



# HHS Public Access

Author manuscript

*Value Health*. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

*Value Health*. 2023 September ; 26(9): 1321–1324. doi:10.1016/j.jval.2023.02.019.

## Use of Big Data and Ethical Issues for Populations with Substance Use Disorder

**Elizabeth Evans, PhD, MA [Professor],**

Department of Health Promotion and Policy, School of Public Health and Health Sciences, University of Massachusetts Amherst, 312 Arnold House, 715 North Pleasant Street, Amherst, MA 01003, USA

**Kimberley H. Geissler, PhD\* [Associate Professor]**

Department of Health Promotion and Policy, School of Public Health and Health Sciences, University of Massachusetts Amherst, 337 Arnold House, 715 North Pleasant Street, Amherst, MA 01003, USA

### Abstract

With expanding data availability and computing power, health research is increasingly relying on big data from a variety of sources. We describe a state-level effort to address aspects of the opioid epidemic through public health research, which has resulted in an expansive data resource combining dozens of administrative data sources in Massachusetts. The Massachusetts Public Health Data Warehouse is a public health innovation that serves as an example of how to address the complexities of balancing data privacy and access to data for public health and health services research. We discuss issues of data protection and data access and provide recommendations for ethical data governance. Keeping these issues in mind, the use of this data resource has the potential to allow for transformative research on critical public health issues.

### Précis:

We review issues of data protection and provide recommendations for ethical data governance for a state administrative data warehouse created to address the opioid epidemic.

---

Availability of data and accompanying computing power have expanded health research in all areas – basic science, biological basis of disease, precision medicine, clinical care, and public health.<sup>1–6</sup> In public health research, much of the data related to health care use, health insurance coverage, and use of social services has historically been siloed in individual health systems, individual state agencies, or individual insurers, and are not available for research purposes despite broad public support for data sharing for research.<sup>7</sup> However, with expanded interest in use of these data and nationwide efforts to improve interoperability for both ongoing clinical care and data availability for research,<sup>8,9</sup> there are an increasing number of private, local, state, and federal data providers who are interested in the use of

---

\*Corresponding Author: Kimberley H. Geissler, kgeissler@umass.edu.

**Financial Disclosure:** None reported.

Initial Submission Responsive to Call for Papers for Value in Health's 25th Anniversary Issue

data for research purposes including for those with substance use disorder (SUD).<sup>10</sup> We describe a state-level effort to address aspects of the opioid epidemic through public health research, which has resulted in an expansive statewide data resource in Massachusetts.

## Massachusetts Public Health Data Warehouse – A statewide data resource in Massachusetts

In Massachusetts, the Massachusetts Public Health Data (PHD) Warehouse is a public health innovation that serves as an example of how to address the complexities of balancing data privacy and access to data for health services research. Established by legislative mandate in 2015 and constructed and managed by the Massachusetts Department of Public Health (MDPH), the PHD Warehouse is a novel US example of using “administrative big data” from a variety of state government sources for policy-relevant research on health services utilization and health outcomes, including measures such as mortality and out of hospital non-fatal overdoses.<sup>11,12</sup>

Among the many strengths for advancing research,<sup>11</sup> the PHD Warehouse is comprised of individually linked administrative data from more than twenty state sources on all Massachusetts residents aged 11 and older with public or private health insurance. The legislation that initiated the creation of the PHD Warehouse prioritized the analysis of fatal and non-fatal opioid overdoses.<sup>13,14</sup> Thus, events recorded in the PHD Warehouse include treatment for opioid and other substance use disorders and overdose events, along with information from the prescription drug monitoring program and cash purchases of opioid prescriptions. However, other individual-level demographics and events are included as well, such as diagnosis and treatment for physical and mental health conditions, public welfare benefit receipt, and mortality. At a contextual level, the PHD Warehouse also includes data from the U.S. Census, the MDPH Naloxone Distribution Program, Drug Seizure data, and neighborhood level measures of privilege and deprivation.

