International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Critical Integrative Review of Machine Learning Approaches for Predicting Opioid Overdose Risk

Godwin Tetteh Wayoe[1], Godson Teye Apaflo[2]

[1]School of Computing, Applied Science, & Engineering – Carolina University, US
[2]Information Technology and Decision Sciences, University of Energy and Natural Resources

Abstract

Background: Machine learning has been increasingly applied to predict opioid overdose risk, yet the extent to which these models can support clinical decision-making and population-level overdose prevention remains unclear.

Objective: This critical integrative review synthesizes evidence on machine learning approaches for predicting opioid overdose risk. Studies were synthesized using a critical integrative approach emphasizing comparative evaluation of model types, predictors, validation practices, and implementation considerations.

Methods: Studies applying machine learning to predict opioid overdose, opioid use disorder, or related outcomes were identified. Twelve primary modeling studies were included and analyzed using a critical integrative review approach focused on comparative interpretation and identification of structural patterns across the literature.

Results: Tree-based ensemble methods frequently demonstrated moderate to high discrimination, with reported AUC values ranging from 0.72 to 0.79 in studies that provided this metric. Ecological and surveillance-enhanced models (particularly those incorporating unstructured emergency medical services data) showed value for geographic risk targeting. Key predictors included prescription patterns, mental health and substance use history, and prior overdose events. However, external validation was limited, fairness considerations were underdeveloped, and evidence of real-world implementation and impact remained scarce.

Conclusions: Although machine learning models can identify individuals and populations at elevated overdose risk, a substantial gap persists between technical development and effective translation into practice. Greater attention to external validation, fairness, implementation science, and multi-level integration is needed to realize the potential of these tools for overdose prevention.

Keywords: machine learning, opioid overdose, risk prediction, external validation, fairness

Introduction

The opioid overdose epidemic remains one of the most significant and persistent public health crises facing the United States. Over the past two decades, opioid-involved deaths have risen dramatically, initially driven by prescription opioids and more recently by the widespread presence of illicitly manufactured fentanyl and other synthetic opioids in the drug supply. Despite substantial national investments in prevention, treatment expansion, harm reduction, and policy reform, overdose mortality rates have remained stubbornly high in many regions (Ahmad et al., 2026; National Institute on Drug Abuse [NIDA], 2024). This persistent burden highlights the limitations of existing strategies and underscores the urgent need for more precise tools to identify individuals and communities at highest risk before fatal or nonfatal events occur.

In response to these challenges, researchers have turned to machine learning as a promising approach for improving risk prediction (Parton et al., 2026). Unlike traditional statistical methods, machine learning techniques can process large, high-dimensional datasets, detect complex non-linear interactions, and generate individualized or population-level risk estimates that may not be apparent through conventional modeling. Applications in this domain have included prediction of fatal and nonfatal overdose, development of opioid use disorder, and treatment retention or failure (Hayes et al., 2024; Balise et al., 2026). The resulting body of research draws on diverse data sources, including administrative claims, electronic health records, emergency medical services narratives, and publicly available ecological data.

However, this rapidly expanding literature is highly heterogeneous. Studies differ substantially in their outcome definitions, modeling architectures, validation strategies, intended applications, and level of analysis. Some focus on individual-level clinical risk stratification within healthcare systems, while others operate at county or neighborhood levels to support public health resource allocation. This diversity reflects the multifaceted nature of overdose risk but also creates difficulties in synthesizing findings and assessing the overall readiness of these models for real-world use. Without a critical integrative review, it is challenging to determine which approaches show the most promise, under what conditions they perform best, and what barriers currently limit their translation into effective clinical or public health practice.

A critical integrative review is therefore essential. Previous systematic reviews in this field have largely focused on model performance metrics, specific data sources such as electronic health records, or algorithmic approaches such as deep learning. These reviews have provided valuable summaries but have generally stopped short of critically examining cross-cutting methodological tensions, fairness and equity considerations, and the persistent gap between technical performance and real-world implementation. A critical integrative approach is needed to move beyond descriptive summaries and instead identify structural patterns, comparative strengths and weaknesses across model types, and the conditions necessary for meaningful translation into practice.

This review addresses these objectives by synthesizing evidence from studies that have applied machine learning to predict opioid overdose risk and related outcomes. It examines the data sources and computational architectures employed, assesses reported performance and validation practices, identifies consistently important risk factors, and explores the implications of these models for clinical care and population-level prevention. Particular attention is given to the gap between technical capability and practical utility, as well as to emerging concerns regarding fairness, equity, and governance.

