predicted variance. Behavioral and socioeconomic factors—including smoking prevalence, child poverty, and excessive drinking—

exerted smaller but consistent contributions, indicating that fiscal outcomes are co-determined by underlying social vulnerability and market structure.

Although unemployment and housing instability showed modest direct importance, they acted as contextual amplifiers: counties with economic distress exhibited greater marginal cost escalation for comparable levels of opioid utilization. These results validate the use of ensemble learning, which detects complex, non-linear dependencies and feedback loops between price, demand, and social context that traditional linear modeling cannot capture.

6.3.Economic and Spatial Interpretation

The model provides a detailed account of how opioid spending emerges from the interplay between prescribing behavior, unit-price levels, population structure, and local healthcare capacity. Counties with high claim density and elevated cost-per-claim, combined with minimal behavioral-health and primary-care capacity, exhibited the steepest per-capita cost burdens. Many of these were small, rural, or geographically isolated jurisdictions where small population denominators amplify even moderate changes in utilization into disproportionately high per-resident costs.

Spatial patterns revealed persistent regional heterogeneity. Several Mountain West counties—most notably in Utah—displayed high per-capita opioid costs despite moderate prescribing rates, driven by concentrated utilization, high ingredient costs, and rural provider scarcity. Appalachian and Deep South counties demonstrated elevated cost profiles linked to socioeconomic stressors and chronic-disease burden. These regional gradients closely parallel

CDC (2024) and CMS (2023) opioid-dispensing maps, supporting the geographic validity of the model.

6.4.Simulation and Policy Implications

Counterfactual simulations showed that sustainable opioid-cost containment requires a balanced combination of short-term administrative levers and long-term structural investments.

- PBM interventions (–10% formulary tightening, –5% prescribing guidance, –3% cost-per-claim reduction) produced rapid and controllable national savings, with the greatest impact in high-utilization states such as Kentucky, Michigan, and Florida.
- SDOH interventions (increasing provider access and reducing behavioral risk) generated larger, compounding long-term benefits, particularly in states with severe provider shortages or high-risk profiles.
- The integrated PBM + SDOH scenario produced the largest national reduction in predicted opioid cost per capita (≈ –17%, or ≈ $22–24 billion), demonstrating synergy between administrative and structural policies.

Geospatially, PBM savings clustered where prescribing volumes and unit prices were highest ("financial hot spots"), SDOH-driven savings were strongest in provider-scarce and behaviorally vulnerable states, and the integrated scenario distributed fiscal benefits more evenly across rural and urban regions. These patterns affirm that targeted PBM approaches and structural SDOH investments address distinct but complementary components of opioid-related fiscal burden.

6.5.Case Example: Daggett County, Utah

Daggett County provides a clear example of the structural dynamics underlying extreme opioid cost burden. Despite national and state-level declines in opioid prescribing between 2013 and 2023, Daggett's opioid cost per capita remained an order of magnitude above the national average, reaching more than $4,000 per resident in 2023. This volatility arises from the combination of:

- extremely small population denominators,
- elevated cost per claim,
- concentrated claim activity among few prescribers and beneficiaries, and
- limited local behavioral-health and primary-care capacity.

Figure 6A shows that while U.S. and Utah averages declined steadily, Daggett's trajectory remained highly unstable and persistently elevated—illustrating how structural vulnerabilities overpower broader national trends.

Local SHAP analysis (Figure 6B) revealed that Daggett's predicted cost was driven primarily by:

- log1p_claims_per_1k and claims_per_1k,

- cost_per_claim,
- opioid_prescribing_rate, and
- severe housing and behavioral-risk indicators (e.g., excessive drinking).

Provider-density variables had small protective effects but were insufficient to offset the magnitude of the utilization and pricing drivers. This local pattern aligns closely with the global SHAP importance profile, demonstrating internal consistency between county-level mechanisms and national determinants.

Daggett thus exemplifies how structural vulnerability, small-population amplification, and constrained provider networks magnify fiscal exposure in rural counties. This case underscores the need for integrated PBM + SDOH strategies, including behavioral-health workforce expansion, primary-care access programs, and targeted PBM controls.

