Mixed-Integer Predictive Optimization for Opioid Settlement Allocation in West Virginia

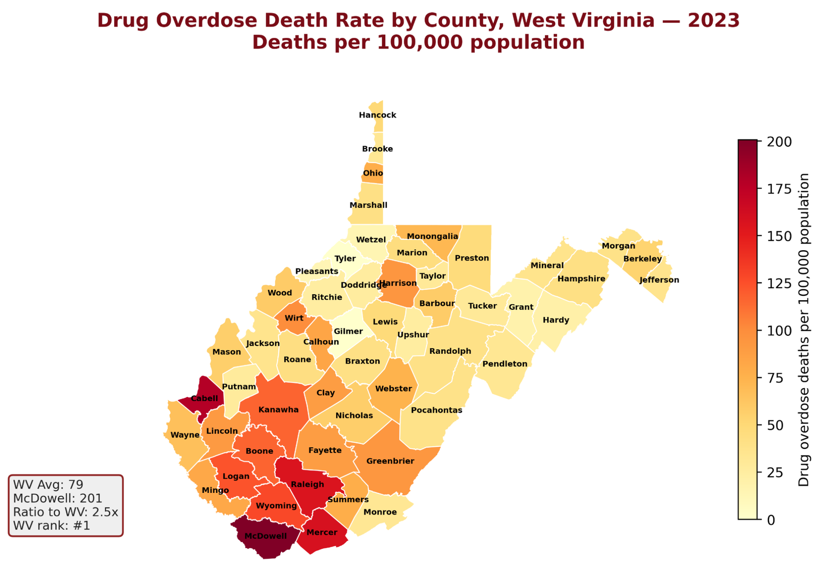
Bouchra Er-Rabbany* (errab003@umn.edu) and Prof. Jean Philippe Richard (jrichar@umn.edu)   Dept. of Industrial & Systems Engineering, University of Minnesota–Twin Cities

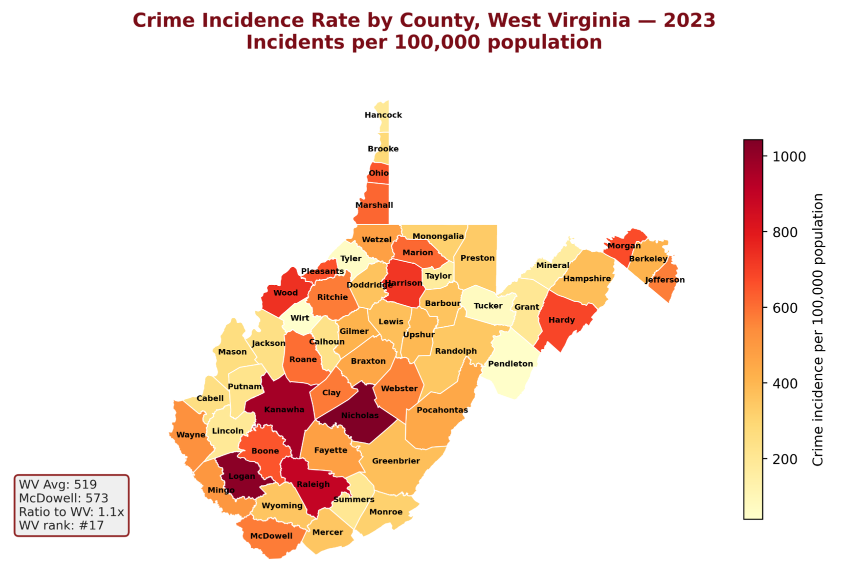
Joint work with A. Bansal, G. Smith, H. Brownstein, and S. Workman

Motivation & The Crisis in West Virginia

West Virginia has 55 counties and remains one of the states most affected by the opioid crisis. As new public resources from state settlement funds become available, the state must decide how to allocate funds across counties and spending categories. This creates a difficult tradeoff: reducing predicted harm while maintaining feasibility, and policy relevance.