By integrating findings across individual-level clinical prediction and ecological public health surveillance paradigms, this review aims to provide a balanced and analytically rigorous assessment of the current evidence base. In doing so, it seeks to inform both future research priorities and the responsible development and deployment of predictive tools in a domain where the human and societal stakes are exceptionally high.

Machine Learning Architectures Used for Opioid Overdose Risk Prediction

Tree-based ensemble methods have become the dominant architectural choice in machine learning approaches for individual-level opioid overdose risk prediction. Random forest and gradient boosting machine models have been widely applied because they can effectively manage high-dimensional clinical and prescription data while handling class imbalance through techniques such as Synthetic Minority Over-sampling Technique (SMOTE). These characteristics make them particularly suitable for predicting relatively rare but high-stakes events such as opioid overdose (Parton et al., 2026; Hayes et al., 2024; Balise et al., 2026).

Although SMOTE improved minority class representation in several studies, few evaluated whether synthetic oversampling introduced clinically implausible patterns or inflated performance estimates. Where performance metrics were reported, tree-based models demonstrated moderate to high discrimination. For example, Hayes et al. (2024) reported AUC values ranging from 0.72 to 0.79 when predicting retention, overdose, and mortality outcomes among veterans receiving buprenorphine.

More advanced architectures have been applied in contexts involving complex spatial or temporal dependencies. Chen et al. (2026) used a spatio-temporal graph neural network to predict county-level overdose mortality and reported an AUC of 0.738. This suggests that graph-based models may offer advantages when the prediction task requires modelling geographic and temporal patterns in ecological overdose data. In contrast, studies focused on individual-level prediction have more commonly relied on tree-based ensembles, partly due to their balance of performance and interpretability in high-dimensional clinical datasets (Hayes et al., 2024; Parton et al., 2026).

Hybrid architectures have also been explored in specific settings. Rock et al. (2026) combined natural language processing with machine learning classifiers to improve the detection of suspected opioid overdoses from emergency medical services free-text data, demonstrating the value of hybrid approaches when structured data alone is insufficient. Benitez-Aurioles et al. (2026) incorporated competing risks modelling to better account for mortality outcomes. These examples indicate that architectural choices in opioid overdose research are often influenced by the type of available data and the specific prediction objective.

As the field develops, attention is increasingly being given to the practical trade-offs between model complexity, interpretability, and implementation feasibility. The systematic review by Song et al. (2024) observed that while more complex models can capture intricate patterns, simpler and more interpretable models such as tree-based ensembles often remain attractive in operational settings. This suggests that the selection of machine learning architectures for opioid overdose risk prediction is shaped not only by predictive accuracy but also by considerations of transparency and real-world applicability.

Predictive Performance and Model Evaluation

Table 1. Comparative Performance, Validation Characteristics, and Implementation Status of Machine Learning Models for Opioid Overdose Risk Prediction

Study (Author, Year)	Primary ML Method(s)	Data Source	Main Outcome	Key Performance	Validation Type	Implementation Status
Parton et al. (2026)	Gradient Boosting Machine +	Alabama Medicaid claims	Opioid overdose	Recall improved; AUC not	Temporal	Model developed but not implemented

	SMOTE			reported		in clinical or public health workflows
Hayes et al. (2024)	Random Forest, GBM, DNN	VHA veterans EHR	Retention, overdose, mortality	AUC range: 0.72–0.79	Internal split	Model developed but not implemented in clinical workflows
Balise et al. (2026)	Tree-based ensembles	Claims/EHR	MOUD treatment failure	Discrimination reported as moderate to good; exact AUC not reported	Internal	Model developed but not implemented in clinical workflows
Li et al. (2025)	Ensemble + phenotype rules	Real-time EHR	OUD phenotype / overdose risk	AUC 0.99 (retrospective evaluation)	Temporal + prospective	Implemented in real-time within one healthcare system
Chen et al. (2026)	Spatio-temporal graph neural network	County-level geospatial	County overdose death	AUC 0.738	Temporal	Model developed but not implemented in public health practice
Kumar & Butler (2025)	XGBoost + SHAP	County-level US data	County overdose death	Discrimination described as good; exact AUC not reported	Internal	Model developed but not implemented in public health practice
Rock et al. (2026)	NLP + ML hybrid	EMS free-text	Suspected opioid overdose identification	Improved case detection compared with rule-based methods; exact performance metrics not reported	Internal	Enhanced surveillance case identification; not deployed as a clinical prediction tool
Schell et al. (2022)	LASSO + Random Forest	ACS + mortality (CBG level)	Neighborhood overdose death	$R^2 = 0.17$ †	Double cross-validation	Model developed but not implemented in public health practice
Benitez-Aurioles	Competing risks + ML	UK linked data	Opioid-related death	Calibration improved;	Internal	Model developed but