Over the past five years, PHD Warehouse data have been used extensively to generate a significant and growing body of policy-relevant public health and health services research related to opioid use disorder. For example, studies using PHD Warehouse data have documented: (1) prevalence of opioid use disorder and variation in prevalence rates by specific factors including changes over time, population characteristics, and geographic location<sup>15,16</sup>; (2) potentially inappropriate opioid prescribing practices by individual characteristics<sup>17–19</sup>; (3) information about treatment with medications for opioid use disorder (MOUD)<sup>20,21</sup>; and (4) prevalence of fatal and non-fatal overdoses by vulnerable populations, including known (e.g., Veterans, pregnant people, and adolescents)<sup>16,22,23</sup> and newly-identified (e.g., construction workers).<sup>24</sup> These and other findings have been used to design health surveillance efforts, allocate resources, conduct community outreach, and plan interventions.<sup>25–27</sup>

Currently, the PHD Warehouse is earmarked for research on new and emergent public health issues for priority populations with opioid use disorder, including maternal health equity research and research on the health impacts of COVID-19.<sup>26</sup> As a notable example, these data are critical to evaluating the implementation, outcomes, and costs of providing access

to medications to treat opioid use disorder in jail settings,<sup>28</sup> a major policy change that is currently underway in Massachusetts.

## Leveraging administrative big data for research purposes – data protection and data access

This volume of research activity has been made possible by the policies and procedures that MDPH established to create the PHD Warehouse and manage it for research purposes. Immediately following passage of the legislative mandate, MDPH worked to form the necessary legal, contractual, and data use agreements with multiple data contributors to be able to receive data. Critically for data privacy and security, the legislative act mandating the creation of the PHD Warehouse noted “Such information or data shall not be considered a public record, shall be exempt from disclosure under section 10 of chapter 66 and shall not be subject to subpoena or discovery or admissible as evidence in any action of any kind in any court or before any other tribunal, board, agency or person.”<sup>13</sup> MDPH also developed with its partners a plan for the computing environment and for the data architecture.<sup>29</sup> MDPH planned for the PHD Warehouse “backbone” to be built on the Massachusetts all-payer claims database (APCD), to which state datasets were to be linked.<sup>29</sup> Massachusetts has near universal health insurance coverage, which means that the APCD – at the time of creation of the initial PHD Warehouse – was a near-complete census of the state’s population. The decision to use APCD as the “backbone” allowed for a uniquely comprehensive view of the opioid crisis. MDPH created processes to link individual-level data through deterministic match protocols that required exact identifier matches, followed by validity checks and then deidentification of the data for analysis while avoiding redisclosure.

Innovative solutions were used to protect data privacy, even beyond the standards set by federal and state law, and to create mechanisms for data sharing with researchers. Plans included a secure analytic environment through which only de-identified data would be made available to qualified researchers for analysis. As additional technical safeguards, the individual level data are stored in separate datasets and linked only temporarily for analysis, with resultant datasets automatically destroyed at the completion of an analytic session; small cell sizes are automatically suppressed.<sup>26,29</sup> A year after the legislative mandate went into effect, MDPH had executed the needed data use agreements and received and linked data provided by various state agencies.<sup>29,30</sup>

Also critical to the successes of the PHD Warehouse have been the mechanisms that MDPH has used to solicit research proposals and review and approve them for implementation (see <https://www.mass.gov/public-health-data-warehouse-phd>). MDPH hosts community-engaged forums to generate topics and set priorities for the analyses to be conducted. MDPH also established a data governance structure and shared several resources to facilitate the processes for submitting data proposals and conducting data analysis. These included, for example, a PHD User Manual, a webinar, a list of questions and answers regarding approval processes and protocols for PHD data projects, and a forum for data users to share statistical code and solutions to data challenges. These protocols and resources have created the ability

to access, integrate, and cooperatively use these data in a coordinated way. However, as noted below, expansion of topics that could be studied could substantially improve the utility of the PHD Warehouse.

