PNAS
www.pnas.org

Supplementary Information for

Predicting High-Risk Opioid Prescriptions Before they are Given

Justine S. Hastings, Mark Howison, Sarah E. Inman

Justine S. Hastings.
E-mail: connect@ripl.org

This PDF file includes:

Supporting Information Text

1. Experimental Design

Our objective was to define a panel of Rhode Island Medicaid recipients who received an initial opioid prescription under Medicaid coverage; define adverse outcomes of opioid dependence, abuse, or poisoning; and model and assess the accuracy of predictions of adverse outcomes using information known only prior to the initial prescription. Data were split into randomly-sampled training, validation, and testing sets using the ratio 50:25:25 at the beginning of the study. We report the results of model predictions on the testing set, which was withheld from analysis prior to the preparation of the manuscript.

Data are from the period 2005-2017, and include Rhode Island administrative records from the Department of Human Services (DHS), Department of Labor and Training (DLT), Department of Corrections (DOC), Medicaid program (under the Executive Office of Health and Human Services), and police agencies (including the Rhode Island State Police and eight municipal police departments).

Although our data span the years 2005 to 2017, we construct a panel of individuals with initial prescriptions between 2006 and 2012 to allow for the construction of variables a year before the initial prescription and to define outcomes up to five years after the initial prescription.

A. Data Availability. Data are available through individual data sharing agreements with each of the following Rhode Island agencies and municipal police departments: RI Department of Corrections, RI Department of Labor and Training, RI Executive Office of Health and Human Services, RI State Police, Central Falls Police Department, Cranston Police Department, Cumberland Police Department, Middletown Police Department, Narragansett Police Department, Providence Police Department, Warwick Police Department, Woonsocket Police Department.

2. Panel and Outcome Definitions

A. Opioid Prescriptions. To define our panel, we first establish which pharmacy claims correspond to opioid prescriptions. The primary identifier for the dispensed drug is a standardized 11-digit National Drug Code (NDC) from the U.S. Food and Drug Administration's NDC Directory (1). This directory is only available as a current snapshot, and because our claims data start in 2005, there are many unmapped NDCs to the current directory. Out of approximately 14.8 million pharmacy claims between 2006 and 2012, only 66.8 percent join to the current directory. Therefore, we construct a historical NDC directory using a data mining framework that downloads and collates all available Internet Archive snapshots of the FDA's NDC website since 2000 (2). This historical directory also includes full ingredient lists for each NDC, standardized to milligrams. Using this improved directory, 88.1 percent of pharmacy claims between 2006 and 2012 map to an NDC entry.

We define an opioid prescription as any claim for a drug containing an opioid ingredient at or above the recommended starting dose when initiating opioid therapy for chronic pain management, as established in Washington State's 2015 prescribing guideline and further cited in the Centers for Disease Control's 2016 prescribing guideline (3, 4). Table S10 lists these ingredients and the minimum amounts we use to define an opioid drug. Of the 4,359 drugs containing one of these ingredients, 4,175 meet the minimum threshold amount and appear in 3.9 percent of claims.

Additionally, we define a recovery prescription as any NDC containing one of four ingredients commonly used in medication-assisted treatment of an opioid use disorder, which identifies 412 such drugs that appear in 0.5 percent of claims. These prescriptions may indicate that an individual has a pre-existing opioid use disorder.

B. Opioid Injections. In addition to identifying opioid prescriptions in pharmacy claims, we also identified procedure codes for epidural or intraveneous opioid injections in all of the Medicaid claims. Table S11 lists these codes, which were identified by searching the descriptions of all procedure codes for each of the opioid ingredients in the Washington State prescribing guidelines used to identify prescription drugs.

C. Outcomes. For each individual in our panel, we examine all of the Medicaid claims following their initial opioid prescription to construct indicator variables for four types of adverse outcomes: opioid dependence, opioid abuse, prescription-opioid poisoning and heroin poisoning. We include heroin poisoning as an outcome given the increasing use of heroin among those who abuse opioids, and the high proportion (greater than 80 percent) of joint heroin-prescription-opioid users who abused opioids prior to using heroin (5).

We determine these outcomes from the claim's International Classification of Diseases (ICD) diagnosis codes, which are used by medical professionals to classify a patient's health conditions following an encounter. Because our data span the transition from the ICD-9 to ICD-10 classification, we include diagnosis codes from both. Table S12 lists the codes used to indicate each of these four diagnosis-related outcomes.

Not everyone with an opioid use disorder receives a diagnosis code (6, 7). Therefore, we define a fifth treatment outcome using procedure codes related to the treatment of opioid use disorder, and more generally for drug rehabilitation and detoxification (see "Treatment" in Table S12). Finally, we define a sixth "any" outcome as the union of any of the diagnoses or treatment outcomes, to capture as broad a population of individual with opioid use disorder as possible. Data and measurement limitations notwithstanding, our model demonstrates that administrative data can be combined to form an accurate prediction of these outcomes, suggesting a feasible path forward for utilizing data to inform prescription risk. Figure S1 shows the accumulating fraction of adverse outcomes over the five year period following initial prescription.

D. Final Panel. Out of 400,024 distinct Medicaid enrollees between 2006 and 2012, our panel initially contains 85,377 individuals who received at least one opioid prescription or injection in that period. We exclude 500 individuals who received a recovery prescription before their initial opioid prescription or injection, since this indicates they may have been seeking treatment for an opioid use disorder. We exclude 4,109 individuals with an adverse outcome prior to their initial opioid prescription or injection, since we assume they were already receiving opioids from another source, such as through private insurance before enrolling in Medicaid. Our final panel includes 80,768 individuals. Table S1 shows the incidence of adverse outcomes among these individuals by baseline characteristics.

3. Variable Construction

We construct variables that summarize information known in the 12 months prior to the individual's initial prescription.

Using the demographics from the integrated RI 360 database (8), we construct variables for (modal) age, sex, race, marital status, body mass index, and median income and fraction below the federal povery line in the home Census block group. Using DHS data, we construct variables for household size and new births in the household, and monthly payments for the Supplemental Nutrition Assistance Program (SNAP), the Temporary Assistance for Needy Families (TANF), the General Public Assistance (GPA), the Child Care Assistance Program (CCAP), and State Supplemental Payment portions of Supplemental Security Income benefits. Using DLT data, we construct indicators for sector of work derived from the first two digits of industry codes assigned according to the North American Industry Classification System (NAICS); monthly payments for Temporary Disability Insurance (TDI) and Unemployment Insurance (UI); and quarterly wage history, including average quarterly wages and variance, the number of employers and the number of hours worked (for hourly employees); the monthly unemployment rate in Rhode Island; and the annual national unemployment rate for two-digit NAICS industries that the individual has worked in. Using DOC data, we construct indicators for charges, seven categories of sentencing, and commitments and releases from prison. Using police data, we construct variables for arrests; the number of car crashes involved and injured in; and the number of and total fines for citations.

The largest set of variables comes from the Medicaid data. These include indicators for enrollment eligibility categories, plan type, and payer codes; number of claims and total bill and payment amounts for all claims and for Emergency Department claims; indicators for prescriptions in 262 drug categories from the AHFS Pharmacologic/Therapeutic Classification;* and topic models summarizing the concatenated text descriptions for all of the individual's ICD-9 diagnosis codes and HCPCS procedure codes. We also include summary counts of the number of distinct diseases using the Clinical Classifications Software (9), of distinct chronic conditions using the Chronic Condition Indicators (10), and of distinct procedure codes.

A. Topic Modeling. We construct the topic models using a technique called non-negative matrix factorization (NMF), which is commonly used in text analysis to discover latent topic structure in documents (11). In this application, we treat each individual's concatenated text descriptions of diagnosis and procedure codes as a document to learn the latent topic structure across individuals' health histories. Our topic models summarize 80,768 documents comprised of 16,367 distinct words from the code descriptions, after removing 173 uninformative words using a stopword list. The total corpus consists of over 20.5 million words.

NMF works by factorizing the non-negative $d \times w$ matrix of the documents' word frequencies into non-negative matrices $d \times t$ and $t \times w$, where d is the number of documents, w is the number of distinct words, and t is the number of topics. We apply a term frequency-inverse document frequency (TF-IDF) transformation to the $d \times w$ matrix to reweight the word frequencies by their overall frequencies in the entire corpus, which is common practice when implementing NMF. The $d \times t$ matrix represents the weighting of topics for each document, and the $t \times w$ matrix represents the weighting of words for each topic. We summarize each topic using the 10 words with the greatest frequency in the $t \times w$ matrix.