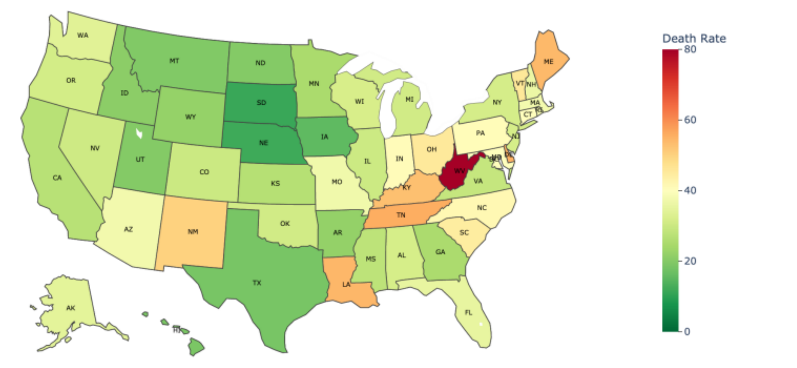
Research Question.
How should additional state resources be allocated across counties and spending categories to minimize the statewide predicted drug deaths and crime incidence?





Notation: $i \in C$ county, $j \in J$ expenditure category.

Expenditure Categories	Challenges
Capital Projects	Heterogeneous counties
Culture & Recreation	Endogenous responses
General Government	Efficiency vs. fairness
Health & Sanitation	Non-linear prediction
Public Safety	Feasibility
Social Services	



Drug Overdose Mortality by State in 2022 (Centers for Disease Control and Prevention, 2025)

Integrated Predictive Optimization Pipeline

Three-layer MIP framework: historical data → learned models → integrated optimization model.

Historical County Data
55 counties · 2012–2023 · Expenditure + socioeconomic features · drug deaths · crime

↓

Learned Spending Allocation
Raw spending scores s_{ij} learned from lagged budget data and county features

↓

Projection / KKT Mechanism
$\widetilde{x}_{ij}$ = feasible projected budgets satisfying balance and lower-bound constraints

↓

Harm Prediction Trees
Drug deaths · Crime incidence

↓

MIP Solver (Gurobi)
100k vars · 98% binaries · 200k constraints

↓

Optimal Allocation X^*
County-category allocation minimizing predicted statewide harm

Data & Predictive Models

We train predictive models for drug deaths, crime incidence, and county spending patterns.

Model	Target	Key Inputs	Role
Drug Boosted Tree	Drug deaths	$\widetilde{x}_{ij}$ + socioeconomic features	Objective
Crime Boosted Tree	Crime incidence	$\widetilde{x}_{ij}$ + socioeconomic features	Objective
Linear Regressions	Spending pattern	Lagged budgets + socioeconomic features	Learned expenditures

Fixed socioeconomic features: Population · Year · Uninsured rate · Food stamp recipient ratio · rent-to-income ratio· Unemployment rate

Expenditure-response model:

$$s_{ij} = \widehat{\alpha}^j b_i + \widehat{\beta}^j b_i^{t-1} + \widehat{\gamma}^j b_i^{t-2} + R_{ij}(\text{Socio-Economic}), \quad \forall i \in C,\ j \in J.$$

Key insight: Decision variables $\widetilde{x}_{ij}$ appear directly as predictors inside the boosted-tree harm models.

Data	Source
Socioeconomic features	American Community Survey (ACS), U.S. Census Bureau
County expenditures	WV State Auditor's Office, Local Government Division
Drug deaths	West Virginia Forensic Drug Database, WVU Health Sciences Center
Crime incidents	West Virginia Incident-Based Reporting System / FBI NIBRS

Resource Allocation Model

Decision variables: $x_{ij} \geq 0$ new state allocation; $b_i = \bar{b}_i + \sum_{j \in J} x_{ij}$ updated budget; $\widetilde{x}_{ij}$ projected category budget.

$$\min \sum_{i \in C} \alpha_D D_{\text{Forest}}(\widetilde{x}_{i1}, \ldots, \widetilde{x}_{i6}, \text{socio}_i) + \sum_{i \in C} \alpha_C C_{\text{Forest}}(\widetilde{x}_{i1}, \ldots, \widetilde{x}_{i6}, \text{socio}_i)$$

s.t.