## Recommendations for ethical data governance

While a valuable research resource, the PHD Warehouse also presents ethical concerns which, if unaddressed, may undermine its benefits.<sup>12</sup> We have documented how concerns regarding administrative big data on opioid use are rooted in potential privacy infringements due to linkage of previously distinct data systems, increased profiling and surveillance capabilities, the limitless lifespan of data, and the lack of explicit informed consent.<sup>12</sup> Also problematic is the inability of affected groups to control how big data are used, the potential of big data to increase stigmatization and discrimination of those affected despite data anonymization and uses that ignore or perpetuate biases.<sup>12</sup>

Given these concerns, we examined the perspectives of big data stakeholders (which we defined as patient advocates, researchers, and data gatekeepers) with knowledge of the PHD Warehouse to identify critical aspects of ethical big data governance.<sup>12</sup> Based on this work, ethical big data governance should offer ways to narrow the big data divide. This might include, for example, prioritizing research addressing health equity among racial/ethnic groups and/or geographic areas, setting off-limits topics/methods, and recognizing blind spots in the data. Regarding blind spots, one important limitation of these data is that they generally omit detailed information on criminal-legal-carceral experiences, thereby precluding full assessment of the relationship between these experiences and opioid use disorder treatment and outcomes.<sup>31</sup> Such omissions contribute to significant “blind spots” in our understanding of how to address the opioid overdose epidemic.

Additionally, as data quality changes among different data contributors, attention should be paid to how this changes the population and conclusions drawn. For example, with the Supreme Court ruling in *Gobielle v Liberty Mutual*,<sup>32</sup> self-insured employers were not required to report data to the MA APCD as of 2016. This has resulted in changes to the composition of the included population over time; if these changes are not accounted for in the analysis and description of results, specific populations could be disproportionately impacted by this omission. Additionally, deidentification of the data to protect privacy to a higher level than state and federal standards may result in some loss of detail needed for specific types of research questions.

As another recommendation for ethical data governance, ways to enact shared data governance are needed, for example via community advisory boards. This shared data governance is important for the cultivation of public trust, which is needed to earn social license for big data uses. This includes, for example, instituting technical safeguards and other data stewardship responsibilities, engagement of the public, and communication of the greater good. In the case of the PHD Warehouse, expansion of research topics beyond those with opioid use disorder could result in a substantially better understanding of public health priority areas. Although the legislation mandating its creation gave some discretion to the use of the constructed data set for “additional priorities for the reduction of morbidity and

mortality,”<sup>13</sup> additional topics have not been available for research. Shared data governance could allow for broader use of such a dataset with appropriate checks and balances to achieve the primary goals while expanding research use appropriately over time.

A final recommendation is to refocus ethical approaches. This means an examination of protection from presumed harms, consent, and individual control of data uses, as is typically done as a part of usual regulatory review, and also a consideration of respect for patients and society, issues of equity and justice, and methods to cultivate patient and public trust in public institutions. Finally, it also means giving primacy to community engagement, which extends concerns beyond individual-level health to also consider population health and the public’s interests.

## Other settings can learn from Massachusetts

The PHD Warehouse is similar to the comprehensive nationwide population-based registries that have been used to conduct public health research for decades in other countries.<sup>33,34</sup> New insights can be generated by connecting multiple data sources, thereby offering invaluable data and tools for identifying what is working and for whom.<sup>35,36</sup> MDPH’s has effectively communicated findings generated from this resource to researchers, clinicians, the legislature, and the general public. Lessons from Massachusetts can be used by other settings, including states, regions, and nations to build similar administrative big data warehouses on population health and use them to advance research for the public good. One key takeaway is that having a dataset such as a true all-payer claims dataset or equivalent that includes health care interactions for the full population of interest is a useful starting point. Other administrative datasets that provide a census (e.g., tax records) that could be linked with other individual level data providing key indicators could be an alternate strategy in the absence of such health encounter data.

## Conclusion

Focusing on the PHD Warehouse from Massachusetts, we note that although major initiatives such as this require intense focus on technical security for data safeguards and the ethical production and use of data, they may result in expansive new knowledge at the intersection of health and social welfare that could not be gained otherwise. This knowledge has the potential to transform public health initiatives and expand the capacity of public health and health systems to identify high impact targets for change.

## Acknowledgements:

The authors thank Valerie Evans for research assistance.

## Funding Support:

Research reported in this publication was supported by the Greenwall Foundation, the National Institute on Drug Abuse of the National Institutes of Health under award number 1UG1DA050067-01, and the National Institute of Mental Health of the National Institutes of Health under award number 5R34MH123628. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Greenwall Foundation.