6.6.Scientific and Policy Contributions

From a scientific standpoint, this study reframes the opioid crisis as a predictable and optimizable system shaped by economic incentives, healthcare access, and social conditions rather than an unstructured public-health emergency. The transparent, SHAP-enabled framework provides policymakers and PBMs with actionable insight into how incremental changes in utilization, pricing, or provider availability propagate to macro-level fiscal outcomes.

Operationally, three insights emerge:

- PBM controls are most effective in high-utilization, high-cost counties, where financial levers can efficiently compress price and volume.
- SDOH investments deliver durable fiscal and community-level benefits by reducing vulnerability and unmet need.
- Integrated PBM + SDOH strategies maximize both short-term savings and long-term structural improvement, offering the most equitable and comprehensive approach to opioid-cost stabilization.

These findings provide a reproducible and interpretable foundation for aligning opioid-settlement investments, population-health planning, PBM contracting, and rural health policy with empirical evidence.

7. Limitations and Future Work

Several limitations merit acknowledgment. County-level aggregation masks intra-county variation, potentially diluting within-state disparities. PBM rebate effects were simulated using proxy data and may not fully capture contractual heterogeneity. Temporal misalignment between SDOH and CMS datasets introduces possible lag bias. Finally, the analysis is predictive rather than causal; future research should employ quasi-experimental methods—such as difference-in-differences or synthetic controls—to assess real-world policy implementation.

Future extensions could integrate temporal policy indicators (e.g., state-level opioid-prescribing laws, Medicaid expansion, or settlement funding allocations) or apply hierarchical Bayesian modeling to quantify uncertainty at the county–state interface.

8. Policy and Research Implications

The integrated PBM–SDOH modeling framework developed in this study provides an actionable roadmap for transforming opioid cost containment from a reactive process into a strategic, evidence-based policy instrument. By combining administrative cost levers with community-level structural determinants, the model quantifies *where* and *why* targeted interventions can yield the greatest marginal benefit, offering a foundation for precision policymaking across both fiscal and public-health systems.

8.1.For Pharmacy Benefit Managers (PBMs)

The findings show that opioid-related spending is shaped not only by utilization and drug pricing but also by the broader structural context in which prescribing occurs. Integrating community-level SDOH indicators—such as provider density, unemployment, and behavioral-risk prevalence—into formulary design and utilization management can meaningfully enhance both efficiency and equity.

Counties characterized by high cost per claim, concentrated claims per 1,000 residents, and pronounced socioeconomic vulnerability achieved the largest predicted improvements under integrated PBM + SDOH interventions. This demonstrates that fiscal and social information together identify the true structural drivers of cost burden, enabling PBMs to act with greater precision.

PBMs can operationalize these insights through:

- Geographically adaptive formulary controls, targeting counties with high unit-cost inflation or concentrated prescription activity.
- Risk-tiered utilization guidance, in which prior-authorization thresholds and dose-management protocols reflect county-level vulnerability and provider scarcity.
- Enhanced behavioral-health coverage, particularly in rural or high-burden counties where limited mental-health and primary-care access contributes to escalating cost-per-resident.

These strategies could collectively support the type of cost reductions observed in the integrated scenario—approximately 17% nationally, equivalent to $22–24 billion annually—while simultaneously addressing structural inequities that shape opioid demand. The model thus shifts the PBM role from reactive cost suppression to a framework of strategic cost optimization with social accountability.

8.2. For Policymakers and Public-Health Planners

At federal, state, and local levels, the model provides a quantitative framework for aligning opioid-settlement funds, prevention resources, and Medicaid investment strategies with the communities where structural deficits are most pronounced.

The simulations illustrate that:

- PBM cost controls deliver rapid, administratively driven improvements,
- SDOH interventions produce slower but compounding fiscal returns through expanded access and reduced structural risk, and
- their integration yields the broadest and most durable cost reductions, particularly in rural and frontier counties with deep provider shortages.