Because the number of topics t is not known *a priori*, we tune this parameter by finding the t with the best out-of-sample area under the operating-receiver characteristic curve (AUC) in a logistic regression that includes only the topic model variables. We use only the training set for this tuning, and further subdivide it in half into topic training and topic validation sets. We consider an increasing number of topics and terminate the tuning procedure when the AUC does not improve by more than 0.001. The tuning achieves AUCs on the topic validation set of 0.663 for 10 topics, 0.670 for 20 topics, 0.684 for 50 topics, 0.685 for 100 topics. Therefore, we select the model with 50 topics for the final variables.

B. Low-Dosage Opioids. Within the prescription drug categories, there is a category for opiate agonists. By construction of our panel, no individuals should have previously received an opioid prescription. However, the opiate agonist category includes 152 drugs that were not identified in the 4,175 opioid drugs from our historical NDC directory, and which are listed in Table S3. These drugs either contain an opioid ingredient at a lower amount than the minimum thresholds defined by the Washington State prescribing guidelines, or contain an ingredient not identified in those guidelines (e.g., "opium"). Therefore, the opiate agonist variable indicates that the individual received a drug that was not likely for initiating opioid therapy, but nonetheless contains a small amount of an opioid ingredient. Most of these drugs are over-the-counter cough syrups or painkillers combined with small amounts of an opioid ingredient. Of the 152, there are eight that are not present in the historical NDC directory, possibly because they were on the market for a short enough time that they do not occur in any of the available historical snapshots of the NDC directory.

C. Tensors. For our neural network models, we construct tensors of monthly values for a given variable for each of the individuals in our panel in the 12 months prior to the individual's initial prescription. Missing values are imputed using mean values from the training population.

The DHS tensor includes 13 variables for demographics (age and indicators for sex, race, and Spanish or Portuguese as a primary language) and monthly payments for the Supplemental Nutrition Assistance Program (SNAP), the Temporary Assistance for Needy Families (TANF), the General Public Assistance (GPA), the Child Care Assistance Program (CCAP), and State Supplemental Payment portions of Supplemental Security Income benefits.

The DLT tensor includes 31 variables for indicators for sector of work derived from the first two digits of industry codes assigned according to the North American Industry Classification System (NAICS); monthly payments for Temporary Disability Insurance (TDI) and Unemployment Insurance (UI); and quarterly wage history, including wage amount, the number of employers and the number of hours worked (for hourly employees).

The DOC tensor includes 16 variables for demographics (age and indicators for sex, race, Spanish as a primary language), and indicators for charges, seven categories of sentencing, and commitments and releases from prison.

The Medicaid tensor includes 683 variables for demographics (age and indicators for sex, race, and Spanish or Portuguese as a primary language); indicators for eligibility categories, plan type, and payer codes at each month of enrollment; number of claims and total bill and payment amounts for all claims and for Emergency Department claims; the number of prescriptions in each of 265 categories from the AHFS Pharmacologic/Therapeutic Classification; and indicators for ICD-9 diagnosis codes and HCPCS procedure codes for all codes that are correlated >0.02 with any adverse outcome in the training population.

The police tensor includes 42 variables for demographics (age and indicators for sex and officer-observed race); indicators for all arrests, DUI arrests, and domestic-offense arrests; the number of car crashes involved and injured in; the number of and total fines for citations; and the spatio-temporal intensity of calls for service in the individual's home Census block group for 29 categories of calls.

Finally, we construct an integrated tensor including all of the 785 variables from the DHS, DLT, DOC, Medicaid, and police tensors. The dimension of this integrated tensor are 70,153 individuals x 12 months x 785 variables.

4. Models

We estimate a range of predictive models using modern machine learning algorithms, which vary in both their complexity and interpretability. For example, a class of models called "regularized regression models" estimate standard linear models, but search over many potential explanatory variables, potentially more explanatory variables than available data observations, to maximize out-of-sample predictive fit and minimize overfitting. Like ordinary least squares or logistic models, the model results are easy to interpret, but the complexity is limited to functions of variables the researcher specifies in advance. At the other extreme are artificial neural network models where the algorithm searches over non-linear transformations of layers of local linear regressions. The increased complexity allows the algorithm to search for arbitrary non-linearities and interactions between variables, but at a cost of greatly reducing the interpretability of the model (e.g., it is difficult to simply measure which variables contribute most to predictive fit).

A. Regularized Regression. For our regularized regression, we use an algorithm called Bootstrap Least Absolute Shrinkage and Selection Operator (BOLASSO) (12). This algorithm is a generalization of the popular LASSO algorithm which is able to consistently identify a model even when predictors are highly correlated. The BOLASSO selects the predictors with non-zero coefficients that appear in at least 90% of bootstrapped LASSO models.

Following convention, we use BOLASSO to select the variables from among 560 variables which are persistently the strongest predictors of future adverse opioid outcomes, and we present results from a second-stage logistic regression of an indicator for future adverse outcomes on these selected variables, to describe the predictive power of each variabe. Exhibit A6 lists the variables selected by the BOLASSO as occuring with a non-zero coefficient in more than 90 of the 100 LASSO bootstrap replicates, along with the regression results from the second-stage logistic regression. In addition to the second-stage logistic regression, we also construct a regression ensemble model that averages the predictions of all 100 bootstrap replicates in the BOLASSO.

We fit each LASSO bootstrap replicate on the training set using a regularized logistic regression implementation called the gamma LASSO, which was developed specifically to address the challenges of modeling sparse, high-dimensional data (13). Since a predictive model fits idiosyncratic noise through increased complexity in the model's structure, machine learning techniques commonly penalize complexity in the models they produce through a process called regularization. We tune the regularization parameters for the gamma LASSO model through a parameter search over gamma values in [0, 1, 10] and a path of 100 lambda values, and we select the model with the best area under the receiver-operating characteristic curve (AUC) on the validation set. Regularization helps prevent overfitting to the training data and thus improves out-of-sample fit. We are primarily interested in out-of-sample performance since our goal is to use the model to inform successful policy interventions, which require making predictions on new observations (14).

B. Neural Networks. We train a neural network model for each tensor using the Python package Keras (15), which provides an interface to the TensorFlow library (16). Specifically, we train a recurrent neural network (RNN), since RNNs have the ability to model temporal patterns in the input data. We input our training data into a two-layer network of 12x12 Long Short-Term Memory (LSTM) (17) units with the tanh activation function. We input the last LSTM layer into a dense layer that

applies a sigmoid activation function to the weighted sum of the 10 inputs in order to produce a single predicted probability of adverse outcome. We employ regularization prior to each layer in the form of a dropout factor of 0.25, which causes a random deactivation of units within the layer during training with a fixed probability of 0.25 (18).

The neural networks are optimized to minimize the binary cross-entropy, also known as log-loss, on the training data. We use the Adam optimization algorithm (19), training with a batch size of 16. We tune the model on the validation set by allowing the neural network to train for as many epochs as needed until the area under the receiver-operating curve (AUC) from predictions on the validation set does not improve by 0.001. Table S13 shows the AUC from predictions on the testing set for each data source and each individual outcome.

5. Estimation of Adverse Outcome Cost

In 2015, 33,091 people died from drug overdoses involving opioids (20), and 2,375,000 individuals over the age of 12 had an opioid use disorder (21). The U.S. Department of Transportation's Value of a Statistical Life is $10.1 million. Florence *et al.* (22) estimate the aggregate annual societal cost of an opioid use disorder to be $61,297 (including additional cost of health care, substance abuse treatment, lost productivity, and criminal justice activities). Weiss and Rao (23) estimate a 50 percent recovery probability after one year of medication-assisted treatment. Using these statistics, with the simplifying assumption that once an individual receives a prescription, they either overdose resulting in death, become dependent but successfully recover after one year of treatment, or continue to be dependent for ten years, we estimate a ballpark present discounted value of $450,000 for C_A (see Table S4).

6. Propensity Score Models for Opioid Injection

To explore if rational addiction may drive first-time prescriptions for opioids, we examine data on adverse outcomes as a function of the patient's degree of knowledge that they are receiving an opioid. We use the fact that patients may receive opioids through epidural or intravenous injections during inpatient procedures. Under the assumption that these opioid recipients were less likely to be informed they were receiving an opioid than those receiving and filling a prescription from a physician, we would expect fewer adverse outcomes from opioids received through inpatient procedures than through prescriptions in a rational addiction framework.

In a simple comparison of means, we find the mean adverse outcome rate for those with an initial opioid through injection is 6.1 percent over the subsequent five years compared to 3.3 percent for an initial opioid through prescription (see Table S14). While this difference in means does not support a rational addiction model, it is likely that these two groups differ on many baseline characteristics and a comparison of means is biased.

Therefore, we estimate a propensity scoring model that makes use of the rich baseline data we have to understand and address the potential systematic differences between these two groups. First, we specify a logistic regression $I = \beta X + \epsilon$ for opioid-injection status I, where X is the set of variables selected by the main BOLASSO model (which excludes I) and ϵ is an error term. The propensity score is the predicted probabilities $\hat{p}$ from this model.