et al. (2026)				discrimination metrics not reported		not implemented in clinical workflows

Reported model performance has varied substantially across studies depending on outcome definition, data source, and validation approach. Area under the receiver operating characteristic curve values have commonly ranged from 0.70 to 0.79 in internal validation settings when reported, with gradient boosting and random forest models frequently performing at the higher end of this spectrum in structured claims and electronic health record data (Hayes et al., 2024; Parton et al., 2026). However, performance has often declined when models were tested using temporal or external validation, reflecting the challenge of temporal drift in an evolving epidemic (Song et al., 2024; Schell et al., 2022).

Calibration assessment has been less consistently reported than discrimination metrics but remains important for practical decision-making. Where evaluated, calibration has been mixed, with some models showing reasonable alignment between predicted probabilities and observed outcomes while others exhibited systematic miscalibration, particularly at higher risk thresholds. Techniques to address class imbalance, such as SMOTE, have been shown to improve recall in several investigations (Parton et al., 2026; Benitez-Aurioles et al., 2026). However, few studies assessed whether synthetic oversampling introduced clinically implausible patterns or inflated performance estimates.

External validation remains relatively uncommon. While some studies conducted temporal validation, true geographic or multi-system external validation is rare. This limitation restricts conclusions about model transportability across different healthcare systems, regions, and evolving drug environments (Song et al., 2024). Because opioid overdose is a rare event, reliance on AUC alone can be misleading, as it may not adequately reflect performance on the minority class. Greater attention to precision-recall trade-offs and other threshold-dependent metrics would provide a more complete picture of model utility in this context.

These patterns collectively indicate that while technical performance has advanced in internal validation settings, the gap between reported metrics and real-world utility remains a critical concern. Greater emphasis on prospective evaluation, external validation, and metrics appropriate for imbalanced outcomes is needed in future work.

Risk Factors and Predictive Features

A core set of predictors has shown remarkable consistency across studies despite heterogeneity in populations, data sources, and modeling approaches. Prescription-related variables, including higher morphine milligram equivalents, longer duration of opioid therapy, concurrent benzodiazepine or gabapentinoid use, and patterns suggestive of multiple prescriber or pharmacy use, have emerged as strong and recurrent predictors in claims-based studies (Parton et al., 2026; Hayes et al., 2024). These variables appear to capture both pharmacological exposure and behavioral indicators of elevated risk.

History of mental health conditions and substance use disorders has also demonstrated robust predictive value. Prior diagnoses of depression, anxiety disorders, post-traumatic stress disorder, and non-opioid substance use disorders have been consistently associated with increased overdose risk (Song et al., 2024). Prior overdose events represent one of the strongest single predictors identified, suggesting that individuals who have experienced an overdose remain at substantially elevated risk for subsequent events (Parton et al., 2026).

At the population level, ecological studies have identified complementary contextual predictors. Lower educational attainment, economic disadvantage, residential instability, and social isolation at the neighborhood or county level have been associated with higher rates of overdose mortality (Schell et al., 2022; Chen et al., 2026). These findings indicate that overdose risk is shaped by both individual-level clinical and behavioral factors and broader social and environmental conditions.

The convergence of predictors across individual and ecological studies supports the potential value of multi-level approaches to risk assessment and intervention, although few studies have yet integrated these perspectives within a single analytical framework. The systematic review by Song et al. (2024) and the neighborhood-level work by Schell et al. (2022) further emphasize that these predictors are stable across different data ecosystems. This convergence also highlights the need for continued research into how these predictors interact in real-time clinical and public health contexts.