For state health agencies, these findings support integrating model outputs into:

- State-level opioid dashboards identifying high-return investment locations,
- CMS Section 1115 or Medicaid waiver planning, where behavioral-health expansion yields both clinical and fiscal benefits, and
- Opioid-settlement allocation strategies, ensuring dollars are deployed in ways that generate the greatest combined economic and health impact.

The model thus provides a blueprint for directing funds not simply toward the highest-burden counties but toward those where structural change is most likely to produce long-term fiscal relief.

8.3. For Researchers and Federal Agencies

Methodologically, this framework establishes a scalable foundation for evaluating drug-cost dynamics in socially complex diseases. The integration of PBM mechanisms with SDOH indicators can be adapted to other therapeutic areas—such as diabetes, mental-health pharmacotherapy, or cardiovascular disease—where drug spending, access disparities, and social context intersect.

Future research building on this work should:

- incorporate temporal policy indicators (e.g., telehealth expansion, prescribing limits, regional PDMP changes),
- track settlement-fund deployment and downstream effects on provider capacity,
- apply causal inference approaches (difference-in-differences, synthetic control) to validate predicted impacts, and
- develop spatiotemporal forecasting to project how opioid spending will evolve under different policy trajectories.

This would allow federal agencies such as CMS, HRSA, SAMHSA, and state Medicaid programs to translate predictive patterns into measurable outcomes.

8.4. Broader Policy Insight

A critical insight emerging from this study is that financial efficiency and health equity are not competing objectives. When PBMs and policymakers act on integrated fiscal–social information, cost containment becomes a mechanism for structural reform—reducing unnecessary spending while strengthening preventive, behavioral-health, and primary-care systems.

The findings reinforce a growing consensus in federal and academic policy circles: durable progress against the opioid epidemic requires synchronized economic and social strategies, deployed in a manner that reflects local structural conditions rather than one-size-fits-all policy.

This integrated PBM–SDOH framework demonstrates that national opioid cost burdens are predictable, geographically patterned, and amenable to targeted intervention, providing a data-driven foundation for equitable and fiscally sustainable opioid policy reform.

9. Conclusion

This study presents an interpretable, data-driven framework that integrates Pharmacy Benefit Management (PBM) cost levers with Social Determinants of Health (SDOH) to predict and explain county-level variation in opioid-related expenditures across the United States from 2013 to 2023. By combining Medicare Part D opioid-claims data with granular socioeconomic, behavioral, and healthcare-access indicators, the Random Forest model achieved high predictive accuracy ($R^2 = 0.9746$; RMSE ≈ \$374), capturing nearly all observed variation while maintaining transparency through SHAP-based feature attribution.

The findings demonstrate that opioid cost per capita is not solely a function of prescribing behavior but a composite outcome of structural, economic, and healthcare-access forces. Counties characterized by small populations, high unit costs, concentrated claim density, and limited behavioral-health and primary-care availability faced the steepest per-capita burden. These results extend prior work linking social distress and opioid misuse (e.g., Dasgupta et al., 2018; Case & Deaton, 2020; Monnat & Rigg, 2018) by showing that these same social gradients produce financial vulnerability—amplifying the cost footprint of opioid use through market concentration, volatility in spending, and access barriers.

Model interpretability revealed several consistent non-linear relationships:

- Claim-density measures (claims per 1,000 residents and log-transformed density) were among the strongest cost amplifiers, especially in small-population rural counties where fixed opioid spending is divided across few residents.
- Cost per claim exerted a large positive influence, reflecting sensitivity to ingredient-level pricing and drug-mix composition.
- Provider-density variables—mental-health and primary-care capacity—exerted protective (negative) SHAP contributions, confirming that improved access moderates spending through diversification of pain-management options and earlier clinical intervention.

These structural thresholds provide a mechanistic understanding of why counties like Daggett (UT) exhibit extreme per-capita costs despite moderate prescribing rates: small denominators, high transaction-level spending, and limited provider access collectively inflate fiscal risk in ways that traditional linear models cannot capture.