Next, we use the proposensity score to create a new sample through inverse probability of treatment weighting, with individual weights:

$$w_i = \begin{cases} 1, & I_i = 1 \\ \frac{\hat{p}}{1-\hat{p}_i}, & I_i = 0 \end{cases}.$$

To evaluate the balance of the weighted sample, we individually regressed each variable that was significant in the propensity scoring LASSO model with I in both an unweighted and weighted regression (see Table S15). None of these variables have significant coefficients in the weighted regressions, indicating that the sample is now balanced.

Finally, we estimate the weighted logistic regressions $Y = \gamma I + \xi$ (see Table S5) and $Y = \eta X + \theta I + \zeta$ (see Table S6), with weights w_i and error terms ξ and ζ. The significance of the coefficients γ and θ for the opioid-injection status indicator I test whether there is a systematic difference in outcomes between those who receive an opioid prescription versus an opioid injection, flexibly controlling for baseline characteristics. We find that the coefficients are not significant and have positive point estimates.

7. Simulated Correlation between Policy Effectivness and Risk

Consider the scenario where the policy effictiveness rate α_i for an individual i is negatively correlated with the individual's probability of a true positive TP_i, through the linear relationship $\alpha_i = 1 - \rho\hat{Y}_i$, where $\alpha_i \in [0, 1]$ and $\hat{Y}_i$ is the predicted risk for individual i. The parameter ρ, which measures the degree of the negative correlation, lies in the range:

$$\left[0, \frac{1}{\max(\hat{Y}_i)}\right].$$

Figure S3 shows the break-even cost ratio across cumulative risk deciles for selected values of ρ, using the averaged predicted risk $E[\hat{Y}_i]$ across the individuals in the decile. In the case of no correlation ($\rho = 0$), the break-even cost ratio is the same as in the case $\alpha = 1$ from Figure 2. It is 0.232 for the top risk decile, which corresponds to a diversion cost of $104,400, assuming an adverse-outcome cost of $450,000.

Intermediate values of ρ have similar impacts on the break-even cost ratio as lowering the homogeneous policy effectiveness rate α. For example, $\rho = 0.5$ has a break-even cost ratio of 0.209 for the top risk decile, corresponding to a diversion cost of \$94,050. This is similar to the break-even cost ratio of 0.221 for the homogeneous $\alpha = 0.893$.

At the extreme value $\rho = 4.02$, which occurs for $E[\hat{Y}_i] = 0.249$ in the top risk decile, $\alpha = 0$ and diversion is completely ineffective among the highest-risk individuals. However, it is effective for lower-risk individuals. For the second risk decile, the ratio is 0.047, which corresponds to a diverstion cost of \$21,150. Under this assumption of strong negative correlation between α_i and TP_i, the break-even cost ratio of the policy increases by including lower-risk individuals.

8. Population Estimates

To estimate population-level characteristics of Medicaid enrollees in Rhode Island, we constructed a second panel of longterm Medicaid enrollees. We included all enrollees who were enrolled for at least six out of 12 months in each of the five years between 2007 and 2011. This panel comprises 120,584 enrollees, who were enrolled with a median of 60 months (interquartile range of 59 to 60 months). Using this panel, we estimated the fraction of adverse outcomes, race/ethnicity, and median age among all enrollees and only those who received an opioid prescription, an opioid injection, both, or neither (see Table S14). For those with an opioid prescription, we calculated the average number of visits in the 30 days prior to the prescription, and the average distance to the five closest providers based on the Census block group of the last known home address before the prescription (see Table S8).

References

1. U.S. Food & Drug Administration, National Drug Code Directory (2019) Available at https://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm (Accessed 02 August 2019).
2. Research Improving People's Lives, Assembling a Historical National Drug Code Directory from the Internet Archive (2019) Available at https://github.com/ripl-org/historical-ndc (Accessed 02 August 2019).
3. Washington State Agency Medical Directors' Group, Interagency Guideline on Prescribing Opioids for Pain, Technical report (2015).
4. D Dowell, TM Haegerich, R Chou, CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. *MMWR. Recomm. Reports* **65** (2016).
5. CM Jones, Heroin use and heroin use risk behaviors among nonmedical users of prescription opioid pain relievers – United States, 2002–2004 and 2008–2010. *Drug Alcohol Dependence* **132**, 95–100 (2013).
6. DS Carrell, et al., Using natural language processing to identify problem usage of prescription opioids. *Int. J. Med. Informatics* **84**, 1057–1064 (2015).
7. JA Barocas, et al., Estimated Prevalence of Opioid Use Disorder in Massachusetts, 2011–2015: A Capture–Recapture Analysis. *Am. J. Public Heal.* **108**, 1675–1681 (2018).
8. JS Hastings, M Howison, T Lawless, J Ucles, P White, Unlocking Data to Improve Public Policy. *Commun. ACM* **62**, 48–53 (2019).
9. Agency for Healthcare Research and Quality, Clinical Classifications Software (CCS) for ICD-9-CM (2017).
10. Agency for Healthcare Research and Quality, Chronic Condition Indicator (CCI) for ICD-9-CM (2016).
11. C Févotte, J Idier, Algorithms for Nonnegative Matrix Factorization with the β-Divergence. *Neural Comput.* **23**, 2421–2456 (2011).
12. FR Bach, Bolasso: Model Consistent Lasso Estimation Through the Bootstrap in *Proceedings of the 25th International Conference on Machine Learning*. (Helsinki, Finland), pp. 33–40 (2008).
13. M Taddy, One-Step Estimator Paths for Concave Regularization. *J. Comput. Graph. Stat.* **26**, 525–536 (2017).
14. J Kleinberg, J Ludwig, S Mullainathan, Z Obermeyer, Prediction Policy Problems. *Am. Econ. Rev.* **105**, 491–495 (2015).
15. F Chollet, Keras: The Python Deep Learning library (2015) Available at https://keras.io/ (Accessed 02 August 2019).
16. M Abadi, et al., TensorFlow: A System for Large-Scale Machine Learning in *12th USENIX Symposium on Operating Systems Design and Implementation*. (Savannah, GA, USA), pp. 265–283 (2016).
17. S Hochreiter, J Schmidhuber, Long Short-Term Memory. *Neural Comput.* **9**, 1735–1780 (1997).
18. N Srivastava, G Hinton, A Krizhevsky, I Sutskever, R Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
19. DP Kingma, J Ba, Adam: A Method for Stochastic Optimization (2014) arXiv:1412.6980 (22 December 2014).
20. National Institute on Drug Abuse, National Overdose Deaths from Select Prescription and Illicit Drugs (2017) Available at https://www.drugabuse.gov/sites/default/files/overdose_data_1999-2015.xls (Accessed 25 March 2019).
21. Center for Behavioral Health Statistics and Quality, 2016 National Survey on Drug Use and Health: Detailed Tables (2017) Available at https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf (Accessed 25 March 2019).
22. CS Florence, C Zhou, F Luo, L Xu, The economic burden of prescription opioid overdose, abuse, and dependence in the United States, 2013. *Med. Care* **54**, 901–906 (2016).
23. RD Weiss, V Rao, The Prescription Opioid Addiction Treatment Study: What have we learned. *Drug & Alcohol Dependence* **173**, S48–S54 (2017).
24. U.S. Department of Transportation, Guidance on Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analysis – 2016 Adjustment (2016) Available at https://www.transportation.gov/sites/dot.gov/files/docs/2016%20Revised%20Value%20of%20a%20Statistical%20Life%20Guidance.pdf (Accessed 05 August 2019).
25. Center for Behavioral Health Statistics and Quality, Results from the 2016 National Survey on Drug Use and Health: Detailed Tables (2017) Available at https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf (Accessed 05 August 2018).

Table S1. Descriptive statistics for the final panel.

Variable	*Value*	*N*	*% Outcome*
Age	<18	8880	2.32%
	18-45	45563	6.88%
	45-60	14503	7.78%
	61+	11723	1.32%
	NA	99	4.04%
Race/Ethnicity	White	47385	7.75%
	Black	8142	4.41%
	Hispanic	8545	2.06%
	Other	5449	2.20%
	NA	11247	2.69%
Sex	Female	54520	4.95%
	Male	26205	7.37%
	NA	43	2.33%
Marital status	Married	13226	4.67%
	Not married	44284	7.75%
	NA	23258	2.50%
Body mass index	Underweight (<18.5)	1270	9.37%
	Normal (18.5-25)	22653	7.44%
	Overweight (25-30)	17907	5.96%
	Obese (>30)	14291	5.17%
	NA	24647	4.13%
Blockgroup fraction of residents below FPL	At least 16.4 percent	27717	5.47%
	Otherwise	53051	5.87%
Average quarterly wages in previous year	<$2500	15578	7.07%
	$2500-$7500	14835	4.49%
	$7500-$15000	1547	4.33%
	>$15000	70	2.86%
	$0 or NA	48738	5.73%
Received SNAP in previous year	Yes	44632	7.42%
	No	36136	3.65%
Received SSI in previous year	Yes	1597	7.51%
	No	79171	5.70%
Received UI in previous year	Yes	6478	6.53%
	No	74290	5.66%
Received TDI in previous year	Yes	5710	6.73%
	No	75058	5.66%
Children in DHS household in previous year	0 or NA	4866	3.12%
	1	23188	7.85%
	2+	52714	5.04%

Table S2. Regression output for the post-BOLASSO regression (a logistic regression of variables selected by BOLASSO as occurring with a non-zero coefficient in more than 90% of LASSO bootstrap replicates).

