Preventive Medicine Reports 28 (2022) 101847

Contents lists available at ScienceDirect

Preventive Medicine Reports

journal homepage: www.elsevier.com/locate/pmedr

The Massachusetts public health data warehouse and the opioid epidemic: A qualitative study of perceived strengths and limitations for advancing research

Elizabeth A. Evans *, Elizabeth Delorme, Karl D. Cyr, Kimberley H. Geissler

Department of Health Promotion and Policy, School of Public Health and Health Sciences, University of Massachusetts Amherst, Amherst, MA, USA

ARTICLE INFO

Keywords:
Big data
Administrative data linkage
Opioid epidemic
Population health
Qualitative research

ABSTRACT

Due to the opioid overdose epidemic, Massachusetts created a Public Health Data Warehouse, encompassing individually-linked administrative data on most of the population as provided by more than 20 systems. As others seek to assemble and mine big data on opioid use, there is a need to consider its research utility. To identify perceived strengths and limitations of administrative big data, we collected qualitative data in 2019 from 39 stakeholders with knowledge of the Massachusetts Public Health Data Warehouse. Perceived strengths included the ability to: (1) detect new and clinically significant relationships; (2) observe treatments and services across institutional boundaries, broadening understanding of risk and protective factors, treatment outcomes, and intervention effectiveness; (3) use geographic-specific lenses for community-level health; (4) conduct rigorous "real-world" research; and (5) generate impactful findings that legitimize the scope and impacts of the opioid epidemic and answer urgent questions. Limitations included: (1) oversimplified information and imprecise measures; (2) data access and analysis challenges; (3) static records and substantial lag times; and (4) blind spots that bias or confound results, mask upstream or root causes, and contribute to incomplete understanding. Using administrative big data to conduct research on the opioid epidemic offers advantages but also has limitations which, if unrecognized, may undermine its utility. Findings can help researchers to capitalize on the advantages of big data, and avoid inappropriate uses, and aid states that are assembling big data to guide public health practice and policy.

1. Introduction

The U.S. opioid overdose epidemic has led to extraordinary numbers of accidental injuries, premature deaths, and comorbid infectious diseases, contributing to marked decreases in life expectancy. (Case and Deaton, 2015; Murphy et al., 2017; Scholl et al., 2018) In Massachusetts, a key component of the response to the opioid overdose epidemic has been the creation of the Public Health Data (PHD) Warehouse. (Massachusetts Department of Public Health, 2017; Massachusetts Department of Public Health, 2016) Established by legislative mandate in 2015 and managed by the Massachusetts Department of Public Health, the PHD Warehouse encompasses individual-level linked administrative data as provided by more than twenty sources on all Massachusetts residents aged 11 and older with public or private health insurance, covering more than 98% of the state's population. (Bharel et al., 2020) Events recorded in the PHD Warehouse include treatment for opioid and other substance use disorders, diagnosis and treatment for physical and mental health conditions, receipt of public welfare benefits, insurance claims, involvement with the criminal justice system, and mortality. (Geissler et al., 2021) Findings generated thus far from PHD data have informed public health surveillance efforts, resource allocation, community outreach, and intervention planning. (Bharel et al., 2020).

For more than 20 years, individually-linked administrative data have been used to support research in the US, primarily to understand health services utilization, treatment outcomes, and costs. (Anglin et al., 2013; Ettner et al., 2006; Evans et al., 2010; Krebs et al., 2018; Weissman, 2020) In relation to addiction research, an established strength of such data is the ability to investigate interactions with healthcare, social services, and criminal justice systems, and related outcomes, as they arise and influence each other over time. (Evans et al., 2010; Smart et al.,

* Corresponding author at: 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.
E-mail address: eaevans@umass.edu (E.A. Evans).

https://doi.org/10.1016/j.pmedr.2022.101847
Received 20 January 2022; Received in revised form 24 May 2022; Accepted 28 May 2022
Available online 31 May 2022