## References

1. McCarthy D State All-Payer Claims Databases: Tools for Improving Health Care Value, Part 1 -- How States Establish an APCD and Make It Functional. Commonwealth Fund 2020.
2. Blewett LA, Mac Arthur NS, Campbell J. The Future of State All Payer Claims Databases. *J Health Polit Policy Law*. 2022.
3. Carrasco-Ramiro F, Peiró-Pastor R, Aguado B. Human genomics projects and precision medicine. *Gene Therapy*. 2017;24(9):551–561. [PubMed: 28805797]
4. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nature Reviews Cardiology*. 2016;13(6):350–359. [PubMed: 27009423]
5. Amarasingham R, Audet AM, Bates DW, et al. Consensus Statement on Electronic Health Predictive Analytics: A Guiding Framework to Address Challenges. *EGEMS (Wash DC)*. 2016;4(1):1163. [PubMed: 27141516]
6. Wu PY, Cheng CW, Kaddi CD, Venugopalan J, Hoffman R, Wang MD. -Omic and Electronic Health Record Big Data Analytics for Precision Medicine. *IEEE Trans Biomed Eng*. 2017;64(2):263–273. [PubMed: 27740470]
7. Hensley S Poll: Most Americans Would Share Health Data For Research. NPR. January 9, 2015.
8. Frakt AB, Bagley N. Protection or Harm? Suppressing Substance-Use Data. *New England Journal of Medicine*. 2015;372(20):1879–1881. [PubMed: 25875196]
9. Butler JM, Becker WC, Humphreys K. Big Data and the Opioid Crisis: Balancing Patient Privacy with Public Health. *Journal of Law, Medicine & Ethics*. 2018;46(2):440–453.
10. Geissler KH, Evans EA, Johnson JK, Whitehill JM. A Scoping Review of Data Sources for the Conduct of Policy-Relevant Substance Use Research. *Public Health Reports*. 2022;137(5):944–954. [PubMed: 34543133]
11. Evans EA, Delorme E, Cyr KD, Geissler KH. The Massachusetts public health data warehouse and the opioid epidemic: A qualitative study of perceived strengths and limitations for advancing research. *Prev Med Rep*. 2022;28:101847. [PubMed: 35669857]
12. 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(1):101. [PubMed: 33087123]
13. Commonwealth of Massachusetts. FY2018 Budget Summary: Outside Section 48. [https://budget.digital.mass.gov/bb/gaa/fy2018/os\\_18/h48.htm](https://budget.digital.mass.gov/bb/gaa/fy2018/os_18/h48.htm). Published 2018. Accessed February 10, 2023.
14. Commonwealth of Massachusetts. Public Health Data Warehouse (PHD). <https://www.mass.gov/public-health-data-warehouse-phd>. Accessed February 10, 2023.
15. Barocas JA, White LF, Wang J, et al. Estimated Prevalence of Opioid Use Disorder in Massachusetts, 2011–2015: A Capture-Recapture Analysis. *Am J Public Health*. 2018;108(12):1675–1681. [PubMed: 30359112]
16. Schiff DM, Nielsen T, Terplan M, et al. Fatal and Nonfatal Overdose Among Pregnant and Postpartum Women in Massachusetts. *Obstet Gynecol*. 2018;132(2):466–474. [PubMed: 29995730]
17. Rose AJ, Bernson D, Chui KKH, et al. Potentially Inappropriate Opioid Prescribing, Overdose, and Mortality in Massachusetts, 2011–2015. *J Gen Intern Med*. 2018;33(9):1512–1519. [PubMed: 29948815]
18. Rose AJ, McBain R, Schuler MS, et al. Effect of Age on Opioid Prescribing, Overdose, and Mortality in Massachusetts, 2011 to 2015. *J Am Geriatr Soc*. 2019;67(1):128–132. [PubMed: 30471102]
19. Stopka TJ, Amaravadi H, Kaplan AR, et al. Opioid overdose deaths and potentially inappropriate opioid prescribing practices (PIP): A spatial epidemiological study. *Int J Drug Policy*. 2019;68:37–45. [PubMed: 30981166]
20. Laroche MR, Bernson D, Land T, et al. Medication for Opioid Use Disorder After Nonfatal Opioid Overdose and Association With Mortality: A Cohort Study. *Ann Intern Med*. 2018;169(3):137–145. [PubMed: 29913516]