Variable	*Odds Ratio*	*95% C.I.*	*p-value*	*Freq.**
Released from a corrections facility	2.188	(1.795 - 2.667)	0.000	100%
Arrested during lookback period	1.756	(1.460 - 2.111)	0.000	93%
Prior prescription for Benzodiazepines	1.513	(1.341 - 1.707)	0.000	100%
Prior prescription for Centrally Acting Skeletal Muscle Relaxants	1.393	(1.219 - 1.592)	0.000	100%
Prior prescription for Opiate Agonists	1.357	(1.192 - 1.544)	0.000	96%
Rhode Island monthly unemployment rate	1.212	(1.151 - 1.275)	0.000	100%
Topic 32 (single substance group ethanol screen cocaine cannabis abuse dependence drug)	1.128	(1.097 - 1.159)	0.000	100%
Topic 26 (canal displacement degeneration spine spinal myelopathy lumbosacral intervertebral disc lumbago)	1.108	(1.073 - 1.145)	0.000	100%
Topic 6 (structures contusion supporting abscess evaluation management visit key comp department)	1.096	(1.056 - 1.137)	0.000	100%
Topic 38 (defiant psychotherapy management prescription review pharmacologic medication disorder attention deficit)	1.085	(1.049 - 1.123)	0.000	100%
Number of AHRQ CCS diseases	1.085	(0.991 - 1.188)	0.079	0%
Number of opioid prescriptions in household	1.076	(1.039 - 1.113)	0.000	100%
'Number of distinct procedures' X 'Number of unique Medicaid IDs'	1.065	(1.024 - 1.107)	0.002	96%
Total Food Stamps payments	1.059	(1.008 - 1.111)	0.022	18%
Topic 11 (pharmacologic moderate behavior severe psychotic disorder episode recurrent major depressive)	1.057	(1.021 - 1.095)	0.002	100%
Topic 30 (depressive mental classified psychotherapy episodic mixed anxiety adjustment depressed disorder)	1.054	(1.016 - 1.093)	0.005	97%
Total Medicaid pharmacy payments	1.053	(1.016 - 1.090)	0.004	95%
Topic 49 (tendons thigh minimum sprains strains upper arm pain joint region)	1.042	(1.004 - 1.082)	0.032	92%
Total TANF payments	1.041	(0.988 - 1.096)	0.133	4%
'Total TDI payments' X 'Average quarterly wages'	1.039	(1.019 - 1.059)	0.000	99%
Total Unemployment Insurance payments	1.037	(0.994 - 1.082)	0.089	16%
Topic 41 (supportive prescription review pharmacologic medication manic episode disorder current recent)	1.030	(0.998 - 1.063)	0.067	95%
'Total Food Stamps payments' X 'Number of unique Medicaid IDs'	1.027	(1.001 - 1.054)	0.042	94%
Count of car crashes involved in during lookback period	1.006	(0.970 - 1.043)	0.754	47%
Number of distinct procedures	0.983	(0.902 - 1.071)	0.696	0%
Number of unique Medicaid IDs	0.977	(0.935 - 1.021)	0.309	30%
Total TDI payments	0.976	(0.929 - 1.026)	0.344	14%
'Total TDI payments' X 'count of car crashes involved in during lookback period'	0.959	(0.904 - 1.018)	0.166	98%
'Total Unemployment Insurance payments' X 'Share of block group below poverty line'	0.955	(0.912 - 1.001)	0.057	92%
Age is 40-54	0.953	(0.852 - 1.065)	0.392	10%
Share of block group below poverty line	0.950	(0.906 - 0.995)	0.031	95%
Topic 47 (asthma tests physicians allergenic extracts provision professional immunotherapy allergic rhinitis)	0.928	(0.869 - 0.991)	0.026	92%
Topic 33 (abortion skin microscopic gross female pathology surgical biopsy breast malignant)	0.909	(0.857 - 0.963)	0.001	99%
Average quarterly wages	0.898	(0.846 - 0.953)	0.000	100%
Topic 16 (facility periodic general reevaluation history gynecological individual examination healthy routine)	0.897	(0.843 - 0.955)	0.001	99%
Size of household	0.886	(0.833 - 0.941)	0.000	99%
'Total TANF payments' X 'Age is 40-54'	0.873	(0.781 - 0.976)	0.017	91%
Topic 45 (collection hypercholesterolemia pure amino transferase serum lipid lipoprotein direct cholesterol)	0.870	(0.820 - 0.923)	0.000	100%
Topic 10 (conjunctivitis routine myopia gynecological deductible coinsurance office visits copay share)	0.866	(0.800 - 0.938)	0.000	98%
Topic 13 (monofocal purchases frames aphakia fitting plano spectacles sphere vision minus)	0.856	(0.796 - 0.921)	0.000	100%
'Number of AHRQ CCS diseases' X 'Age is 40-54'	0.840	(0.770 - 0.916)	0.000	96%
Topic 4 (evaluation periodic films bitewings oral application included topical fluoride adult)	0.808	(0.740 - 0.882)	0.000	100%
Topic 15 (adult procedures program monthly temporary mhrh offline mr dd intellectual)	0.709	(0.605 - 0.832)	0.000	100%
Primary language is Spanish	0.581	(0.459 - 0.734)	0.000	92%
Enrolled in Medicaid managed care	0.577	(0.508 - 0.655)	0.000	100%
Race is missing	0.576	(0.486 - 0.683)	0.000	96%
Married is missing	0.512	(0.446 - 0.588)	0.000	100%
Race is African American	0.502	(0.423 - 0.596)	0.000	96%
Age is 55-64	0.431	(0.352 - 0.528)	0.000	100%
Ethnicity is Hispanic	0.405	(0.319 - 0.514)	0.000	99%
Age is 65+	0.125	(0.091 - 0.172)	0.000	100%

* The frequency can be less than 90% for variables that were included in the post-BOLASSO as base terms of a selected interaction term.

Table S3. Low-dosage prescription opioids identified by the AHFS Pharmacologic/Therapeutic Classification category for opiate agonists.