2019) Unique to the PHD Warehouse, it encompasses 98% of the Massachusetts population, not only individuals with opioid or other substance use disorders. Also, it links together data from a broader than ever set of agencies and institutions, encompassing data from more than two dozen sources (see appendix). These strengths offer opportunities to comprehensively assess the causes and consequences of substance use disorders. For example, the PHD Warehouse now includes data on vital statistics (births and deaths), acute care hospitalizations, ambulance services, prescriptions, treatment for addiction and mental health conditions, all payer claims data, incarcerations in jails and prisons, and receipt of public welfare benefits. Moreover, the PHD Warehouse was developed by legislative mandate (Chapter 55 of the Acts of 2015; amended by Chapter 133 of the Acts of 2016) and it is managed by the state's public health department, establishing it as a potential public good. Individuals are able to access PHD Warehouse data for research purposes via a process that involves a periodic call for proposals and assessment of applications according to criteria such as eligibility and fit with research priorities. These characteristics mean that the PHD Warehouse approximates the comprehensive nationwide population health registries that have been used to conduct research for some time in other countries. (Hsing and Ioannidis, 2015; Lee et al., 2021; Thygesen et al., 2011; Winickoff, 2006) Recently, states and other jurisdictions have sought to learn from Massachusetts how to build and manage similar administrative big data warehouses. (Hallvik et al., 2021; KPMG, 2017; Saloner et al., 2018; Smart et al., 2018; Smart et al., 2019).

The PHD Warehouse is one of the first US examples of using administrative big data, combining information from diverse sources (e.g., medical claims, social services, and justice systems) on an entire state's population to address opioid use disorders and advance public health. The PHD Warehouse has been mined to support significant research, (Bharel et al., 2020; Evans et al., 2019) but an assessment of its advantages and disadvantages for research has not been conducted. To address this knowledge gap, we aim to identify the perceived strengths and limitations of big data for conducting research on opioid use disorders. Findings can inform future research opportunities and aid other states that are working to assemble administrative big data to guide public health practice and policy.

2. Methods

2.1. Conceptual framework

We analyzed qualitative data that were collected as part of a study on the ethical issues of using big data to address the opioid overdose epidemic. We used the Kass Public Health Ethics Framework (Kass, 2001) which specifies that public health officials should communicate with and involve constituent communities, along with experts, to understand the benefits and risks of strategies to address public health threats. Thus, we solicited perspectives on the strengths and limitations of big data on opioid use, along with harms and benefits, as perceived by key groups: researchers who conduct analysis of big data on opioid use, gatekeepers of these data, and patient advocates.

2.2. Participants

We interviewed 39 key informants. Researchers were recruited from PHD Warehouse users (i.e., biomedical researchers, clinician-researchers, epidemiologists, data scientists), with priority given to authors of peer-reviewed publications. Gatekeepers (i.e., data managers, regulatory specialists, legal counsel, ethicists) were recruited from staff at the Massachusetts Department of Public Health who manage the PHD Warehouse and also from local agencies that create county-level big data repositories on opioid use. Patient advocates were recruited from community forums held by peer-led support networks for parents and families coping with opioid overdose, some of whom reported having been treated for opioid use. Individuals were invited to participate via flyers distributed at public meetings and direct email outreach.

2.3. Data collection

Semi-structured focus groups (5–10 participants/group) or individual interviews (to avoid power differentials or accommodate scheduling conflicts) were conducted in-person or by teleconference, followed by a socio-demographic questionnaire. Data were collected separately from researchers, gatekeepers, and advocates. Our findings on the ethical issues of big data are provided elsewhere. (Evans et al., 2020) Here, we analyze data on the uses of big data for research. Participants were asked, for example, to reflect on the limitations, strengths, and concerns of using big data on opioid use for research purposes and identify phenomena that shape opioid use but are not captured in big data. Participants were mostly asked to reflect on the Massachusetts PHD Warehouse and, when relevant, any other similar sources (e.g., the Veterans Administration, local data-to-action efforts).

Data were collected March-December 2019. Discussions lasted 1–2 h and were held privately in-person or by video-conference. Participants were consented, assured that findings would be anonymized, and compensated $100. Interviews were digitally recorded, professionally transcribed, and transcripts were reviewed for accuracy. All procedures were approved by the relevant Institutional Review Board.

2.4. Data analysis

We conducted thematic analysis. (Braun and Clarke, 2006; Braun and Clarke, 2014) Three team members reviewed transcripts and developed codes and definitions. Then, two researchers coded each transcript independently in ATLAS.ti Version 8, and met with the Principal Investigator to compare and refine the codes and definitions, and resolve minor discrepancies. Each team member identified themes inductively, identifying analytical categories from the data. The team examined patterns within and across transcripts and grouped consistent responses with illustrative quotations. The entire team reviewed the resulting synthesis of themes.

3. Results

We examined data from a purposive sample (Table 1). More researchers and gatekeepers than advocates had direct experience with the PHD Warehouse and with using big data for research. However, most participants acknowledged living in a new information era in which

Table 1
Sample characteristics.