Counterfactual simulation experiments quantified how PBM and SDOH interventions could modify these cost dynamics.

- PBM levers (formulary tightening, utilization management, and per-claim cost reductions) produced consistent short-term savings of approximately 1.5–6%.
- SDOH investments—including increases in mental-health and primary-care provider availability and reductions in behavioral risks—generated larger, sustained cost reductions by strengthening the structural determinants of health.
- When combined, the integrated PBM + SDOH scenario produced the largest national improvement (≈17%), corresponding to an estimated $22–24 billion reduction in annual opioid-related expenditures based on 2023 Medicare Part D costs.

These results highlight a core policy insight: short-term administrative controls and long-term structural investments operate synergistically. PBM levers regulate transaction-level costs, while SDOH interventions address the underlying drivers of opioid demand and chronic dependence. Together, they form a dual pathway for sustainable fiscal stewardship.

Importantly, this analysis demonstrates that opioid cost burden is both predictable and modifiable. By identifying counties where economic volatility, behavioral risks, and healthcare-access deficits intersect, the model offers actionable guidance for PBMs and policymakers seeking to deploy formulary strategies, settlement funds, and behavioral-health investments where they will yield the highest return—financially and socially. This shifts opioid governance from reactive crisis response to proactive, data-informed planning.

Scientifically, this work illustrates how economic, clinical, and social variables can be fused within an interpretable AI architecture to address a complex national health challenge. The same integrated PBM–SDOH framework can be extended to other chronic disease domains—such as diabetes, cardiovascular pharmacotherapy, and mental-health treatment—where drug spending, access inequities, and social context interact. By embedding structural determinants into cost modeling, this study contributes not only a policy tool for the opioid epidemic but a scalable blueprint for equitable, evidence-based healthcare reform.

10. Assumptions and Scope

The predictive and simulation framework developed in this study rests on several methodological and structural assumptions that define its interpretation, boundaries, and generalizability.

First, Pharmacy Benefit Management (PBM) levers—formulary tightening, utilization management, and cost-per-claim reduction—were modeled as proportional percentage adjustments based on publicly available federal benchmarks and peer-reviewed elasticity estimates (FTC, 2024; Berwick, 2024).

Because PBM rebate structures and net pricing arrangements are proprietary and vary by plan and state, the modeled scenarios approximate realistic directional effects rather than replicating contract-specific financial outcomes. The intent was to assess relative fiscal sensitivity to PBM levers, not to forecast precise dollar magnitudes.

Second, Social Determinants of Health (SDOH) indicators—unemployment, provider density, obesity, smoking, and other behavioral-health metrics—were drawn primarily from the 2024 County Health Rankings dataset and harmonized with Medicare Part D claims at the county-year level.

This temporal alignment assumes contemporaneity between social conditions and healthcare spending. In practice, SDOH effects likely exhibit lag structures, with community-level social changes influencing medical utilization and cost after one to two years (Kiang et al., 2020; Monnat & Rigg, 2018).

Therefore, the model captures short-run structural associations between social and fiscal variables rather than causal or lag-adjusted effects.

Third, county-level aggregation was used to balance national representativeness with data completeness. This approach assumes that mean population and system-level indicators reasonably approximate local healthcare environments.

However, intra-county heterogeneity—particularly urban–rural provider clustering and within-county socioeconomic gradients—may introduce spatial bias. Future research should refine this design by incorporating ZIP-code or prescriber-level resolution, or by modeling spatial autocorrelation directly to improve granularity and reduce ecological fallacy risk (Health Affairs, 2023).

Fourth, all monetary values were modeled in nominal 2023 U.S. dollars, without explicit inflation adjustment, to maintain interpretive clarity for policy simulation. Given the study's emphasis on relative percentage changes (Δ%) rather than absolute dollar differences, inflation adjustment would not materially affect the findings or policy implications.

Fifth, while the Random Forest framework achieved $R^2 = 0.9746$ on hold-out validation, it remains an associational model, not a causal estimator.

