Complications in Using Real-World Data to Study the Health of People Who Use Drugs

Mary C. Figgatt,[a,b] Asher J. Schranz,[c]
Juan M. Hincapie-Castillo,[a] Yvonne M. Golightly,[a,c,d]
Stephen W. Marshall,[a,b] Nabarun Dasgupta[b,e]

For the past two decades, the United States has suffered dramatic increases in a fatal drug overdose.[1] Overdoses represent one major health concern facing people who use drugs, in a milieu including injection drug use-related infections.[2] Accordingly, the use of real-world data to study drug-related harms, such as overdoses and injection drug-related infections, has become more common in the medical literature.[3–5] Real-world data sources include electronic health records and administrative claims that can be leveraged to understand health needs and longitudinal health utilization patterns in specific groups of patients, including people who use drugs. Researchers developed various algorithms that use combinations of drug-related diagnosis codes for substance use disorders and overdoses. However, limited attention to date has been paid to the methodologic and ethical implications arising from the limitations of these algorithms.

People who are perceived to use drugs in an illicit manner are frequently treated poorly within the health care system, perpetuating marginalization. More consideration should be given to the ethical implications of using healthcare data to identify and study marginalized groups of people. Additionally, the choice of some algorithms over others may deliver results that are misleading. Poor validity of these algorithms can bias study results, potentially leading to erroneous changes in policy or practice. Further examination of the methodology, applications, and validity of drug use-related algorithms is needed.

We sought to provide a brief review of the definitions of drug use-related algorithms, to appraise existing validation studies, and to discuss the limitations of these methods. The reviewed drug use-related classifications focus on people using substances such as opioids, stimulants, and other psychotropic drugs. Although some of these drugs may be prescribed, the large majority of drug overdoses are due to illicitly manufactured fentanyl.[1] We excluded algorithms focused on tobacco or alcohol.

TERMINOLOGY

Throughout this article, we use the terms "people who use drugs" and "drug use." We acknowledge the ambiguity implicit in these terms. They are used purposefully to avoid

From the [a]Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC; [b]Injury Prevention Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; [c]Division of Infectious Diseases, University of North Carolina School of Medicine, Chapel Hill, NC; [d]College of Allied Health Professions, University of Nebraska Medical Center, Omaha, NE; and [e]University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC.

This work was supported by the National Institute On Drug Abuse of the National Institutes of Health under Award Numbers F31DA055345 (Figgatt) and K23DA049946 (Schranz). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
The authors report no conflicts of interest.

Correspondence: Mary C. Figgatt, 135 Dauer Drive, Chapel Hill, NC 27599. E-mail: mfiggatt@unc.edu.

ISSN: 1044-3983/23/342-259-264
DOI: 10.1097/EDE.0000000000001564

historically stigmatizing and harmful labels while acknowledging the considerable heterogeneity in drug use behaviors and outcomes.

Among people who use drugs, individual drug use characteristics and outcomes vary widely. For many, no adverse health events will occur nor will they seek healthcare for drug use-related concerns. For some, adverse drug-related health events, such as an infection or overdose, may result in the healthcare system interaction (Figure). This subgroup is both the population with a greater public health need and the population that is observable within real-world data via algorithms.

We use the term "algorithms" to refer to the use of clinical codes, or other systematically collected health data, in healthcare records. Individuals meet an algorithm's inclusion criteria when they have one or a combination of drug use-related diagnosis (e.g., opioid use disorder and drug poisoning), procedure (e.g., medication initiation such as buprenorphine), and prescription drug codes noted on their healthcare records.

ASCERTAINMENT OF DRUG USE IN HEALTHCARE DATA

Drug use is often highly stigmatized. Therefore, it is often a concealed behavior, hindering systematic documentation in administrative healthcare data sources. Healthcare data result in a classic high-specificity and low-sensitivity situation whereby codes are most accurately recorded in extreme cases when seeking medical care is precipitated by drug use itself (e.g., inpatient treatment for overdose or treatment of injection-related infections).[6] Importantly, national surveys show that only a small fraction of people who use drugs will have a medical encounter that results in observable codes in