	Gatekeepers (n = 8)	Researchers (n = 12)	Advocates (n = 19)
Age, mean (sd)	42.9 (9.6)	42.1 (8.8)	51.2 (15.0)
Race/Ethnicity, %			
White	87.5	83.3	73.7
Asian	0.0	16.7	5.3
Latino	12.5	0.0	5.3
Black or African American	0.0	0.0	5.3
Other	0.0	0.0	10.5
Female, %	75.0	33.3	63.2
Education, %			
HS diploma/GED/ Trade Vocation	0.0	0.0	26.3
Associate's	0.0	0.0	42.1
College	12.5	0.0	15.8
Masters or professional degree	62.5	0.0	15.8
MD	0.0	58.3	0.0
PhD	12.5	41.7	0.0
Other	12.5	0.0	0.0

there are tools, expertise, and opportunities to use big data to learn about and respond to epidemics. Hereafter, we use the terms "big data" and "PHD Warehouse" interchangeably with the understanding that participants mostly reflected on the PHD Warehouse to consider the strengths and limitations of big data.

3.1. Perceived strengths

When asked to identify the potential strengths of PHD Warehouse data for the conduct of research, participants generally used positive words such as "powerful" and "invaluable," adding that the data are "comprehensive and rich" and "offer unique and tremendous opportunities" that provide an "enormous advantage" for public health research (Table 2).

3.1.1. Large sample sizes

Participants noted how the data in the PHD Warehouse offers enough statistical power to detect new and meaningful relationships. Examples focused on new information that had resulted when data were analyzed by gender, age, and other key socio-demographic characteristics. Big data has enabled researchers to challenge assumptions about the nature of opioid use disorder and to research historically understudied subgroups.

3.1.2. Cross-institutional treatment and services

Participants admired how the PHD Warehouse bridges siloed institutions, making it possible to track people over time and geographic place and as patients interact with different health and social systems. This was seen as a strength of big data because it enables a more complete and accurate picture of risk and protective factors, treatment outcomes, and the effectiveness of interventions. A related perceived strength of the boundary-spanning nature of linkage across institutions is that, by design, it brings together collaborative and multidisciplinary teams of experts. Participants highlighted that the PHD Warehouse is "…a very, very unique data set" because it creates a space for different state agencies and scientists to talk to each other, a feature perceived to be an "essential" asset for tackling research questions posed by the complexities of the opioid overdose epidemic.

3.1.3. Geographic-specific information

Participants recognized that the PHD Warehouse permits the identification of opioid overdose "hot spots" and other adverse impacts by zip code or locality. Being labelled as an "at risk" community (e.g., a neighborhood with more opioid overdoses) along with any potential community-level harms associated with that designation (e.g., on real estate, tourism, business development), was thought to be outweighed by the generation of geographic-specific information (e.g., siting of treatment centers where they are most needed) that could be used to move beyond a focus on assessing individual-level health to instead prioritize community-level health. In effect, big data was perceived to facilitate the application of a public health geospatial lens to understand the opioid overdose epidemic which, in turn, could aid more appropriate and equitable distribution of public resources.

3.1.4. "Real-world" research on complex phenomena

Some participants pointed out that the PHD Warehouse offers data that are especially suited for conducting "real world" research. A researcher said in relation to the PHD Warehouse and other sources of big data, "…one of the largest pros…is that it [big data] is more or less unaltered by humans and…you can use it as a natural experiment to get answers…[and identify] where there's need for more research…..." This participant further explained that results generated from these data may be less subject to potential manipulation and thus less vulnerable to self-selection bias and implicit researcher bias than data from randomized controlled trials or other study designs. Also perceived as a strength was the ability to have the same big data set analyzed by different teams

Table 2
Perceived strengths of using big data for research on the opioid epidemic.