Resource capacity	$\sum_{i \in C}\sum_{j \in J} x_{ij} \leq B,$		(1)
Fairness	$l_i \leq \frac{\sum_{j \in J} x_{ij}}{p_i} \leq u_i,$	$\forall i \in C$	(2)
Budget update	$b_i = \bar{b}_i + \sum_{j \in J} x_{ij},$	$\forall i \in C$	(3)
Learned expenditures	$s_{ij} = \widehat{\alpha}^j b_i + \widehat{\beta}^j b_i^{t-1} + \widehat{\gamma}^j b_i^{t-2} + \text{R}_{ij}(\text{Socio-Economic}),$	$\forall i \in C, j \in J$	(4)
Projection	$\sum_{j \in J} \widetilde{x}_{ij} = b_i,$	$\forall i \in C$	(5)
	$\widetilde{x}_{ij} \geq \bar{x}_{ij},$	$\forall i \in C, j \in J$	(6)
	$\mu_{ij} \geq 0,$	$\forall i \in C, j \in J$	(7)
	$2(\widetilde{x}_{ij} - s_{ij}) + \mu_i^0 - \mu_{ij} = 0,$	$\forall i \in C, j \in J$	(8)
	$\mu_{ij}(\widetilde{x}_{ij} - \bar{x}_{ij}) = 0,$	$\forall i \in C, j \in J$	(9)
Nonnegativity	$x_{ij} \geq 0.$	$\forall i \in C, j \in J$	(10)

where $\alpha_D = w_D/R_D$, $\alpha_C = w_C/R_C$, with $w_D = w_C = 0.5$.

Complementarity (9): modeled via SOS1 constraints $\{\mu_{ij}, \widetilde{x}_{ij} - \bar{x}_{ij}\} \in \text{SOS1}$.

Projection Mechanism

Projection-Based Budget Adjustment via KKT. The learned category models produce raw scores s_{ij}, but these scores may not satisfy budget balance or minimum-allocation requirements. We compute a projected allocation $\widetilde{x}_{ij}$ that is close to s_{ij} while remaining feasible.

Projection problem:

$$\min_{\widetilde{x}} \sum_{i \in C}\sum_{j \in J}(\widetilde{x}_{ij} - s_{ij})^2$$
$$\text{s.t.} \sum_{j \in J} \widetilde{x}_{ij} = b_i, \qquad \forall i \in C,$$
$$\widetilde{x}_{ij} \geq \bar{x}_{ij}, \qquad \forall i \in C, j \in J.$$

KKT conditions in the MIP:

$$2(\widetilde{x}_{ij} - s_{ij}) + \mu_i^0 - \mu_{ij} = 0,$$
$$\sum_{j \in J} \widetilde{x}_{ij} = b_i,$$
$$\widetilde{x}_{ij} \geq \bar{x}_{ij}, \quad \mu_{ij} \geq 0,$$
$$\mu_{ij}(\widetilde{x}_{ij} - \bar{x}_{ij}) = 0.$$

Raw/projected values
- s_{ij}: raw score from the expenditure model.
- $\widetilde{x}_{ij}$: final feasible category budget.

Feasibility quantities
- b_i: total county budget after allocation.
- $\bar{x}_{ij}$: lower bound / baseline protection.

The projection step transforms learned spending scores into feasible county-category budgets before they enter the harm prediction models.

Boosted-Tree MIP Embedding

Outcome models D and C are embedded directly in the MIP through binary leaf and threshold variables.

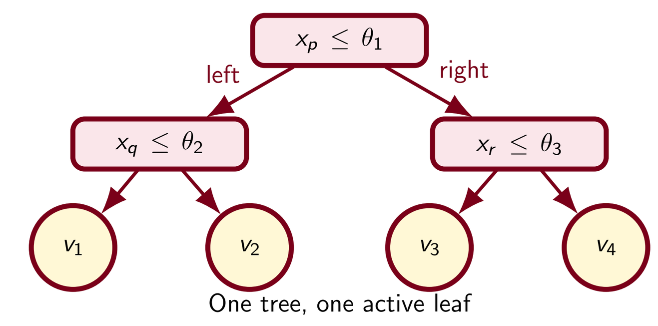
Tree-embedding decision variables:

$$y_{it\ell}^Y \in \{0,1\} : \text{leaf } \ell \text{ selected in tree } t \text{ for county } i, \qquad z_{ipr}^Y \in \{0,1\} : x_p \leq \theta_{pr}^Y, \quad Y \in \{D, C\}.$$

$$\widehat{Y}_i = h_0^Y + \eta^Y \sum_{t \in T^Y} \sum_{\ell \in L_t^Y} v_{t\ell}^Y y_{it\ell}^Y.$$