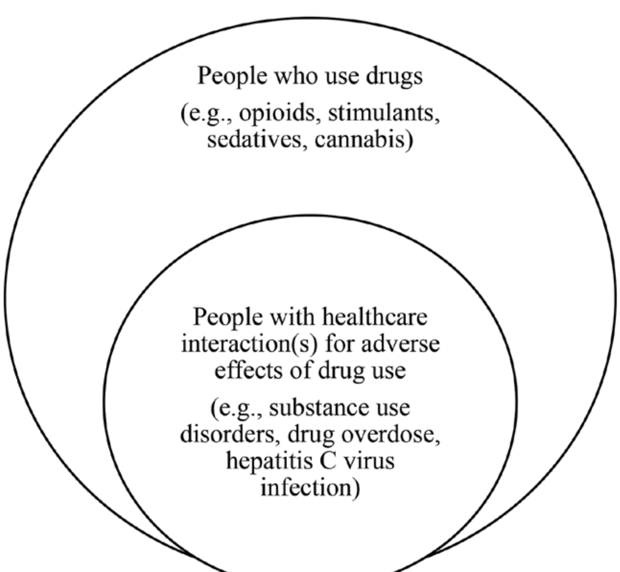

Figure. Possible populations of interest in drug use-related studies.

secondary data analysis.[7] Uninsured individuals using drugs may instead seek care outside of mainstream medical settings, and some patients may be reluctant to disclose their underlying substance use due to stigma. Ultimately, the challenges in ascertaining drug use create a fractured and hidden patchwork record of care for drug-related harms.

Coding schema, such as the International Classification of Disease (ICD), has codes for drug use disorders, dependence, or related complications. However, there is no single indicator of drug use. This distinction is important for research on harms related to drug use in general as not all people who use drugs meet the criteria for a disorder. Using administrative data to operationalize drug use as a single variable is therefore exceedingly difficult. Combinations of proxy diagnosis codes are often used to construct cohorts in research studies. Similar approaches have been applied in more general community settings, outside of the healthcare space.[8,9]

UTILITY AND VALIDITY OF DRUG USE-RELATED ALGORITHMS

Drug use-related algorithms can be used to characterize trends in clinical events, provide insight into treatment experiences and needs, and systematically evaluate prevention strategies (e.g., policies, medications, and treatment programs). These tools also can be used to develop predictive models to identify patients at greatest risk who would benefit most from targeted clinical interventions.[2,10–12]

Although several validation studies have been published recently,[4,13,14] many drug use-related algorithms remain unvalidated. A systematic review highlighted that the lack of validation studies for illicit drug use algorithms leads to issues of interpretability.[3] We expand upon this review to describe challenges in validation studies and how applications of these algorithms relate to the overall public health impact.

We conducted a narrative review of validation studies of drug use-related algorithms in healthcare data by beginning with four pre-identified validation studies in which the algorithm was compared to an external reference standard (i.e., gold standard comparator).[4,13–15] Next, we supplemented our search using the MEDLINE database for a combination of the following: a term for "validation study," drug use-related terms ("people who use drugs," "illicit drug use," "psychoactive drug use," and "overdose") and data-related terms ("claims," "health records," and "electronic medical records"). We included validation studies occurring in any country or year with a focus on drug use-related events. This resulted in one additional study for inclusion.[16]

The list of algorithms summarized in the Table is not comprehensive; rather, it shows examples of topic-specific validation studies, including operational definitions and performance metrics. Overall, algorithm definitions, study populations, and validation metrics varied widely. Validity measures were also highly variable (sensitivity range 47%–97%, specificity 61–99%). Though not necessarily a measure