Theme 1: Large sample sizes
Implications
Detect new relationships
Examine sub-groups
Generate clinically significant results
Challenge assumptions about the nature of opioid use disorder
Quotes and paraphrases
One researcher explained that with a large enough sample it was possible to determine that among adults, men are more likely to have a non-fatal overdose than women, but that among adolescents the reverse is evident, that is, it is actually girls who are more likely than boys to have a non-fatal overdose.
Another researcher commented that the most significant strength of large sample sizes is the ability to find previously unrecognized but clinically significant information that can be readily applied to patients.
A third participant, an advocate, explained how it was an "eye-opener" to learn that many fatal overdoses occur over age 50, a realization that helped her understand that opioid misuse can indeed occur among older adults.
Theme 2: Cross-institutional multidisciplinary research
Implications
Track people over time and place
Create a more complete and accurate picture of risk and protective factors, treatment outcomes, and the effectiveness of interventions
Work in collaborative and multidisciplinary teams of experts, a necessary resource for addressing the opioid epidemic
Quotes and paraphrases
One researcher said, "…big data allows us to better understand what happens to people when they leave our institution…we otherwise make huge claims around, for example, success or failures, treatments, interventions based on just whether or not they came back to our institution, which is really missing so much of what a person's experience likely is…And so, big data allows us to try to capture that and I think that's really invaluable."
Another researcher said, "…we learn a lot more from using bigger data sets, where we can see people moving within and across organizations and insurers and services…than we do from disconnected or really small but deep data sets…this is particularly important in the opioid epidemic, because cross system use is a big risk factor for problematic opioid use, and we've learned a lot of that through the use of big data."
Theme 3: Geographic-specific information
Implications
Go beyond a focus on individual-level health to instead focus on community-level health
Aid more appropriate and equitable distribution of public resources
Quotes and paraphrases
One advocate said that "…for communities to be stigmatized as [opioid] havens may not be a bad thing. Basically, it highlights the need for help, therefore, in policing or in recovery. I would think the federal government…would be able to send extra resources to help that community, instead of letting them suffer all this time, over and over, knowing they're under-policed or they don't have the resources to combat—to help the addicts."
Another advocate noted that having information on the effects of the opioid epidemic by zip code, and especially in rural areas, would be "…a wonderful thing, because government money could be assigned accordingly, rather than have all the money going to big, metropolitan areas…" funds would be allocated according to need, even to "…that little zip code that's going to get wiped out, because of…not having enough government money to fund …treatment…."
Theme 4: "Real-world" research on complex phenomena
Implications
Less vulnerable to self-selection and implicit biases, can replicate results
Encompasses social determinants and interrelationships between multiple related predictors and outcomes over time
Apply and advance innovations in methods and statistics
Quotes and paraphrases
One researcher noted that big data allows researchers to consider how sociodemographic factors may contribute to both opioid overdose and also other related health conditions, such as hepatitis C, HIV, infectious endocarditis, and

(*continued on next page*)

Table 2 (*continued*)

Theme 1: Large sample sizes

Implications
Detect new relationships
Examine sub-groups
Generate clinically significant results
Challenge assumptions about the nature of opioid use disorder
Quotes and paraphrases
One researcher explained that with a large enough sample it was possible to determine that among adults, men are more likely to have a non-fatal overdose than women, but that among adolescents the reverse is evident, that is, it is actually girls who are more likely than boys to have a non-fatal overdose.

Another researcher commented that the most significant strength of large sample sizes is the ability to find previously unrecognized but clinically significant information that can be readily applied to patients.

A third participant, an advocate, explained how it was an "eye-opener" to learn that many fatal overdoses occur over age 50, a realization that helped her understand that opioid misuse can indeed occur among older adults.

sexually transmitted infections. The participant explained that "…there are a range of overlapping epidemics that have the same or similar root causes…it may be social determinants of health…[or] the environmental or biological factors that…are in common for those outcomes to happen…So, we're better [off] not to think about these in siloes, but to think how a host of different factors and determinants could have big impacts on those intertwined and overlapping outcomes."

Another researcher said, "I think it's not just the breadth [of the data], but it's also the depth…we had a longitudinal dataset and you could look at sequencing of events…It's very much 'tracking people over time, looking at their risk after certain events take place, trying to understand individuals who might have had similar risk initially, but had some kind of intervention in between and how did you mitigate that risk?'…[and]…over time [as more data are added], this becomes a richer and richer dataset, where they can look at more and more life events and determine risk, not just for opioids, but [for] chronic disease, infectious disease, it covers the whole gamut of health conditions."

Theme 5: High impact potential

Implications
Viewed as encompassing "all of us"
Legitimize the scope and impacts of the opioid epidemic
Answer urgent research questions
Inform impactful policymaking
Quotes and paraphrases
One advocate observed, "The big pro [of big data]…is that…it adds credibility to the crisis. Because when I read a study or if I watch 20/20 [on television] and they show you maps and [overdose] increases and when you look at the sheer numbers, for me it is a good way to reach the public who feel that it doesn't apply to their neighborhood, their town. And so, when they're faced with facts, it lends credibility." This participant further explained, "I think that I still believe there's a great deal of stigma that's attached to the opioid crisis. And I think it really humanizes it and it puts it across a level playing field. It's not just 'those people' or 'that section' or 'that economic or that ethnic' group. It's everybody."