Constraint	Formula	Meaning
Leaf selection	$\sum_{\ell \in L_t^Y} y_{it\ell}^Y = 1$	One active leaf per tree
Monotonicity	$z_{ip1}^Y \leq z_{ip2}^Y \leq \cdots$	Valid threshold order
Left-path	$y_{it\ell}^Y \leq z_{ipr}^Y$	Left branch: $x_p \leq \theta_{pr}^Y$
Right-path	$y_{it\ell}^Y \leq 1 - z_{ipr}^Y$	Right branch: $x_p > \theta_{pr}^Y$
Link upper	$x_p \leq \theta_{pr}^Y + M(1 - z_{ipr}^Y)$	If $z = 1$, then $x_p \leq \theta$
Link lower	$x_p \geq \theta_{pr}^Y + \epsilon - M z_{ipr}^Y$	If $z = 0$, then $x_p \geq \theta + \epsilon$

Leaf variables select one tree region, while threshold and linking constraints make the selected path consistent with the continuous feature values.

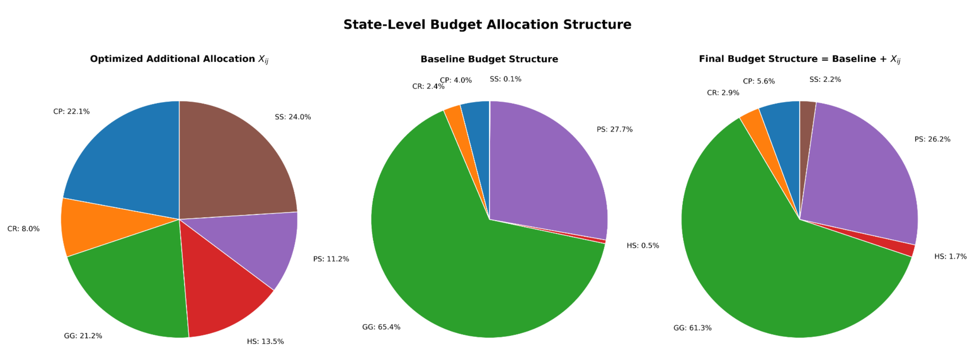


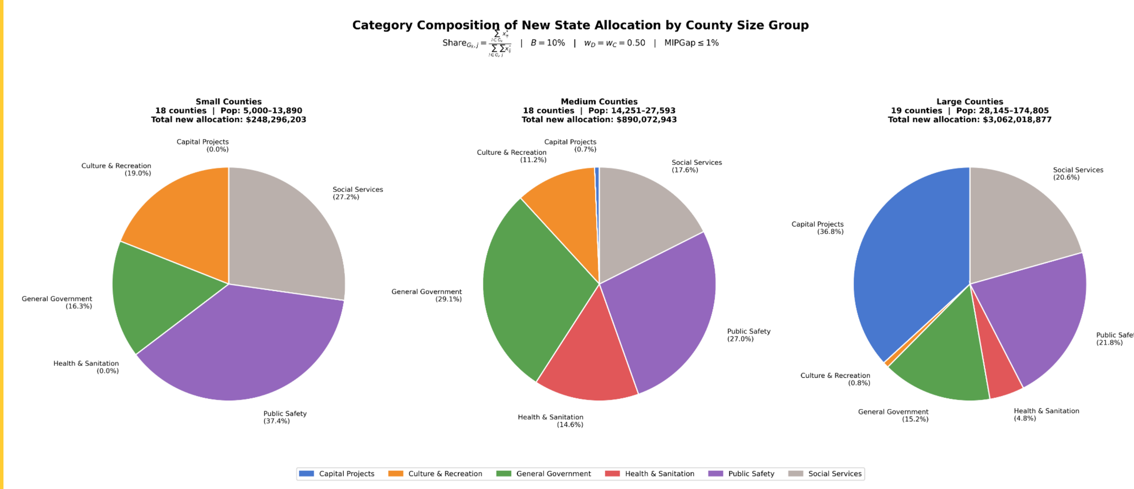
Each feasible path activates one terminal leaf value.

Budget Allocation Results: State & County Level

Resource Distribution Additional budget vs. baseline vs. final budget

- New resources are more diversified than the baseline budget.
- Final budget stays close to baseline.
- Optimized additions shift priorities without replacing existing budgets.