NDC Code	*Opioid Ingredients*	*Other Ingredients*
00037-2403	codeine phosphate (16mg)	aspirin (325mg), carisoprodol (200mg)
00054-0243	codeine sulfate (15mg)	
00054-0386	hydromorphone hydrochloride (1mg)	
00093-0050	codeine phosphate (15mg)	acetaminophen (300mg)
00121-0504	codeine phosphate (12mg)	acetaminophen (120mg)
00121-0775	codeine phosphate (10mg)	guaifenesin (100mg)
00121-1775	codeine phosphate (10mg)	guaifenesin (100mg)
00185-0749	codeine phosphate (16mg)	aspirin (325mg), carisoprodol (200mg)
00378-6117	oxycodone hydrochloride (4.8355mg)	aspirin (325mg)
00378-7103	oxycodone hydrochloride (2.5mg)	acetaminophen (325mg)
00378-8088	tramadol hydrochloride (37.5mg)	acetaminophen (325mg)
00406-0483	codeine phosphate (15mg)	acetaminophen (300mg)
00482-0440	codeine phosphate (10mg)	guaifenesin (300mg)
00482-0441*		
00574-7040	opium (30-60mg), opium, powdered (60mg)	atropa belladonna (16.2mg), belladonna extract (16.2mg)
00574-7045	opium (30-60mg), opium, powdered (1.5%ww)	atropa belladonna (0.81%ww), atropa belladonna (16.2mg), belladonna (0.8-0.81%ww)
00574-7110	morphine sulfate (5mg)	
00591-0617*		
00591-0820	oxycodone hydrochloride (4.5mg)	aspirin (325mg), oxycodone terephthalate (0.38-0.4mg)
00591-3551	oxycodone hydrochloride (4.8355mg)	aspirin (325mg)
00603-1020	codeine phosphate (12mg)	acetaminophen (120mg)
00603-1075	codeine phosphate (10mg)	alcohol, dehydrated (3.68-3.7%vv), guaifenesin (100mg)
00603-1078	codeine phosphate (10mg)	alcohol, dehydrated (1.9%vv), guaifenesin (100mg), pseudoephedrine hydrochloride (30mg)
00603-1329	codeine phosphate (10mg)	guaifenesin (100mg)
00603-1520	codeine phosphate (10mg)	chlorpheniramine maleate (2mg), pseudoephedrine hydrochloride (30mg)
00603-1585	codeine phosphate (10mg)	promethazine hydrochloride (6.25-6.3mg)
00603-1588	codeine phosphate (10mg)	phenylephrine hydrochloride (5mg), promethazine hydrochloride (6.25mg)
00603-2337	codeine phosphate (15mg)	acetaminophen (300mg)
00603-4978	oxycodone hydrochloride (2.5mg)	acetaminophen (325mg)
00603-9013	codeine phosphate (12mg)	acetaminophen (120mg)
00641-1130*		
10135-0519*		
13107-0058	codeine phosphate (15mg)	acetaminophen (300mg)
16571-0301	codeine phosphate (10mg)	guaifenesin (100mg), pseudoephedrine hydrochloride (30mg)
16571-0302	codeine phosphate (10mg)	guaifenesin (100mg)
46672-0561	codeine phosphate (12mg)	acetaminophen (120mg)
49884-0946	tramadol hydrochloride (37.5mg)	acetaminophen (325mg)
50383-0079	codeine phosphate (12mg)	acetaminophen (120mg)
50383-0087	codeine phosphate (10mg)	guaifenesin (100mg)
50383-0804	codeine phosphate (10mg)	promethazine hydrochloride (6.25mg)
50383-0805	codeine phosphate (10mg)	phenylephrine hydrochloride (5mg), promethazine hydrochloride (6.25mg)
53489-0159	codeine phosphate (15mg)	acetaminophen (300mg)
53746-0617	tramadol hydrochloride (37.5mg)	acetaminophen (325mg)
57664-0185	codeine phosphate (10mg)	promethazine hydrochloride (6.25mg)
57664-0537	tramadol hydrochloride (37.5mg)	acetaminophen (325mg)
57963-0103	codeine phosphate (10mg)	guaifenesin (100mg)
58177-0449*		
58177-0620*		
58177-0621*		
58657-0500	codeine phosphate (10mg)	guaifenesin (100mg)
60432-0245	codeine phosphate (12mg)	acetaminophen (120mg)
60432-0606	codeine phosphate (10mg)	promethazine hydrochloride (6.25mg)
60505-2644	tramadol hydrochloride (37.5mg)	acetaminophen (325mg)
60505-7010	fentanyl (12ug)	
60951-0310	oxycodone hydrochloride (4.8355mg)	aspirin (325mg)
60951-0701	oxycodone hydrochloride (2.5mg)	acetaminophen (325mg)
63481-0121	oxycodone hydrochloride (4.8355mg)	aspirin (325mg)
63481-0627	oxycodone hydrochloride (2.5mg)	acetaminophen (325mg)
65162-0617	tramadol hydrochloride (37.5mg)	acetaminophen (325mg)
65162-0694	codeine phosphate (10mg)	phenylephrine hydrochloride (5mg), promethazine hydrochloride (6.25mg)
66594-0333	codeine phosphate (9mg)	pyrilamine maleate (8.33mg)
66689-0024*		
68308-0840	oxycodone hydrochloride (2.5mg)	acetaminophen (325mg)
68308-0845	oxycodone hydrochloride (4.8355mg)	aspirin (325mg)
68382-0334	tramadol hydrochloride (37.5mg)	acetaminophen (325mg)
69543-0252	codeine phosphate (10mg)	guaifenesin (100mg)
69543-0253	codeine phosphate (10mg)	guaifenesin (100mg), pseudoephedrine hydrochloride (30mg)
76439-0252	codeine phosphate (10mg)	guaifenesin (100mg)
76439-0253	codeine phosphate (10mg)	guaifenesin (100mg), pseudoephedrine hydrochloride (30mg)

* NDC code exists in AHFS Pharmacologic/Therapeutic Classification but does not exist in NDC directory.

Table S4. Estimation of the adverse outcome cost.

Cost of poisoning[1]	$140,735
Cost of successful treatment after 1 year[2]	$30,649
Cost of relapsed treatment for 10 years[3]	$269,282
Total adverse outcome cost:	*$440,666*

1. The cost of poisoning is estimated as the product of the probability of poisoning (0.014) and the DOT Value of a Statistical Life ($9,600,000 in 2015 dollars (24); $10,100,770 in inflation-adjusted 2018 dollars). The probability of poisoning is estimated as the number of deaths from drug overdoses related to opioids in 2015 (33,091 (20)) divided by the number of persons aged 12 or older estimated to misuse opioids in 2015 (2,375,000 (25)).
2. The cost of successful treatment is estimated as the product of the probability of successful remission following treatment for an opioid use disorder (0.5 (23)) and the estimated annual societial cost of a non-fatal opioid use disorder ($56,990 in 2013 dollars (22); $61,297 in inflation-adjusted 2018 dollars).
3. The cost of relapsed treatment is estimated as the product of the probability of relapse following treatment (0.5) and the present discounted value of 10 years of the annual societal cost of a non-fatal opioid use disorder ($538,564).

Table S5. Propensity-score weighted regression of opioid injection status on outcome outcomes.

Variable	*Odds Ratio*	*95% C.I.*	*p-value*
Initial exposure was an opioid injection	**1.118**	**(0.847 - 1.475)**	**0.431**
(Intercept)	0.087	(0.071 - 0.107)	0.000***

Table S6. Propensity-score weighted regression of post-BOLASSO variables on outcome outcomes.

Variable	*Odds Ratio*	*95% C.I.*	*p-value*
Arrested during lookback period	2.078	(1.196 - 3.612)	0.009**
Released from a corrections facility	1.986	(0.996 - 3.962)	0.051
Prior prescription for Benzodiazepines	1.559	(1.015 - 2.394)	0.043*
Prior prescription for Centrally Acting Skeletal Muscle Relaxants	1.391	(0.807 - 2.399)	0.235
Prior prescription for Opiate Agonists	1.374	(0.828 - 2.278)	0.218
Rhode Island monthly unemployment rate	1.319	(1.109 - 1.570)	0.002**
Age is 40-54	1.189	(0.805 - 1.757)	0.385
Topic 11 (pharmacologic moderate behavior severe psychotic disorder episode recurrent major depressive)	1.187	(1.060 - 1.329)	0.003**
Topic 32 (single substance group ethanol screen cocaine cannabis abuse dependence drug)	1.163	(1.066 - 1.269)	0.001***
Topic 30 (depressive mental classified psychotherapy episodic mixed anxiety adjustment depressed disorder)	1.150	(1.008 - 1.311)	0.037*
Initial exposure was an opioid injection	**1.138**	**(0.848 - 1.528)**	**0.389**
Number of opioid prescriptions in household	1.130	(0.838 - 1.524)	0.423
Topic 6 (structures contusion supporting abscess evaluation management visit key comp department)	1.125	(0.981 - 1.289)	0.091
Total Unemployment Insurance payments	1.104	(1.011 - 1.206)	0.027*
'Number of distinct procedures' X 'Number of unique Medicaid IDs'	1.101	(0.971 - 1.249)	0.132
Number of distinct procedures	1.081	(0.815 - 1.433)	0.588
'Number of AHRQ CCS diseases' X 'Age is 40-54'	1.063	(0.827 - 1.366)	0.633
Topic 26 (canal displacement degeneration spine spinal myelopathy lumbosacral intervertebral disc lumbago)	1.055	(0.911 - 1.221)	0.477
'Total Unemployment Insurance payments' X 'Share of block group below poverty line'	1.052	(0.952 - 1.163)	0.319
Total Medicaid pharmacy payments	1.052	(0.938 - 1.180)	0.388
'Total TDI payments' X 'Average quarterly wages'	1.039	(0.991 - 1.089)	0.110
Count of car crashes involved in during lookback period	1.014	(0.903 - 1.139)	0.813
Topic 41 (supportive prescription review pharmacologic medication manic episode disorder current recent)	1.007	(0.903 - 1.124)	0.896
Share of block group below poverty line	0.993	(0.847 - 1.165)	0.932
Topic 49 (tendons thigh minimum sprains strains upper arm pain joint region)	0.983	(0.825 - 1.170)	0.844
Topic 38 (defiant psychotherapy management prescription review pharmacologic medication disorder attention deficit)	0.983	(0.797 - 1.212)	0.869
'Total TDI payments' X 'count of car crashes involved in during lookback period'	0.979	(0.905 - 1.059)	0.603
'Total Food Stamps payments' X 'Number of unique Medicaid IDs'	0.970	(0.902 - 1.044)	0.418
Total Food Stamps payments	0.969	(0.822 - 1.142)	0.709
Topic 47 (asthma tests physicians allergenic extracts provision professional immunotherapy allergic rhinitis)	0.961	(0.707 - 1.307)	0.801
Total TANF payments	0.949	(0.701 - 1.284)	0.733
Number of unique Medicaid IDs	0.945	(0.832 - 1.073)	0.383
Total TDI payments	0.896	(0.762 - 1.054)	0.184
Topic 16 (facility periodic general reevaluation history gynecological individual examination healthy routine)	0.894	(0.670 - 1.192)	0.445
Size of household	0.882	(0.713 - 1.093)	0.251
Average quarterly wages	0.876	(0.727 - 1.055)	0.163
Topic 4 (evaluation periodic films bitewings oral application included topical fluoride adult)	0.871	(0.550 - 1.381)	0.557
Topic 45 (collection hypercholesterolemia pure amino transferase serum lipid lipoprotein direct cholesterol)	0.870	(0.723 - 1.047)	0.142
Number of AHRQ CCS diseases	0.862	(0.618 - 1.203)	0.383
Topic 33 (abortion skin microscopic gross female pathology surgical biopsy breast malignant)	0.834	(0.720 - 0.967)	0.016*
Topic 10 (conjunctivitis routine myopia gynecological deductible coinsurance office visits copay share)	0.824	(0.508 - 1.336)	0.432
'Total TANF payments' X 'Age is 40-54'	0.807	(0.451 - 1.446)	0.471
Topic 13 (monofocal purchases frames aphakia fitting plano spectacles sphere vision minus)	0.757	(0.486 - 1.179)	0.218
Enrolled in Medicaid managed care	0.698	(0.430 - 1.133)	0.146
Age is 55-64	0.649	(0.375 - 1.122)	0.122
Race is missing	0.647	(0.384 - 1.089)	0.101
Race is African American	0.628	(0.343 - 1.148)	0.131
Ethnicity is Hispanic	0.593	(0.284 - 1.240)	0.165
Primary language is Spanish	0.575	(0.253 - 1.308)	0.187
Married is missing	0.542	(0.351 - 0.838)	0.006**
Topic 15 (adult procedures program monthly temporary mhrh offline mr dd intellectual)	0.269	(0.040 - 1.804)	0.176
Age is 65+	0.255	(0.048 - 1.363)	0.110
(Intercept)	0.056	(0.034 - 0.090)	0.000***