One researcher felt that "overwhelmingly" big data can be a "source for doing good." Pointing to how big data can be analyzed to answer sophisticated questions quickly, this participant said, "Part of the power of using large, deidentified data sets is that we need answers to questions now. There are epidemics. There are big issues that need quick responses and quick…well-thought-out analyses that can just be put forth."

of researchers across institutions who could attempt to replicate findings, a capacity that could potentially improve research quality.

Participants valued that the PHD Warehouse enables examination of complex phenomena, for example by incorporation of the social determinants of health and the examination of interrelationships between multiple related predictors and outcomes over time. Some participants noted that the desire to better capture these complexities created opportunities to apply and advance innovations in methods and statistics (e.g., machine learning, health informatics, predictive analytics), which was viewed as an added advantage.

3.1.5. High impact potential

A final perceived strength of the PHD Warehouse was its potential for consequential impacts. Particularly impactful was the idea that these data are not based on relatively small samples but instead are viewed as encompassing "all of us." Participants shared how the results generated from PHD Warehouse studies have been used to educate the public about the scope and impacts of the opioid overdose epidemic, answer urgent research questions, and inform impactful policymaking. The ability to conduct sophisticated analyses quickly to answer urgent research questions was perceived to be an advantage of the PHD Warehouse that differentiated it from most cohort study designs and randomized trials.

Other participants shared experiences with how findings from PHD Warehouse research had changed laws that guide clinical practice. A researcher gave the example of findings that had documented how few opioid overdose patients are initiated onto methadone or buprenorphine in emergency department settings and, when they are, patient lives are saved. The participant felt that the need-care gap that had been revealed by research had helped to pass a recent law mandating that non-fatal overdose patients treated in emergency departments be offered the option to start these medications, a significant change in practice that this participant expected would increase the availability of evidence-based treatment for opioid use disorder in emergency departments in the state and nationwide.

3.2. Perceived limitations

Participants noted several limitations of using big data from the PHD Warehouse for research (Table 3) and highlighted the value of transparency in communicating these limitations to avoid inappropriate uses.

3.2.1. Oversimplification of information

To protect participant confidentiality, some information within the PHD Warehouse is aggregated or "scrubbed" (removed altogether). These methods enhance data security, but also mask useful detail. A researcher said in relation to the PHD Warehouse, "…the restrictions that can happen in trying to protect these data, can actually make it really hard for us to learn information." Participants also observed that big data derived from administrative sources is not built for research and it lacks specific measurement of certain constructs. Race and ethnicity was commonly referenced as a construct that is defined differently across data sources or missing altogether. Unclear reliability and validity of measures made big data vulnerable to misinterpretation and, in some cases, precluded analysis of variables.

3.2.2. Challenging to access and analyze

Researchers shared how PHD Warehouse data require time to access and analyze, involving many application "hurdles." Other logistical challenges related to being permitted too little time to complete analyses and lengthy processes for review of findings by gatekeepers prior to dissemination. Researchers described having windows of time within their career, for example as post-doctoral fellows, that were not long enough to complete big data projects.

Researchers noted that PHD Warehouse data are "not user-friendly," "overwhelming" to make sense of, and required advanced statistical and data management skills. Relatively few researchers have the needed expertise and institutional infrastructure to conduct big data research. Participants worried that the volume and complexity of data mean that it can easily be misused. Participants agreed that the steep learning curve is best tackled by individuals who have the time to be trained by gatekeepers and are mentored by colleagues with relevant expertise.

3.2.3. Lag time

The PHD Warehouse data are essentially a static record of historic events. It can take several years from the occurrence of an event to when it is analyzed for research purposes. Participants were concerned that by the time big data from the PHD Warehouse are made available for

Table 3
Perceived limitations of using big data for research on the opioid epidemic.

Theme 1: Oversimplification of information

Implications
Useful detail is masked or scrubbed, which eliminates granularity and makes it hard to learn information
Lacks reliable and valid measurement of certain constructs (e.g., race, ethnicity)
Misses nuances of the patient story
Vulnerable to misinterpretation and overgeneralization of results
Quotes and paraphrases
A researcher shared how monthly data on healthcare utilization was "reduced" to "zeroes and ones" which made it impossible to know whether somebody had more than one treatment episode per month.

A gatekeeper commented that aggregated data on treatment episodes was a "major limitation" because it eliminated the "granularity" needed to pinpoint specific places and times to intervene.