Public Health Surveillance and Population-Level Prediction

Machine learning has demonstrated value in enhancing public health surveillance capabilities. Ecological models operating at county or neighborhood levels have shown the feasibility of using data such as mortality records, emergency medical services calls, census data, and social vulnerability indices to identify geographic areas at elevated risk of overdose mortality. These models offer a scalable approach to support population-level targeting of interventions such as naloxone distribution and treatment expansion (Chen et al., 2026; Schell et al., 2022; Kumar & Butler, 2025).

Natural language processing applied to unstructured emergency medical services data has also improved surveillance. Rock et al. (2026) demonstrated that machine learning-enhanced identification of suspected opioid overdoses significantly outperformed traditional rule-based methods that rely solely on structured data. While these surveillance applications show promise, there is still limited evidence regarding the operational integration of machine learning outputs into public health workflows and the downstream impact of geographically targeted interventions on overdose outcomes.

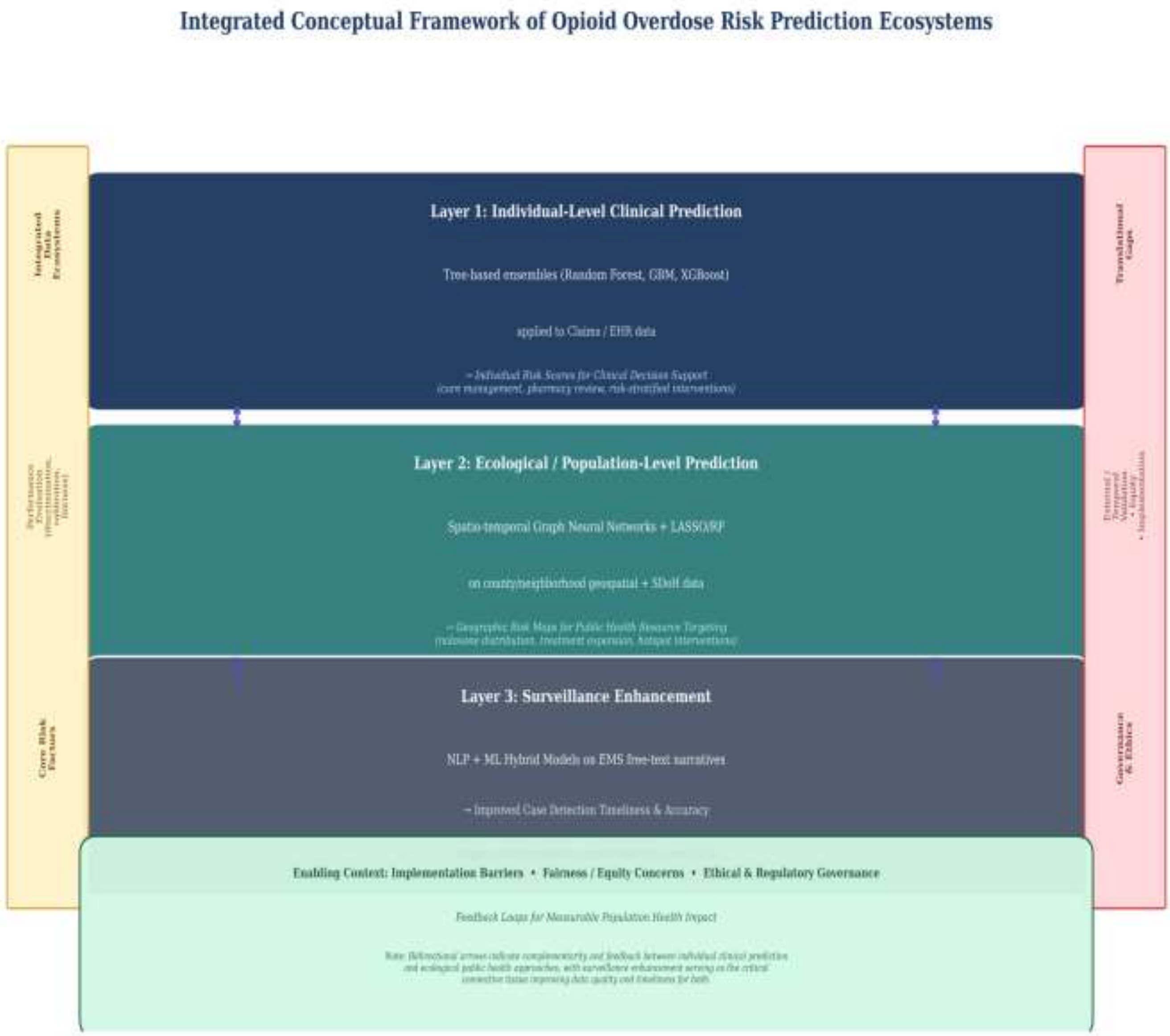
The work of Chen et al. (2026) and Schell et al. (2022) demonstrates clear potential for geographic risk mapping. However, the translation of these maps into actionable public health strategies remains underdeveloped. Similarly, while Rock et al. (2026) showed that unstructured data can enhance case detection, the downstream effects on response times and resource deployment are not yet well documented. This gap between improved surveillance data and effective intervention highlights the need for closer integration between predictive analytics and operational public health systems.