Table S7. Calculations of the difference in False Discovery Rate (FDR) between whites and minorities that can be detected given our sample size at a power of 0.8.

Decile	*N White*	*N Minority*	*FDR White*	*FDR Minority**	*FDR Difference**
1	1,736	218	0.762	0.844	0.082
2	3,319	521	0.815	0.864	0.049
3	4,863	843	0.850	0.886	0.036
4	6,397	1,136	0.875	0.904	0.029
5	7,829	1,483	0.890	0.914	0.024
6	9,098	1,910	0.902	0.922	0.020
7	10,074	2,540	0.909	0.926	0.017
8	10,851	3,336	0.915	0.930	0.015
9	11,391	4,344	0.918	0.931	0.013
10	11,790	5,571	0.921	0.933	0.012

* The FDR for minorities and the difference in FDR was calculated from the other parameters using the `power twoprop` command in Stata version 14.2 (StataCorp LLC, College Station, TX).

Table S8. Population estimates of access to health care providers by minority status and adverse outcome status.

Group	*Average visits in 30 days prior to initial opioid prescription*	*Average distance to closest five providers at the last known home address prior to initial opioid prescription*
White	1.10	1.25km
African-American	1.12	0.89km
Hispanic	1.09	0.85km
Adverse outcome	1.03	1.14km
No adverse outcome	1.04	1.15km

Table S9. Cost of medication-assisted treatment (MAT) for 1,000 individuals assuming a 50% remission rate and annual MAT cost of $6,552 (low) to $14,112 (high).

Year	*Cost (Low)*	*Cost (High)*	*In Remission*	*% In Remission*
1	$3,276,000	$7,056,000	500	50%
2	$1,638,000	$3,528,000	750	75%
3	$819,000	$1,764,000	875	88%
Total	*$5,733,000*	*$12,348,000*		

Table S10. Minimum amounts of ingredients in a drug to classify it as an opioid prescription or a recovery prescription.

Opioid Ingredient	*Minimum Amount (mg)*
Codeine	30.0
Fentanyl	0.0125
Hydrocodone	5.0
Hydromorphone	2.0
Meperidine*	0.0
Morphine	10.0
Oxycodone	5.0
Oxymorphone	5.0
Tapentadol	50.0
Tramadol	50.0
Recovery Ingredient°	*Minimum Amount (mg)*
Buprenorphine	0.0
Methadone	0.0
Naloxone	0.0
Naltrexone	0.0

* Meperidine has no recommended starting dose for treatment of chronic pain because of its risk for complications in older adults; therefore, we consider any amount as evidence that the drug is an opioid.
° We consider any amount of a recovery ingredient as evidence that the drug may have been used to treat a prior opioid use disorder.

Table S11. Procedure codes used to identify opioid injections.

Code	*Description*
J2270	Injection, morphine sulfate, up to 10 mg
J2271	Injection, morphine sulfate, 100mg
J2275	Injection, morphine sulfate (preservative-free sterile solution), per 10 mg
Q9974	Injection, morphine sulfate, preservative-free for epidural or intrathecal use, 10 mg
S0093	Injection, morphine sulfate, 500 mg (loading dose for infusion pump
J2274	Injection, morphine sulfate, preservative-free for epidural or intrathecal use, 10 mg
J2410	Injection, oxymorphone hcl, up to 1 mg
J1170	Injection, hydromorphone, up to 4 mg
S0092	Injection, hydromorphone hydrochloride, 250 mg (loading dose for infusion pump)
J0745	Injection, codeine phosphate, per 30 mg
J3010	Injection, fentanyl citrate, 0.1 mg
J1810	Injection, droperidol and fentanyl citrate, up to 2 ml ampule
J2175	Injection, meperidine hydrochloride, per 100 mg
J2180	Injection, meperidine and promethazine hcl, up to 50 mg

Table S12. Diagnosis and procedure codes used to indicate adverse outcomes when occurring in any claim after the initial opioid prescription.

Outcome	*Code*	*Description*
Opioid Dependence	304.0	Opioid type dependence
	304.7	Combinations of opioid type drug with any other drug dependence
	F11.2*	Opioid dependence
Opioid Abuse	305.0	Nondependent opioid abuse
	F11.1*	Opioid abuse
Prescription-Opioid Poisoning	965.00	Poisoning by opium (alkaloids), unspecified
	965.02	Poisoning by methadone
	965.09	Poisoning by other opiates and related narcotics
	970.1	Poisoning by opiate antagonists
	E850.1	Accidental poisoning by methadone
	E850.2	Accidental poisoning by other opiates and related narcotics
	E935.1	Methadone causing adverse effects in therapeutic use
	E935.2	Other opiates and related narcotics causing adverse effects in therapeutic use
	E940.1	Opiate antagonists causing adverse effects in therapeutic use
	T400*	Poisoning by, adverse effect of and underdosing of opium
	T402*	Poisoning by, adverse effect of and underdosing of other opioids
	T403*	Poisoning by, adverse effect of and underdosing of methadone
Heroin Poisoning	965.01	Poisoning by heroin
	E850.0	Accidental poisoning by heroin
	E935.0	Heroin causing adverse effects in therapeutic use
	T401*	Poisoning by and adverse effects of heroin
Treatment	J2310°	Naloxone HCl Injection, per 1 mg
	J2315°	Naltrexone injection, depot form, 1mg
	J0592°	Buprenorphine HCL injection, 0.1mg
	X0305°	Methadone detoxification – outpatient
	X0321°	Methadone maintenance, assessment and evaluation, counseling, treatment and review, and lab testing
	H0020°	Alcohol and or drug services; methadone administration and or service
	J1230°	Injection, methadone, up to 10mg
	83840°	methadone
	946°	Alcohol and drug rehabilitation and counseling
	9464°	drug rehabilitation
	9465°	drug detoxification
	9466°	drug rehabilitation and detoxification
	9467°	combined alcohol and drug rehabilitation
	9468°	combined alcohol and drug detoxification
	9469°	combined alcohol and drug rehabilitation and detoxification

* ICD-10 diagnosis code
° HCPCS procedure code

Table S13. Area under the receiver-operating characteristic curve (AUC) of neural network models using different subsets of administrative data and outcome definitions. Confidence intervals are calculated from 100 bootstrap replicates.