A researcher said, "…we don't necessarily always have a complete understanding of what a particular code or flag or variable in these records might mean…and I think there's a potential for misuse, misinterpretation as a result….…"

A researcher said, "It's being generated for another purpose, not for research. And so there's going to just be this inherent messiness…you're using proxies or…making some assumptions…you can be upfront about those assumptions, but they're still pretty real limitations."

A researcher called for integration of quantitative and qualitative data saying, "I don't think [big] data is enough. I think we need some interviews, we need to hear the stories, we need to hear the experiences…big data answers one question, but…it needs to be…extended…that would give more of a human touch…we might be missing a lot of things by not hearing them out…."

Theme 2: Challenging to access and analyze

Implications
Length of time needed to complete big data projects is longer than what most researchers can devote
Steep learning curve, requires specialized training and mentorship
Few researchers have the needed expertise and institutional infrastructure to conduct big data research

Quotes and paraphrases
A gatekeeper said, "That's always my concern…it's a lot of data and there's a lot of caveats and…places where things could go wrong…there's a lot of nuance…." This participant went on to explain, "…and that's why we hold hands very tightly with the people working with us…[why] we don't give the data away and [instead] people really have to come here and work with us, so we can see [the work] along the way…."

A gatekeeper shared current proactive efforts to "open up" the process to make big data more accessible and "broaden the playing field" to encompass a greater diversity of researcher interests and expertise.

Theme 3: Lag time

Implications
Static record of old events
Results are out of sync with the current state of the opioid epidemic
Limited utility for the provision of healthcare in the moment
If were to create capacity for real-time data, would need to address data safeguards and ethical issues
Quotes and paraphrases
A researcher noted, "…if there really is…going to be like a five to six-year turnaround [to generate findings] for things as fast moving as…the opioid epidemic, it may not be that the data we get can meaningfully…change practice on a timescale that's helpful."

A researcher said, "Right now they have a static dataset…that hasn't been updated in three years. They're going to have closer to real-time updates coming up…with a three-month lag, as opposed to making a static data set that never changes."

A researcher said, "There is this thought that if we could deploy resources now, instead of based on five years ago data, [then] that would be a lot better."

An advocate felt that more appropriate healthcare could be provided if big data could be used to ensure that patients' history of addiction was shared across

Table 3 (*continued*)

Theme 1: Oversimplification of information

Implications
Useful detail is masked or scrubbed, which eliminates granularity and makes it hard to learn information
Lacks reliable and valid measurement of certain constructs (e.g., race, ethnicity)
Misses nuances of the patient story
Vulnerable to misinterpretation and overgeneralization of results
Quotes and paraphrases
A researcher shared how monthly data on healthcare utilization was "reduced" to "zeroes and ones" which made it impossible to know whether somebody had more than one treatment episode per month.

A gatekeeper commented that aggregated data on treatment episodes was a "major limitation" because it eliminated the "granularity" needed to pinpoint specific places and times to intervene.

A researcher said, "…we don't necessarily always have a complete understanding of what a particular code or flag or variable in these records might mean…and I think there's a potential for misuse, misinterpretation as a result….…"

A researcher said, "It's being generated for another purpose, not for research. And so there's going to just be this inherent messiness…you're using proxies or…making some assumptions…you can be upfront about those assumptions, but they're still pretty real limitations."

A researcher called for integration of quantitative and qualitative data saying, "I don't think [big] data is enough. I think we need some interviews, we need to hear the stories, we need to hear the experiences…big data answers one question, but…it needs to be…extended…that would give more of a human touch…we might be missing a lot of things by not hearing them out…."

healthcare systems. This participant explained, "I could go to the hospital in [town 1], they know I'm an addict. I'll go to the hospital in [town 2] and they don't know I'm an addict, because it's two different hospitals. I think if we're in that system…it [addiction history] should be [known] to every hospital you go to."

Another advocate, who was a military veteran, shared "I'm a Veteran and…I know I get frustrated, every time I go to…[different VA hospitals] that my records don't follow me throughout the state, and…I get so frustrated at the process…'Why aren't these computers linked? Why do I have to go through this process every time?'…I'm in recovery…and when…I use again, I'm so ashamed…that…I might not want to mention [my substance use]…[when] the doctor [says] 'How many drinks do you drink?'…you really drink 12 but you tell him 2…because you're embarrassed to say the truth. So, this [real-time big data made available across healthcare systems] would make the truth come forward and then I think that would be the benefit."

A researcher observed, "so [with real-time data], are you going to then go out and intervene and track them [patients] down? And if you are…then it's human subjects research. And then all of the safeguards…[for] human subjects research need to apply…."