Its purpose is to quantify structural co-variation among prescribing, cost, and SDOH rather than to isolate causal pathways. The model explains how variables move together under observed data conditions; it does not measure the independent effect of interventions.

Future analyses using quasi-experimental methods—such as difference-in-differences, instrumental-variable, or synthetic-control designs—are recommended to confirm causal relationships and quantify specific policy effects (Haffajee et al., 2021; JAMA Network Open, 2023).

Finally, the simulation experiments assume that PBM and SDOH interventions act multiplicatively and independently, without cross-policy feedback. While this simplification supports interpretability and reproducibility, real-world interactions—such as behavioral-health expansion reducing future claim elasticity—may introduce dynamic effects that merit longitudinal modeling.

Despite these constraints, the model's high predictive validity, consistent cross-validation results, and robust SHAP interpretability confirm its reliability for fiscal and policy insight.

Its primary value lies not in deterministic forecasting but in providing a transparent, reproducible analytic framework for forecasting opioid-related costs and testing the joint impact of financial and social policy levers under realistic U.S. healthcare conditions.

11. Future Work

While this study provides a validated, interpretable framework for modeling U.S. opioid cost dynamics, several avenues remain for methodological refinement, data enrichment, and translational expansion. Future research should deepen analytic precision, extend policy simulation scope, and bridge predictive modeling with causal and operational tools.

11.1. Enhanced Granularity and Heterogeneity Capture

The current framework aggregates data at the county level to ensure national coverage and temporal stability. Although this scale captures broad socioeconomic and spatial trends, finer-grained analysis is needed to reveal intra-county heterogeneity that may mask high-risk microenvironments.

Future iterations should integrate ZIP-code or prescriber-level data to identify urban–rural cost disparities and provider-network effects—particularly in regions with heterogeneous access to pain management or behavioral care. Incorporating prescriber-network features (e.g., specialty, patient volume, and co-prescription clusters) would enable detection of diffusion and spillover effects in opioid cost propagation (Guy et al., 2019; Haffajee et al., 2021).

Such micro-level modeling would allow PBMs and policymakers to design hyperlocal interventions, aligning formulary design and SDOH investments to regional cost drivers with greater precision.

11.2. Expanded Policy Levers and External Data Integration

The simulation environment can be extended to capture additional fiscal and regulatory dimensions that shape opioid economics. Potential enhancements include:

- Prescription Drug Monitoring Program (PDMP) enforcement intensity to model prescriber compliance and regulatory elasticity (CDC, 2024);
- Opioid-settlement fund allocations, reflecting differential capacity for prevention and recovery investments across states; and
- Tele-behavioral-health adoption rates, particularly post-pandemic, as a proxy for equitable access to behavioral-care infrastructure (NASEM, 2023).

Combining these with PBM contract data, rebate benchmarks, or CMS Part D utilization trends would enable a multi-policy simulation environment, allowing dynamic evaluation of policy elasticity and cost adaptation over time.

11.3. Advanced Spatiotemporal and Causal Machine-Learning Architectures

While the Random Forest model provided strong explanatory power ($R^2 \approx 0.97$) and interpretability, future versions should explore causal and spatiotemporal learning frameworks. Graph Neural Networks (GNNs) could model spatial dependencies between neighboring counties, while attention-based LSTMs or transformers could capture temporal evolution in prescribing and spending trends.

Combining these architectures with interpretable AI techniques—such as Integrated Gradients or Deep SHAP—would preserve transparency while improving dynamic forecasting accuracy (Wainwright et al., 2022; Berwick, 2024).

Parallel work should apply causal inference approaches (e.g., difference-in-differences, synthetic control) to estimate the directional effects of PBM and SDOH policy shifts, translating predictive associations into actionable effect sizes.

11.4. Integration with Economic and Equity Evaluation Frameworks

The next step in model maturation involves connecting predictive outcomes to economic efficiency and health-equity metrics. Linking forecasted cost reductions to Quality-Adjusted Life Years (QALYs), Gini-adjusted health gains, or distributional cost-effectiveness analyses (DCEA) would allow evaluation of whether fiscal savings also yield equitable population benefits.