Table. Published Algorithms Definitions and Validation Metrics

Target Measure	Study Population	Reference Group	Algorithm Definitions[a]	Validation Metrics
Patients with diagnosis of "drug abuse"[15,37]	Random selection of patients admitted to four hospitals in Alberta, Canada during 2003	Medical chart mention of drug use	Drug abuse (ICD-9): drug dependence or abuse, drug withdrawal, drug-induced mental disorders	Drug abuse (ICD-9): 55% sensitivity, 99% specificity, and 74% PPV
			Drug abuse (ICD-10): drug use disorders for opioids, stimulants, cannabis, sedatives, hallucinogens, inhalants, and other psychoactive drugs; counseling for drug use	Drug abuse (ICD-10): 47% sensitivity, 99% specificity, and 81% PPV
People who inject drugs[4]	People with hepatitis C virus testing, HIV testing, hepatitis B or C virus infections, HIV, or active tuberculosis that was reported during 1990-2015 in British Columbia, Canada	Public health interviews with mention of injection drug use	Drug misuse: claims for a medical visit or hospitalization that contained a drug-related code for drug use disorders or poisonings related to opioids, stimulants, sedatives, and other drugs	Drug misuse (≥1 claim in any setting): 90% sensitivity, 73% specificity, and 65% PPV
			Injection drug use: drug misuse classification for drugs that are known injectables, excluded cannabis and solvents.	Injection drug use (≥1 claim in any setting): 80% sensitivity, 81% specificity, and 71% PPV
			Injection-related infections: injection drug use classification with endocarditis, bacteremia or sepsis, osteomyelitis, or a skin or soft tissue infection	Injection-related infection: 60% sensitivity, 90% specificity, and 78% PPV
People with opioid overdose events[16]	People enrolled in a health plan in Oregon and Washington state during 2008–2014	Medical chart mention of opioid use	Opioid-related overdose: claims with code for opioid poisoning using ICD-9 for nonfatal events and ICD-10 for fatal events	Opioid-related overdose: 97% sensitivity, 85% specificity, and 87% PPV
Patients with injection drug use-associated infective endocarditis[14]	Randomly selected population of adults who were admitted to US Veterans Administration hospitals during 2010–2018 with a drug use-related diagnosis	Medical chart mention of drug use	Drug use during endocarditis admissions: code recorded for drug use/dependence/disorders and poisonings involving opioid, stimulants, and sedatives.	Drug use during the endocarditis admissions: 87% PPV
			Drug use within 6 months of the endocarditis admission: where a code was recorded for drug use/dependence/disorders and poisonings involving opioid, stimulants, and sedatives was recorded in the 6 months before or after endocarditis admission.	Drug use within 6 months of the endocarditis admission: 83% PPV
			Drug use within the endocarditis admission or hepatitis C virus during endocarditis admission: hepatitis C virus infection code during endocarditis admission or drug use/dependence/disorders and poisonings involving opioid, stimulants, and sedatives in the 6 months before or after endocarditis admission	Drug use within 6 months of the endocarditis admission or hepatitis C virus during endocarditis admission: 77% PPV

[a]Several studies evaluated more algorithm definitions than what is presented. For brevity, the algorithms presented here are not exhaustive of what was evaluated in each study. ICD indicates international classification of disease; PPV, positive predictive value.

of validity, due to its inherent reliance on prevalence,[17] the heterogeneity in positive predictive value (PPV) across studies is likely tied to the wide range of source populations. Some studies had very low specificity values. In these algorithms, the comparison group drew from populations who inherently have a much higher prevalence of drug use than the general patient population, such as those with endocarditis[5,14] or people screened for viral hepatitis.[4] Thus, a major barrier to understanding the validity of these algorithms is the lack of a gold standard comparison group.[13] Gold standard comparison groups (i.e., the population who truly has a characteristic) are difficult to obtain, particularly for marginalized populations. Self-reported drug use, if captured in a safe and confidential manner, would likely be a more reliable gold standard than medical chart review.

When applying validated algorithms to other populations, transportability issues may arise due to temporal and geographic differences in drug use prevalence[18] and coding practices.[19] For example, several validation studies have been conducted in Canada,[4,15] which use country-specific ICD-9 and ICD-10 dictionaries and have national health insurance coverage, potentially limiting algorithm portability to other countries. Notably, one US-based validation study examined the transportability of a drug overdose algorithm and found consistent performance measures across states and insurance providers.[16]

Many other drug use-related algorithms in use that are not validated have varying operational definitions, leading to potential issues with comparability across studies. For example, the Centers for Medicare and Medicaid Services developed an algorithm for drug use disorders that contains a wide range of diagnosis and procedure codes, including infants receiving care for prenatal substance exposure and family members receiving psychotherapy for another person's substance use.[20] To our knowledge, no validation studies have assessed the performance of this algorithm in real-world data.

Overall, the variability in algorithm sensitivity, specificity, and PPV underscores the need to consider the impact of information bias on the study results and looking one step beyond, public health policy. For example, consider a study of the incidence of all-cause mortality among a cohort of people with drug-use-associated endocarditis. The study results will be used to guide governmental resource allocation to address this issue. The algorithm, applied to cohort inclusion criteria, has high PPV but low sensitivity. Given that PPV is an important measure for cohort classification,[21] the study estimates of mortality might be accurate for the intended target population. However, low sensitivity would suggest an underestimation of the true burden of disease among the population. Depending on how these results are used, it could potentially impact public health resource allocation.