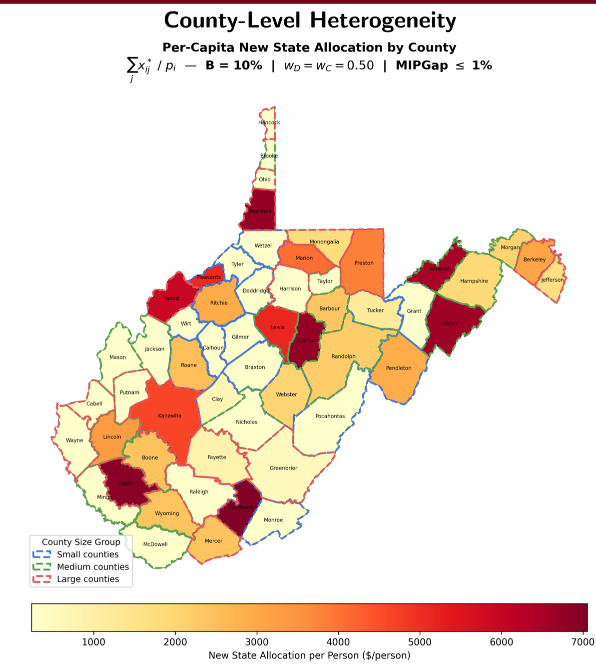


- Small counties prioritize Public Safety and Social Services.
- Medium counties receive the most balanced category mix.
- Large counties receive a larger Capital Projects share.

Optimized resources are heterogeneous across both counties and spending categories.

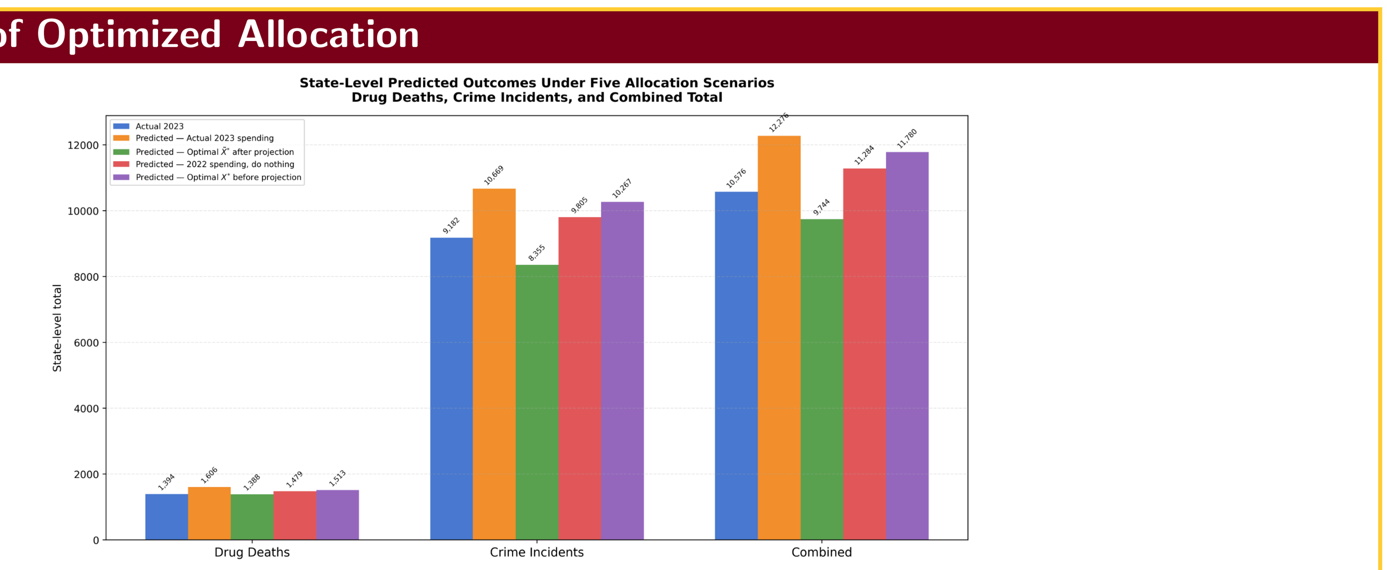
Budget Allocation Results: County Level

County-Level Heterogeneity



Per-capita new state allocation
- The optimized policy is spatially heterogeneous.
- High allocations appear in counties of different sizes.
- The model is not simply allocating by population.