	Any	*Dependence*	*Abuse*	*Prescription-Opioid Poisoning*	*Heroin Poisoning*	*Treatment*
DHS	0.727 (0.710-0.742)	0.735 (0.715-0.753)	0.710 (0.683-0.736)	0.652 (0.593-0.695)	0.594 (0.520-0.662)	0.739 (0.720-0.757)
DLT	0.529 (0.515-0.541)	0.554 (0.533-0.575)	0.526 (0.493-0.554)	0.499 (0.460-0.533)	0.446 (0.354-0.510)	0.550 (0.529-0.568)
DOC	0.691 (0.678-0.708)	0.714 (0.696-0.732)	0.725 (0.691-0.758)	0.638 (0.588-0.679)	0.724 (0.567-0.830)	0.695 (0.673-0.717)
Medicaid	0.773 (0.757-0.787)	0.783 (0.762-0.799)	0.770 (0.742-0.800)	0.732 (0.698-0.770)	0.689 (0.574-0.800)	0.760 (0.739-0.777)
Police	0.647 (0.631-0.663)	0.644 (0.621-0.662)	0.639 (0.604-0.673)	0.611 (0.554-0.658)	0.500 (0.500-0.500)	0.667 (0.641-0.688)
Integrated	0.801 (0.785-0.812)	0.814 (0.799-0.828)	0.786 (0.765-0.815)	0.755 (0.718-0.796)	0.703 (0.535-0.847)	0.792 (0.767-0.809)

Table S14. Population estimates based on a five-year panel of longterm Medicaid enrollees.

	All	*Opioid Rx*	*Opioid Injection*	*Both*	*Neither*
N	120,584	29,623	11,916	4,842	83,887
Adverse Outcome	4,545 (3.8%)	983 (3.3%)	732 (6.1%)	218 (4.5%)	1,114 (1.3%)
White	59,756 (49.6%)	17,465 (59.0%)	7,814 (65.6%)	3,210 (66.3%)	37,687 (44.9%)
African-American	13,480 (11.2%)	3,419 (11.5%)	1,006 (8.4%)	498 (10.3%)	9,553 (11.4%)
Hispanic	22,305 (18.5%)	3,112 (10.5%)	1,001 (8.4%)	418 (8.6%)	18,610 (22.2%)
Median Age	24	34	50	42	12

Table S15. Predictors of injection status before and after propensity-score weighting.

Variable	Unweighted *Odds Ratio (95% C.I.)*	*p-value*	Weighted *Odds Ratio (95% C.I.)*	*p-value*
'Number of distinct procedures' X 'Number of unique Medicaid IDs'	1.253 (1.212 - 1.295)	0.000***	1.002 (0.954 - 1.052)	0.946
Topic 33 (abortion skin microscopic gross female pathology surgical biopsy breast malignant)	1.239 (1.198 - 1.280)	0.000***	0.997 (0.954 - 1.042)	0.903
Rhode Island monthly unemployment rate	1.195 (1.133 - 1.262)	0.000***	1.003 (0.929 - 1.084)	0.936
Number of distinct procedures	1.170 (1.115 - 1.228)	0.000***	0.988 (0.929 - 1.051)	0.707
Total Unemployment Insurance payments	1.100 (1.057 - 1.144)	0.000***	1.000 (0.947 - 1.056)	1.000
Topic 45 (collection hypercholesterolemia pure amino transferase serum lipid lipoprotein direct cholesterol)	1.095 (1.044 - 1.149)	0.000***	0.997 (0.931 - 1.068)	0.931
Average quarterly wages	1.025 (0.973 - 1.079)	0.354	0.983 (0.932 - 1.036)	0.519
Topic 15 (adult procedures program monthly temporary mhrh offline mr dd intellectual)	0.977 (0.919 - 1.039)	0.462	0.999 (0.916 - 1.090)	0.983
Total Food Stamps payments	0.947 (0.895 - 1.003)	0.061	1.001 (0.927 - 1.080)	0.989
Topic 10 (conjunctivitis routine myopia gynecological deductible coinsurance office visits copay share)	0.914 (0.846 - 0.988)	0.024*	0.999 (0.897 - 1.114)	0.992
Topic 13 (monofocal purchases frames aphakia fitting plano spectacles sphere vision minus)	0.795 (0.728 - 0.868)	0.000***	1.000 (0.884 - 1.131)	0.998
Race is African American	0.773 (0.630 - 0.949)	0.014*	1.007 (0.757 - 1.341)	0.960
Topic 4 (evaluation periodic films bitewings oral application included topical fluoride adult)	0.709 (0.636 - 0.791)	0.000***	1.000 (0.867 - 1.154)	0.998
Number of unique Medicaid IDs	0.692 (0.665 - 0.720)	0.000***	0.997 (0.951 - 1.046)	0.914
Number of opioid prescriptions in household	0.620 (0.515 - 0.746)	0.000***	0.999 (0.842 - 1.186)	0.992
Age is 65+	0.321 (0.244 - 0.422)	0.000***	1.002 (0.681 - 1.472)	0.993
Enrolled in Medicaid managed care	0.320 (0.284 - 0.361)	0.000***	1.002 (0.847 - 1.187)	0.978

Fig. S1. Cumulative frequency of adverse outcomes over time since initial opioid prescription. Adverse outcomes are indicated by the diagnosis and procedure codes in Medicaid claims following the initial prescription. An individual may experience multiple types of adverse outcomes, and "any" is the union of the five specific outcome types. Opioid dependence is the most prevalent of the types.

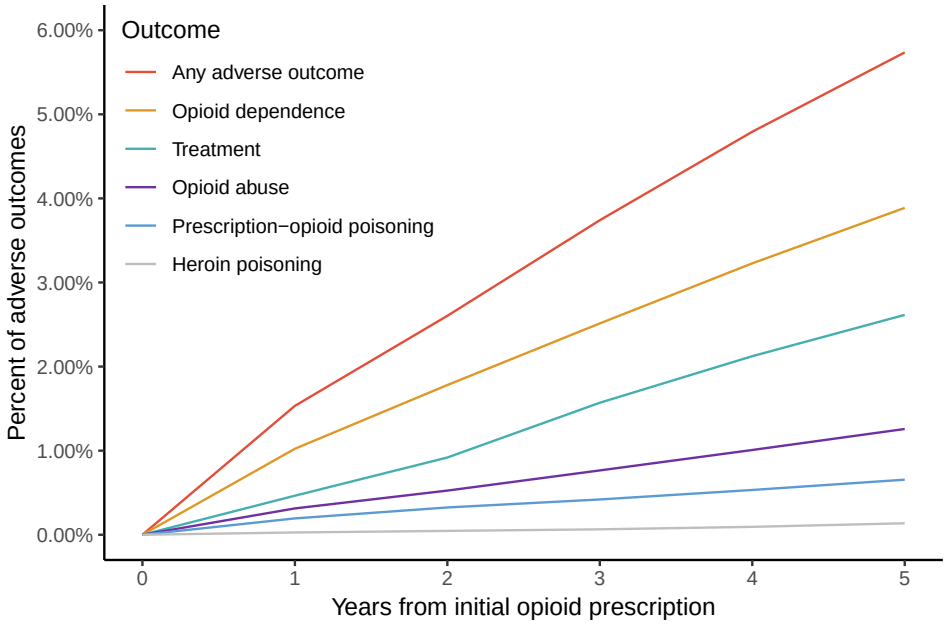

Fig. S2. The fraction of true outcomes in the test sample. The vertical black line indicates the base rate of outcomes among the entire population, which is 0.057.