PUBLIC HEALTH IMPACT ON THE STUDY POPULATION

In the human subject protections paradigm,[22] researchers primarily focus on protecting study participants during the study. There is an often overlooked, yet critical detail: How might research impact the target population after the study? This detail is particularly important among historically marginalized populations. How algorithms are applied to clinical or public health practice is a fundamental consideration if a study intends to inform practice. Below, we discuss algorithms applications and their potential public health impact.

Inappropriate Reliance on Prescription Opioids in Predictive Scores

The number of studies aiming to predict overdose in administrative data have grown in tandem with increasing population-level overdose death rates.[23] These clinical algorithms overestimate overdose risk among low-risk patients, potentially depriving them of adequate pain care.[24] Specialized versions of these algorithms focus on overdose risk among patients receiving opioid analgesics, for example.[25] Conditioning on prescription opioid exposure may be invalid in recent years where prescription opioid deaths were replaced by illicit fentanyl overdose deaths.[26] Therefore, prescription opioid data should not be used as a proxy for drug or opioid use in general, particularly when applied in risk prediction studies. Yet, these data may be useful when explicitly examining prescription opioids as a treatment, particularly if nonprescription opioid use is rare in the target population. Inappropriate use of prescription opioid data may lead to a skewed narrative of the relationship between pain treatment, prescription drugs, and overdose that negatively impacts patients and mischaracterizes the true toxicologic harms facing people who use drugs.

Impact of Predictive Scores on Patient Care

A concerning trend in the use of algorithms is their application tied to access to care.[27–29] With prescription opioid analgesics, in particular, predictive scores or "risk scores" derived from large datasets have proliferated with the intended goal of reducing harms such as overdose.[24,30] A key limitation of these scores derived from administrative data is the lack of important clinical information such as measures of pain and function that preclude a correct contextualization of risk. Automated algorithms to identify "problematic" opioid prescription patterns in patients or "patients needing caution" have become prevalent in the clinical setting in recent years.[30] Despite the wide adoption of some of these risk algorithms, little evidence exists on the real-world performance and validation of these measures to correctly identify patients at high risk for opioid overdose. NarxCare[31] is a software that uses a widely used opioid risk score algorithm. However, it is unknown how often people are falsely flagged with high-risk scores. Additionally, the impact of its implementation has not been thoroughly evaluated in prospective studies. Algorithm audits and formal validation studies are urgently needed if these systems continue to be used in clinical and policy decision-making. Without a comprehensive algorithmic audit for these measures, these risk scores can harm people who use drugs and other patients in need of care.

Strategies to Improve the Public Health Impact

When using drug use-related algorithms, the impact of the study on the population of interest should be carefully considered before initiating a study. If the goal of a study is to improve the health of the population (as opposed to methodologic studies), the relevance to and implications of the research on the study population should be considered. Community involvement in epidemiology studies has existed for some time.[32–34] A promising approach gaining more attention is community-engaged and community-driven research.[35,36] Future research will have a greater public health reach when people with lived experience are directly involved in developing research questions, study procedures, and meaningful inclusion practices.[37,38] In real-world data studies, mixed methods studies can provide a much richer picture of the question at hand. Not only can qualitative interviews provide a better understanding of real-world data limitations but they can also improve the translation of research into practice by understanding the greater public health context.

CONCLUSIONS

Drug use-related algorithms applied to healthcare data are a potential tool to understand the urgent health needs facing people who use drugs. The variability in algorithm definitions and a lack of validation studies generate concern in study validity, comparability, and research implications. However, healthcare data are one of our main ways to understand the burden of disease among large populations. Without these data, we would be more reliant on cohort studies involving primary data collection, which would be neither feasible nor timely enough for the urgency of the health of people who use drugs. While drug use-related algorithms are imperfect, a balance must be struck between maximizing the study's validity and addressing the public health need.

ABOUT THE AUTHORS

MARY C. FIGGATT is a PhD Candidate in Epidemiology at the University of North Carolina Gillings School of Global Public Health. She works in the field of substance use epidemiology and has a specific focus on harm reduction, drug overdose, and infectious diseases. Her research methodologies incorporate a range of approaches including community-engaged studies, small-scale surveys, and analysis of large healthcare data.

NABARUN DASGUPTA is a scientist who studies drugs and infectious diseases. His passion is telling true stories about health, with numbers. Centered in epidemiology, his multidisciplinary approach draws on field research, large database analytics, laboratory investigations, randomized trials, and community-based interventions. Through his work, he aims to amplify community and patient voices in public health.