Overall, ecological and surveillance-enhanced models complement individual-level clinical prediction by providing a broader population perspective. Greater integration between these paradigms could strengthen both risk identification and intervention planning, although this remains an underdeveloped area (Song et al., 2024).

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Figure 1. Integrated Conceptual Framework of Opioid Overdose Risk Prediction Ecosystems



Note: The framework illustrates the complementarity between individual-level clinical prediction (Layer 1), ecological/population-level prediction (Layer 2), and surveillance enhancement (Layer 3). Bidirectional arrows represent feedback loops and data flow. Cross-cutting elements (data ecosystems, performance evaluation, risk factors, and translational gaps) connect all layers, while the outer enabling context emphasizes implementation barriers, fairness/equity, governance, and the ultimate goal of measurable population health impact.

Emerging Gaps and Future Research Directions

Despite progress in model development, several structural gaps limit the translation of predictive capability into improved outcomes. The most significant gap is the limited evidence of real-world implementation and impact. Most studies have focused on retrospective model performance rather than prospective evaluation of clinical or public health utility. This implementation gap represents the primary barrier to realizing the potential benefits of these models.

Fairness and equity considerations remain notably underdeveloped. Systematic examination of differential model performance across demographic or geographic groups is rare, creating risk that deployed models could perform unevenly or exacerbate existing disparities. External and temporal validation also remain insufficient. Most studies have relied on internal validation, with temporal validation conducted in only a minority of cases and multi-system external validation being exceptionally uncommon. This constrains confidence in model generalizability across different populations and evolving conditions.

Data limitations further constrain progress. Most models depend on structured claims or electronic health record data that incompletely capture social determinants, behavioral factors, and exposure to illicitly manufactured synthetic opioids. Governance, regulatory, and ethical frameworks for deployment have also received minimal explicit attention despite their growing relevance. Questions regarding accountability, transparency, privacy, and regulatory pathways for predictive systems in healthcare and public health settings remain largely unaddressed.

To address these gaps, future research should prioritize prospective validation studies and multi-site external validation using temporally and geographically distinct datasets. There is a particular need for studies that link clinical and claims data with social determinants of health data (such as housing, income, and neighborhood-level indices) to improve model fairness and contextual relevance. In addition, hybrid effectiveness-implementation trials are needed to evaluate not only predictive accuracy but also the feasibility, acceptability, and impact of integrating machine learning tools into real-world clinical and public health workflows.

Future studies should also adopt more rigorous evaluation standards, including the routine reporting of calibration, decision curve analysis, and fairness metrics (such as disparate impact and equalized odds). Implementation science frameworks, such as the Consolidated Framework for Implementation Research (CFIR) or the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework, could guide the design and evaluation of deployment studies. Finally, greater collaboration between clinical informatics researchers, public health practitioners, and implementation scientists is needed to ensure that future models are not only technically sound but also actionable and equitable in real-world settings.

Conclusion

Machine learning has emerged as a promising tool for identifying individuals and populations at elevated risk of opioid overdose, yet its current impact remains limited by a persistent gap between technical development and real-world application. Across diverse data sources and methodological approaches, models have demonstrated meaningful predictive capability, particularly when tree-based ensemble methods are applied to structured claims and electronic health record data. However, strong retrospective performance has not consistently translated into improved clinical decision-making or more effective public health interventions.

Several structural challenges constrain progress. External and temporal validation remain insufficient, limiting confidence in model generalizability. Fairness and equity considerations have received inadequate attention, raising concerns that predictive systems may perform unevenly across demographic and geographic groups. Most significantly, evidence of successful implementation and measurable impact on overdose outcomes is scarce. Without deliberate investment in prospective evaluation, workflow integration, and fairness assessment, even sophisticated models risk remaining underutilized or producing unintended consequences.