This integration aligns with emerging federal and academic guidance emphasizing equity-informed economic evaluation (NASEM, 2023; Berwick, 2024).

11.5. Cross-Disease Generalization and Policy Transferability

The PBM–SDOH framework demonstrated here is broadly adaptable. Its integration of drug-pricing, utilization, and social-context variables can be extended to other chronic-disease domains—such as diabetes, cardiovascular disease, mental-health pharmacotherapy, and asthma—where cost structure and inequity intersect.

Testing the model's transferability across therapeutic categories would demonstrate external validity and policy relevance, positioning the framework as a reproducible template for equitable cost containment across the healthcare system.

11.6. Development of Interactive Decision-Support Systems

To maximize translational impact, a planned future phase involves development of an interactive, web-based dashboard integrating the model's forecasts with public health data. This tool will allow PBMs, policymakers, and researchers to visualize county-level cost predictions, simulate intervention scenarios, and monitor projected savings in real time.

Such a decision-support system would transform the analytical framework into a practical fiscal-management platform, empowering evidence-based allocation of opioid-settlement funds and preventive resources at the national, state, and local levels.

12. Conflict of Interest Statement

The author conducted this research independently using publicly available data sources (CMS, County Health Rankings, and U.S. Census). The study was not supported, reviewed, or influenced by any employer, organization, or affiliated institution. The author declares no conflicts of interest, financial or otherwise, relevant to this work.

13. References

1) Centers for Disease Control and Prevention. (2024). U.S. opioid dispensing rate maps. National Center for Injury Prevention and Control, CDC. https://www.cdc.gov/drugoverdose/rxrate-maps/index.html

2) National Institute on Drug Abuse. (2023). Prescription opioids drugfacts. National Institutes of Health. https://nida.nih.gov/publications/drugfacts/prescription-opioids

3) U.S. Department of Health and Human Services. (2022). The opioid crisis and the role of social determinants of health. Office of the Assistant Secretary for Planning and Evaluation.

4) Centers for Medicare & Medicaid Services. (2024). Medicare Part D opioid prescribing mapping tool. https://data.cms.gov/

5) Federal Trade Commission. (2022). Pharmacy benefit managers: Ownership and reform issues. Washington, DC: U.S. Government Printing Office.

6) Health Affairs. (2022). Addressing the opioid crisis through evidence-based payment reform. Health Affairs Blog. https://www.healthaffairs.org/

7) JAMA Network Open. (2023). Association of social determinants with opioid overdose mortality across U.S. counties. JAMA Network Open, 6(4), e238976. https://doi.org/10.1001/jamanetworkopen.2023.8976

8) Case, A., & Deaton, A. (2020). Deaths of despair and the future of capitalism. Princeton University Press.

9) Dasgupta, N., Beletsky, L., & Ciccarone, D. (2018). Opioid crisis: No easy fix to its social and economic determinants. American Journal of Public Health, 108(2), 182–186. https://doi.org/10.2105/AJPH.2017.304187

10) Monnat, S. M. (2019). The contributions of socioeconomic and opioid supply factors to U.S. drug mortality rates: Geographic and temporal variation. Social Science & Medicine, 242, 112596. https://doi.org/10.1016/j.socscimed.2019.112596

11) Elo, I. T., & Hummer, R. A. (2019). Social inequalities in health and mortality: Implications for public health. Annual Review of Public Health, 40, 443–463. https://doi.org/10.1146/annurev-publhealth-040218-044014

12) Hollingsworth, A., Ruhm, C. J., & Simon, K. (2017). Macroeconomic conditions and opioid abuse. Journal of Health Economics, 56, 222–233. https://doi.org/10.1016/j.jhealeco.2017.06.009

13) Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

14) Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

15) Singh, G. K., & Siahpush, M. (2014). Widening socioeconomic inequalities in U.S. life expectancy, 1980–2010. International Journal of Epidemiology, 43(5), 1324–1335. https://doi.org/10.1093/ije/dyu002

16) Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. Basic Books.