ASHER SCHRANZ is an Assistant Professor of Medicine in the Division of Infectious Diseases at the University of North Carolina School of Medicine. He is a practicing infectious diseases physician and clinical researcher. His research focuses on the intersection of infectious diseases and substance use disorders.

JUAN M. HINCAPIE-CASTILLO, PharmD, MS, PhD is an Assistant Professor of Epidemiology at the Gillings School of Global Public Health at the University of North Carolina at Chapel Hill. He works at the intersection of legal and pharmacoepidemiology leveraging large real-world data sources to evaluate and promote evidence-based policymaking, primarily applied to improving prescribing policies and the provision of equitable pain management and safe psychotropic medication use.

YVONNE GOLIGHTLY is a Professor and Assistant Dean for Research at the University of Nebraska Medical Center (UNMC) College of Allied Health Professions. She has appointments in the UNMC Department of Epidemiology and the University of North Carolina Department of Epidemiology. She is an injury epidemiologist and a physical therapist, and her research interests include musculoskeletal injury, chronic musculoskeletal pain, health equity, opioid use and overdose, and community-based research.

STEPHEN MARSHALL is a professor of epidemiology at the Gillings School of Global Public Health, University of North Carolina at Chapel Hill. He is also the Director of the University of North Carolina Injury Prevention Research Center. His research is focused on injury, violence, and overdose epidemiology. Areas of research include overdose prevention, sport-related injuries and musculoskeletal conditions, traumatic brain injury, suicide prevention, transportation injury, and injuries in the military.

REFERENCES

1. Hedegaard H, Miniño A, Spencer MR, Warner M. Drug overdose deaths in the United States, 1999–2020. *NCHS Data Brief*. 2021;428:8.
2. See I, Gokhale RH, Geller A, et al. National Public Health Burden estimates of endocarditis and skin and soft-tissue infections related to injection drug use: a review. *J Infect Dis*. 2020;222(Supplement_5):S429–S436.
3. McGrew KM, Homco JB, Garwe T, et al. Validity of International Classification of Diseases codes in identifying illicit drug use target conditions using medical record data as a reference standard: a systematic review. *Drug Alcohol Depend*. 2020;208:107825.
4. Janjua NZ, Islam N, Kuo M, et al; BC Hepatitis Testers Cohort Team. Identifying injection drug use and estimating population size of people who inject drugs using healthcare administrative datasets. *Int J Drug Policy*. 2018;55:31–39.
5. Ball LJ, Sherazi A, Laczko D, et al. Validation of an algorithm to identify infective endocarditis in people who inject drugs. *Med Care*. 2018;56:e70–e75.
6. Crystal S, Akincigil A, Bilder S, Walkup JT. Studying prescription drug use and outcomes with medicaid claims data strengths, limitations, and strategies. *Med Care*. 2007;45(10 SUPL):S58–S65.
7. Bobashev G, Warren L, Wu LT. Predictive model of multiple emergency department visits among adults: analysis of the data from the National Survey of Drug Use and Health (NSDUH). *BMC Health Serv Res*. 2021;21:280.
8. Aceijas C, Friedman SR, Cooper HLF, Wiessing L, Stimson GV, Hickman M. Estimates of injecting drug users at the national and local level in developing and transitional countries, and gender and age distribution. *Sex Transm Infect*. 2006;82(Suppl 3):iii10–17.
9. Brady JE, Friedman SR, Cooper HLF, Flom PL, Tempalski B, Gostnell K. Estimating the prevalence of injection drug users in the U.S. and in large U.S.metropolitan areas from 1992 to 2002. *J Urban Health*. 2008;85:323–351.
10. Schranz AJ, Fleischauer A, Chu VH, Wu LT, Rosen DL. Trends in drug use-associated infective endocarditis and heart valve surgery, 2007 to 2017: a study of statewide discharge data. *Ann Intern Med*. 2019;170:31–40.
11. Barocas JA, Morgan JR, Wang J, McLoone D, Wurcel A, Stein MD. Outcomes associated with medications for opioid use disorder among persons hospitalized for infective endocarditis. *Clin Infect Dis*. 2020;472–478.
12. Barocas JA, Gai MJ, Amuchi B, Jawa R, Linas BP. Impact of medications for opioid use disorder among persons hospitalized for drug use-associated skin and soft tissue infections. *Drug Alcohol Depend*. 2020;215:108207.
13. McGrew KM, Carabin H, Garwe T, et al. Validity of ICD-based algorithms to estimate the prevalence of injection drug use among infective endocarditis hospitalizations in the absence of a reference standard. *Drug Alcohol Depend*. 2020;209:107906.