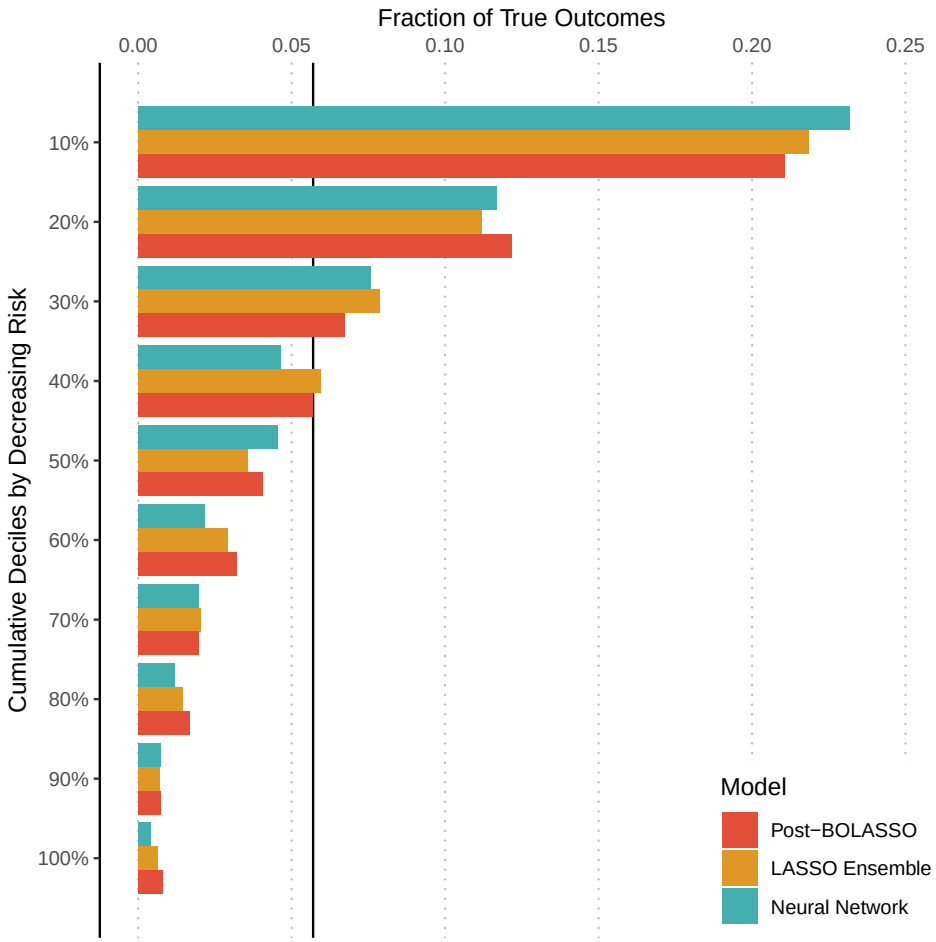

Fig. S3. The break-even cost ratio for four values of the efficacy parameter ρ.

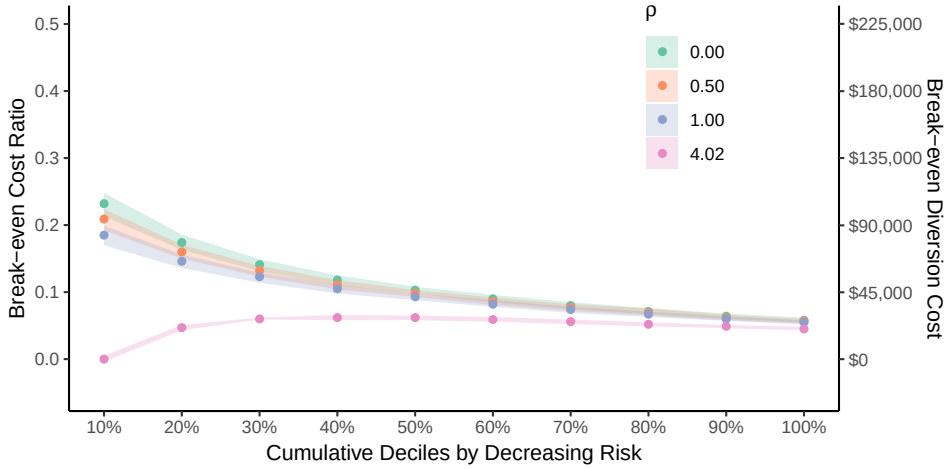

Fig. S4. The break-even cost ratio for minority status (a), incarceration history (b), and disability status (c). Error bars indicate the 95% confidence interval calculated from 100 bootstrap replicates.

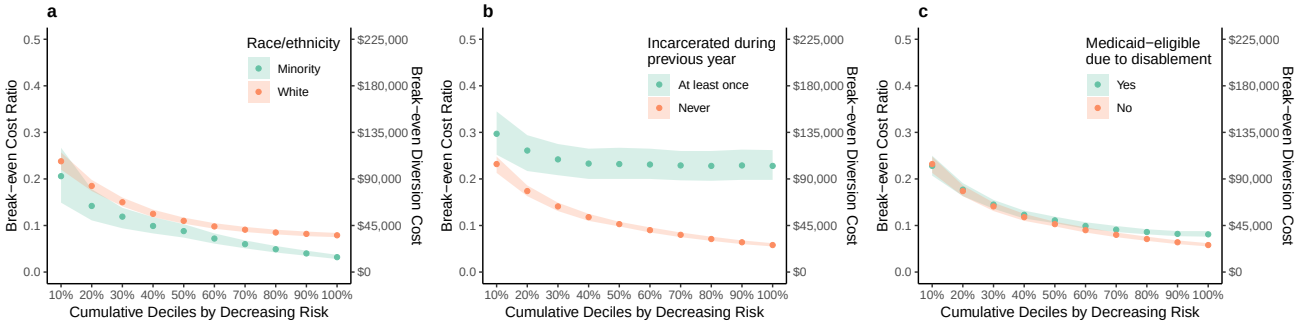

Fig. S5. Odds ratios from the post-BOLASSO regression using variables from Medicaid data only (c.f. Figure 1). Those < 0.75 and > 1.25 are labeled.

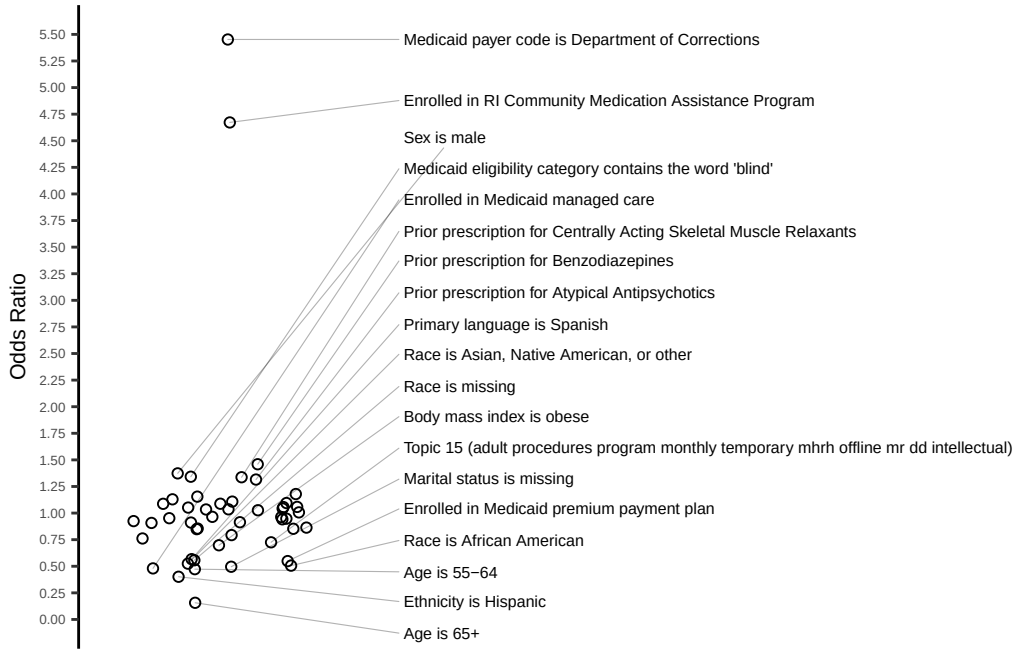

Fig. S6. The break-even cost ratio for three values of the effectiveness rate α for Medicaid data only (c.f. Figure 2).

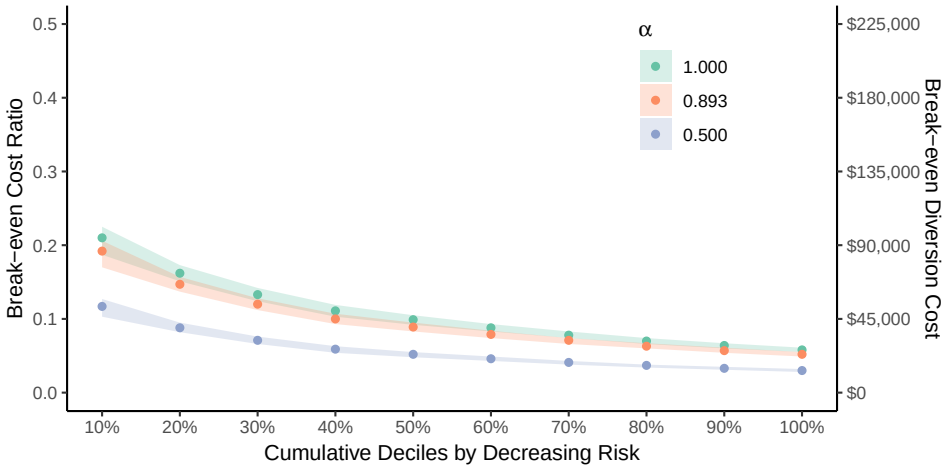

Fig. S7. The false discovery rate for minority status (a), incarceration history (b), and disability status (c) for Medicaid data only (c.f. Figure 3).