21. Larochelle M, Stopka TJ, Xuan Z, Liebschutz JM, Walley AY. Medication for Opioid Use Disorder After Nonfatal Opioid Overdose and Mortality. *Annals of Internal Medicine*. 2019;170(6):430–431.
22. Jasuja GK, Ameli O, Miller DR, et al. Overdose risk for veterans receiving opioids from multiple sources. *Am J Manag Care*. 2018;24(11):536–540. [PubMed: 30452210]
23. Chatterjee A, Larochelle MR, Xuan Z, et al. Non-fatal opioid-related overdoses among adolescents in Massachusetts 2012–2014. *Drug Alcohol Depend*. 2019;194:28–31. [PubMed: 30391835]
24. Hawkins D, Roelofs C, Laing J, Davis L. Opioid-related overdose deaths by industry and occupation-Massachusetts, 2011–2015. *Am J Ind Med*. 2019;62(10):815–825. [PubMed: 31347714]
25. Adams JW, Savinkina A, Fox A, et al. Modeling the cost-effectiveness and impact on fatal overdose and initiation of buprenorphine-naloxone treatment at syringe service programs. *Addiction*. 2022;117(10):2635–2648. [PubMed: 35315148]
26. Bharel M, Bernson D, Averbach A. Using Data to Guide Action in Response to the Public Health Crisis of Opioid Overdoses. *NEJM Catalyst*. 2020;1(5).
27. Larochelle MR, Bernstein R, Bernson D, et al. Touchpoints – Opportunities to predict and prevent opioid overdose: A cohort study. *Drug and Alcohol Dependence*. 2019;204:107537. [PubMed: 31521956]
28. Evans EA, Stopka TJ, Pivovarova E, et al. Massachusetts Justice Community Opioid Innovation Network (MassJCOIN). *J Subst Abuse Treat*. 2021;128:108275. [PubMed: 33483222]
29. Land T, Bernson D, Hood M, et al. Building a prototype of a statewide public health data warehouse: Data privacy and security issues addressed by the Massachusetts Chapter 55 Opioid Initiative. In. Unpublished.2018.
30. Land T, Scurria Morgan E, Bernson D, et al. Developing a collaborative approach to addressing critical public health issues: A case study of the Massachusetts Chapter 55 Opioid Initiative. In. Unpublished.2018.
31. Evans EA. Commentary on Adams et al.: using administrative big data for the public good. *Addiction*. 2022;117(10):2649–2650. [PubMed: 35818700]
32. The Consequences Of Gobeille v. Liberty Mutual For Health Care Cost Control. *Health Affairs Blog Web site*. Updated March 10, 2016. Accessed.
33. Amundsen EJ, Bretteville-Jensen AL, Rossow I. Patients admitted to treatment for substance use disorder in Norway: a population-based case-control study of socio-demographic correlates and comparative analyses across substance use disorders. *BMC Public Health*. 2022;22(1):792. [PubMed: 35443672]
34. Thygesen LC, Daasnes C, Thaulow I, Bronnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand J Public Health*. 2011;39(7 Suppl):12–16. [PubMed: 21898916]
35. Blanco C, Wiley TRA, Lloyd JJ, Lopez MF, Volkow ND. America’s opioid crisis: the need for an integrated public health approach. *Transl Psychiatry*. 2020;10(1):167. [PubMed: 32522999]
36. Evans E, Grella CE, Murphy DA, Hser YI. Using administrative data for longitudinal substance abuse research. *J Behav Health Serv Res*. 2010;37(2):252–271. [PubMed: 18679805]

**Highlights:**

1. Increasing data and computational availability means that clinical, public health, and health services research is increasingly relying on big data, including for questions related to care for populations with substance use disorder (SUD).
2. We describe a state-level effort to address aspects of the opioid epidemic through public health research through the construction of the Massachusetts Public Health Data Warehouse, discuss issues of data protection and data access, and provide recommendations for ethical data governance.
3. Major initiatives such as this require intense focus on technical security for data safeguards and the ethical production and use of data and may result in new knowledge at the intersection of health and social welfare that could not be gained otherwise.