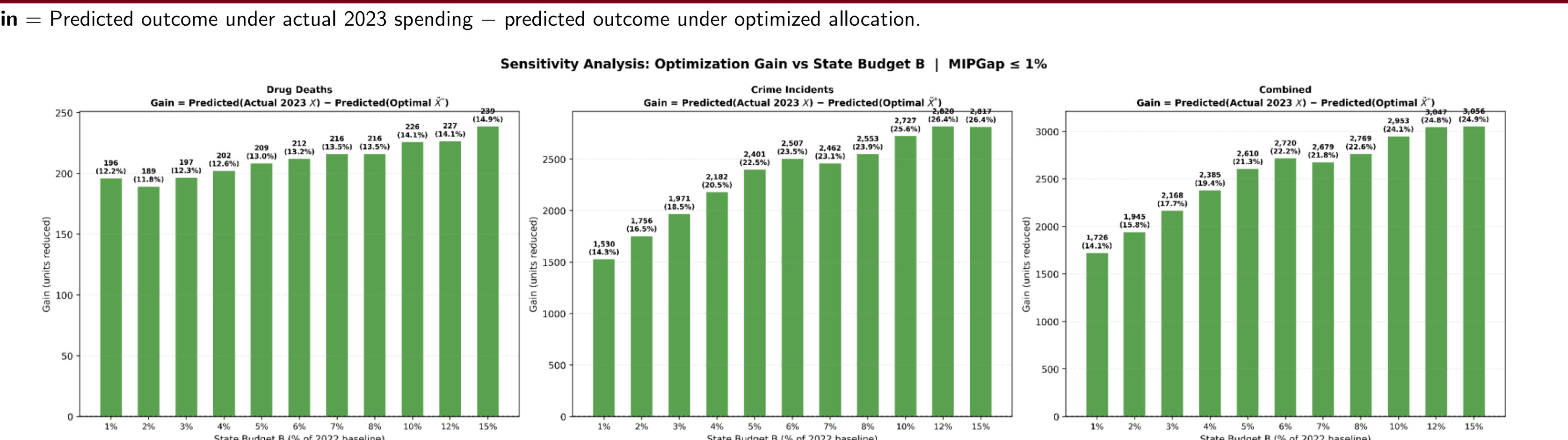
Predicted Impact of Optimized Allocation



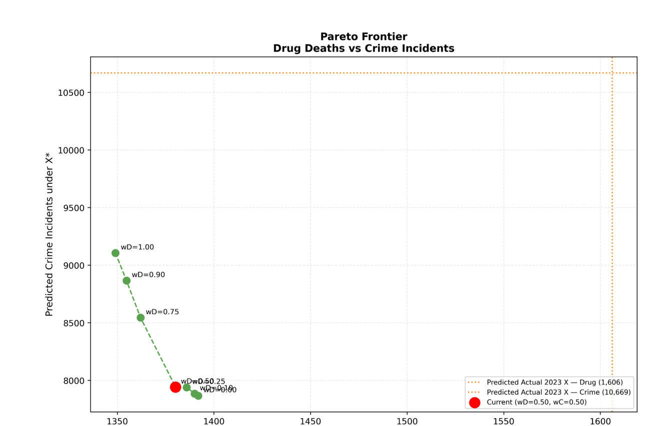
Optimized allocation reduces predicted combined harm from 12,276 to 9,744, mainly by lowering predicted crime incidence; drug deaths decrease modestly. These are model-predicted gains, not causal effects.

Sensitivity Analysis

Gain = Predicted outcome under actual 2023 spending − predicted outcome under optimized allocation.



- More available resources produce larger predicted statewide gains.
- Predicted gains remain positive even with limited additional resources.



Sensitivity confirms the policy tradeoff between scale of funding and objective priorities.

Two-Stage Extension Model

Leader: state chooses new allocations x_{ij}. Followers: counties choose feasible category budgets $\widetilde{x}_{ij}$ after local response.

Leader problem

$$\min \sum_i p_i F_i(\widetilde{x}_{ij}, \text{socio-economic})$$
$$\text{s.t.} \sum_i \sum_j x_{ij} \leq B,$$
$$\widetilde{x}_{ij} = \bar{x}_{ij} + x_{ij}.$$

Follower / county response problem

$$\min \ F_i(\widetilde{x}_{ij}, \text{socio-economic})$$
$$\text{s.t.} \sum_j \widetilde{x}_{ij} \leq b_i.$$

Next step: replace the current single-level response with a leader–follower model that captures strategic county behavior while preserving the tree-based harm predictors.

MIP Workshop 2026, University of Connecticut, Stamford | * Supported by NSF CMMI 2240361