Theme 4: Blind spots

Implications
Contribute to spurious or confounded results, and unjustified conclusions and policy implications
Limits ability to examine the upstream or root causes of opioid use disorder
Incomplete or biased understanding of the opioid epidemic and limited thinking on how to address it
Quotes and paraphrases

<u>Opioid and other substance use</u>
A gatekeeper said, "… if you're not in the system, we don't get your data," explaining the potential implications futher with the example that "…as there is more naloxone, people are not calling 911 as much. And so, we're undercounting non-fatal overdoses, because unless you call 911 or go to the hospital, we don't know about that overdose. And as there's a lot more spread of community naloxone, it's easy for your friend to revive you and then not go to the hospital. So…[events] that are…outside of the system…could be a big blind spot that we just don't know about."

A researcher provided an example of how one dataset censors substance use disorder diagnoses, saying "So…we have a really incomplete understanding of care for these populations..[which] does really give us spurious findings that are not right, because we're not properly accounting for it…[censored data] create some sort of really big hole in what we know about treatment, not just for substance use disorder,

(*continued on next page*)

Table 3 (*continued*)

Theme 1: Oversimplification of information

Implications
Useful detail is masked or scrubbed, which eliminates granularity and makes it hard to learn information
Lacks reliable and valid measurement of certain constructs (e.g., race, ethnicity)
Misses nuances of the patient story
Vulnerable to misinterpretation and overgeneralization of results
Quotes and paraphrases
A researcher shared how monthly data on healthcare utilization was “reduced” to “zeroes and ones” which made it impossible to know whether somebody had more than one treatment episode per month.

A gatekeeper commented that aggregated data on treatment episodes was a “major limitation” because it eliminated the “granularity” needed to pinpoint specific places and times to intervene.

A researcher said, “…we don’t necessarily always have a complete understanding of what a particular code or flag or variable in these records might mean…and I think there’s a potential for misuse, misinterpretation as a result….…”

A researcher said, “It’s being generated for another purpose, not for research. And so there’s going to just be this inherent messiness…you’re using proxies or…making some assumptions…you can be upfront about those assumptions, but they’re still pretty real limitations.”

A researcher called for integration of quantitative and qualitative data saying, “I don’t think [big] data is enough. I think we need some interviews, we need to hear the stories, we need to hear the experiences…big data answers one question, but…it needs to be…extended…that would give more of a human touch…we might be missing a lot of things by not hearing them out….”

Table 3 (*continued*)

Theme 1: Oversimplification of information

Implications
Useful detail is masked or scrubbed, which eliminates granularity and makes it hard to learn information
Lacks reliable and valid measurement of certain constructs (e.g., race, ethnicity)
Misses nuances of the patient story
Vulnerable to misinterpretation and overgeneralization of results
Quotes and paraphrases
A researcher shared how monthly data on healthcare utilization was “reduced” to “zeroes and ones” which made it impossible to know whether somebody had more than one treatment episode per month.

A gatekeeper commented that aggregated data on treatment episodes was a “major limitation” because it eliminated the “granularity” needed to pinpoint specific places and times to intervene.

A researcher said, “…we don’t necessarily always have a complete understanding of what a particular code or flag or variable in these records might mean…and I think there’s a potential for misuse, misinterpretation as a result….…”

A researcher said, “It’s being generated for another purpose, not for research. And so there’s going to just be this inherent messiness…you’re using proxies or…making some assumptions…you can be upfront about those assumptions, but they’re still pretty real limitations.”

A researcher called for integration of quantitative and qualitative data saying, “I don’t think [big] data is enough. I think we need some interviews, we need to hear the stories, we need to hear the experiences…big data answers one question, but…it needs to be…extended…that would give more of a human touch…we might be missing a lot of things by not hearing them out….”

but also for all of the co-occurring physical and mental health conditions that these people have.”

<u>Early life factors</u>
A gatekeeper said, “…sometimes the causes of addiction start very, very young. And we’re not understanding that…[we lack] information on whether they’ve experienced acute trauma as children that might help to point to some of the outcomes or draw some associations. So, we definitely have blind spots around just root causes, especially those that are…only now being better understood….”

One researcher said, “There’s like a gazillion things that happen before somebody actually develops opioid use disorder and those things affect the risk of developing opioid use disorder and then their subsequent likelihood of sustained recovery. And…almost none of these things are captured in databases. Even if we just think about the simple maternal exposures…often moms’ records are not linked with their children in many databases, or at least we don’t have that information. So, we don’t have [data on]…family unemployment or family poverty or homelessness or family incarceration…[and] it means that we sort of entirely miss a full set of factors that are important.”