Moving forward, the field would benefit from greater integration between individual-level clinical prediction and ecological public health approaches, alongside more rigorous attention to implementation science and equity. Realizing the potential of machine learning in this domain will require not only continued technical refinement but also sustained focus on the conditions necessary for responsible and effective translation into practice. Addressing these challenges represents the most pressing priority for advancing overdose prevention through predictive analytics.

References

1. Ahmad FB, Cisewski JA, Rossen LM, Sutton P. Provisional drug overdose death counts. National Center for Health Statistics. 2026. DOI: https://dx.doi.org/10.15620/cdc/20250305008
2. Balise, R. R., Grealis, K., Brandt, L., Luo, S. X., Odom, G., Jean-Baptiste, G., & Feaster, D. J. (2026). Applying machine learning in predicting medication treatment outcomes for opioid use disorder. Journal of Substance Use & Addiction Treatment, 182, Article 209847. https://doi.org/10.1016/j.josat.2025.209847
3. Benitez-Aurioles, J., Medina, C. R. R., Jenkins, D., Peek, N., & Jani, M. (2026). Development and evaluation of machine learning algorithms for the prediction of opioid-related deaths among UK patients with non-cancer pain. PLOS Digital Health, 5(1), e0001190. https://doi.org/10.1371/journal.pdig.0001190
4. Chen, X., Gu, Z., Myers, J., Kim, J., Yin, C., Fareed, N., Thomas, N., Fernández, S., & Zhang, P. (2026). *Opioid overdose death prediction with graph neural networks* [Preprint]. medRxiv. https://doi.org/10.64898/2026.03.18.26348454
5. Hayes, C. J., Noor, N. B., Raciborski, R. A., Martin, B. C., Gordon, A. J., Hoggatt, K. J., Hudson, T., & Cucciare, M. A. (2024). Development and validation of machine-learning algorithms predicting retention, overdoses, and all-cause mortality among US military veterans treated with buprenorphine for opioid use disorder. Journal of Addictive Diseases, 43(3), 207–224.
6. Kumar, V., & Butler, R. (2025). Opioid overdose death prediction using machine learning and risk factor analysis using SHAP values for US counties. International Journal of Mental Health and Addiction. Advance online publication. https://doi.org/10.1007/s11469-025-01563-6
7. Li, H., Iscoe, M., Lutz, J., Hopper, C. D., Fried, S., Minaya, J., King, C. R., Reykhart, O., Paek, H., Meeker, D., Melnick, E. R., & Taylor, R. A. (2025). Implementation of an Opioid Use Disorder (OUD) machine-learning phenotype in real-time for the ADAPT Project. MedRxiv.
8. National Institute on Drug Abuse. (2024). Drug overdose death: facts and figures. https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates
9. Parton, J., Wang, Q., Wang, E. C., Hanson, C., & Higginbotham, J. A. (2026). A machine learning approach for opioid overdose risk prediction among Alabama Medicaid beneficiaries with opioid prescriptions. Scientific Reports, 16, Article 5171. https://doi.org/10.1038/s41598-026-36047-7
10. Rock, P., Slavova, S., Walsh, S. L., Martin, J., & Harris, D. R. (2026). Evaluation and enhancement of suspected opioid overdose definitions in emergency medical services data using machine learning with natural language processing. PLOS ONE, 21(4), e0347589. https://doi.org/10.1371/journal.pone.0347589

11. Schell, R. C., Allen, B., Goedel, W. C., Hallowell, B. D., Scagos, R., Li, Y., Krieger, M. S., Neill, D. B., Marshall, B. D. L., Cerdá, M., & Ahern, J. (2022). Identifying predictors of opioid overdose death at a neighborhood level with machine learning. American Journal of Epidemiology, 191(3), 526–533. https://doi.org/10.1093/aje/kwab279
12. Song, S. L., Dandapani, H. G., Estrada, R. S., Jones, N. W., Samuels, E. A., & Ranney, M. L. (2024). Predictive models to assess risk of persistent opioid use, opioid use disorder, and overdose. *Journal of Addiction Medicine*, *18*(3), 218–239. https://doi.org/10.1097/ADM.0000000000001276