14. Kobayashi T, Beck B, Miller A, Polgreen P, O'Shea AMJ, Ohl ME. Positive predictive values of 2 algorithms for identifying patients with intravenous drug use–associated endocarditis using administrative data. *Open Forum Infect Dis*. 2020;7:ofaa201.
15. Quan H, Li B, Saunders LD, et al; IMECCHI Investigators. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*. 2008;43:1424–1441.
16. Green CA, Perrin NA, Hazlehurst B, et al. Identifying and classifying opioid-related overdoses: a validation study. *Pharmacoepidemiol Drug Saf*. 2019;28:1127–1137.
17. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56:45–50.
18. Substance Abuse and Mental Health Services Administration. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2020 National Survey on Drug Use and Health*. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2020:156.
19. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 2005;40(5 Pt 2):1620–1639.
20. Centers for Medicare and Medicaid Services. Chronic Conditions Data Warehouse. Available at: https://www2.ccwdata.org/web/guest/condition-categories. Accessed February 5, 2022.
21. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65:343–349.e2.
22. National Institute of Health. Human Subjects Protections. NIH Grants Policy Statement. Available at: https://grants.nih.gov/grants/policy/nihgps/HTML5/section_4/4.1.15_human_subjects_protections.htm. Accessed February 16, 2022.
23. Tseregounis IE, Henry SG. Assessing opioid overdose risk: a review of clinical prediction models utilizing patient-level data. *Transl Res*. 2021;234:74–87.
24. Siegel Z. In a World of stigma and bias, can a computer algorithm really predict overdose risk?: A machine-learning algorithm is being deployed across America to prevent overdose deaths. but could it be causing more pain? *Ann Emerg Med*. 2022;79:A16–A19.
25. Metcalfe L, Murrelle EL, Vu L, et al. Independent validation in a large privately insured population of the risk index for serious prescription opioid–induced respiratory depression or overdose. *Pain Med*. 2020;21:2219–2228.
26. Aubry L, Carr BT. Overdose, opioid treatment admissions and prescription opioid pain reliever relationships: United States, 2010–2019. *Front Pain Res*. 2022;3:884674. Available at: https://www.frontiersin.org/articles/10.3389/fpain.2022.884674. Accessed August 15, 2022.
27. Brown S, Davidovic J, Hasan A. The algorithm audit: scoring the algorithms that score us. *Big Data Soc*. 2021;8:2053951720983865.
28. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Int Med*. 2018;178:1544–1547.
29. Metaxa D, Park JS, Robertson RE, Karahalios K, Wilson C, Hancock J, et al. *Auditing Algorithms: Understanding Algorithmic Systems from the Outside In*. Found Trends Hum Comput Interact. Vol 14. 272–344. Available at: https://hci.stanford.edu/publications/paper.php?id=447. Accessed May 3, 2022.
30. Oliva JD. *Dosing Discrimination: Regulating PDMP Risk Scores*. Social Science Research Network; 2021.
31. NarxCare: Substance Abuse Software Solution. Bamboo Health. Available at: https://bamboohealth.com/solutions/narxcare/. Accessed August 8, 2022.
32. Wing S. Whose epidemiology, whose health? *Int J Health Serv*. 1998;28:241–252.
33. Rhodes T, Lancaster K. Evidence-making interventions in health: a conceptual framing. *Soc Sci Med*. 2019;238:112488.
34. Camelo Castillo W, Heath N, Kim J, et al. Engaging stakeholders in pharmacoepidemiology research: current state and recommendations. *Pharmacoepidemiol Drug Saf*. 2019;28:766–776.
35. Simon C, Brothers S, Strichartz K, et al. We are the researched, the researchers, and the discounted: the experiences of drug user activists as researchers. *Int J Drug Policy*. 2021;98:103364.
36. Salazar ZR, Vincent L, Figgatt MC, Gilbert MK, Dasgupta N. Research led by people who use drugs: centering the expertise of lived experience. *Subst Abuse Treat Prev Policy*. 2021;16:70.
37. Dasgupta N. "He is the object of information": the intersection of big data and the opioid crisis. *Am J Public Health*. 2018;108:1122–1123.
38. Evans EA, Delorme E, Cyr K, Goldstein DM. A qualitative study of big data and the opioid epidemic: recommendations for data governance. *BMC Med Ethics*. 2020;21:101.