Another researcher wished for big data that enabled examination of events over the life course with documentation of familial relationships and social contexts. “In a perfect world…I’d like to use sort of Danish [big data] systems, where we can see everything that has ever happened to a person, that’s been measured by the government, from the time they’re born to the time they die…We would have linked health and tax records so that we could accurately measure income. We would be able to follow people over the life course, instead of following them only for the two years they’re insured by their employer. We’d be able to see all these other factors…[like] family linkages…and social context.”

<u>Homelessness, poverty</u>
A researcher explained, “…we only get information by virtue of people’s interactions with these different service systems. So…if you’re not showing up in a particular service system, what does that actually mean? Like if you’re not showing up in the shelter records, are you not using shelter because you’re not homeless or because you’re living in an unsheltered situation or is it you’re couch surfing between places? We don’t get nuanced information about anything. We know that you interacted with a particular service system. We know something about the nature and maybe the duration of that interaction, but we don’t really know much else.”

<u>Contexts</u>
A gatekeeper recalled how patients who received addiction treatment in facilities that had “popped up” in another state due to the opioid crisis were facilities that had “…no system of accountability to report outcomes…and so, as a result…a lot of people started to see that their family members were mistreated. Some of them died, some of them were relapsing like within days or hours of being released from the facilities. And that’s one of the blind spots…we just don’t know…‘where people are accessing drugs or accessing treatment,’ [and] ‘what are the outcomes….’”

<u>Social support, patient perceptions</u>
A researcher said, “We know that you interacted with a particular service system…but we don’t really know much else. We don’t know how you felt about it…we don’t have anything about your perception…we might show that you have an in-patient hospitalization, but that doesn’t tell us anything about how you rate your own health…or how you think about the quality or adequacy of your housing…or your ability to access employment…or transportation. So…people’s perceptions of things, we don’t know.”

<u>Privilege, discrimination, root causes of health inequity</u>
A gatekeeper said, “…often people with the most privilege are not actually represented in the data at all…for instance, if you go to private treatment centers, you’re not in our state-funded treatment system. We don’t actually have your information…[instead] we end up collecting the most information on the most marginalized people. And so…that becomes problematic from a…justice standpoint [and also because]…it can…create trouble with our analyses, when you don’t have a group of the population and we don’t know really what’s going on with them. And I think it’s very easy…[to forget that]…not every data set that becomes part of a system actually is fully representative of the state.”

A gatekeeper said, “my main concern of using big data is how we are using it…we are trying to understand the output [i.e., the opioid epidemic] that comes from an inequitable system…And so, we’re trying to figure out what are those determinants that created the condition, but we are not looking at the original conditions that enabled the creation of that output….we are getting at the rear end of the analysis…we create target populations and we then create interventions to… help people, we are…lacking that angle of going on the offensive and being preventative about how do you avoid people from getting there in the first place.”

A researcher said, “We also don’t get…any sense of people’s everyday lived experience in the world. ‘To what extent have you experienced discrimination in housing or labor markets, because of one characteristic or another?’ By using other proxies in the data, like race, you can kind of get some sense of how that might be at play, but you just don’t get…what might be structural factors acting on people….”

A gatekeeper said, “We need to focus on the policies that exist that have created the situation in which people are prone to become addicted to opioids.”

research, the opioid overdose epidemic has changed. To make findings timely, participants suggested that the PHD Warehouse be updated more frequently.

Advocates focused on the limitations of PHD Warehouse data in relation to its utility for healthcare. Some suggested that these big data be changed to incorporate "real-time" information to inform better healthcare decision-making. Participants were concerned, however, that "real-time" big data might be used to further stigmatize or punish individuals with opioid use disorder. An advocate who wanted his health information shared across VA hospitals also said that big data should not be used by the criminal justice system to target individuals with opioid use disorder or to deny them access to social welfare benefits. Instead, he said that big data "…is supposed to be about saving our lives…." Some participants noted that real-time updates would make the data more timely, but would also transform the data into identifiable human subjects research in need of additional safeguards.

3.2.4. *Blind spots*

Participants noted several factors that influence opioid use but are omitted from the PHD Warehouse. Blind spots were perceived to contribute to inaccurate or confounded results, unjustified conclusions and implications for policy and practice, and an inability to examine the antecedent or root causes of opioid use disorder. Participants shared that consequences of blind spots are an incomplete or biased understanding of the opioid overdose epidemic and limited thinking in how to address it.

Opioid and other substance use. The "hugest blind spot" is limited and indirect measurement of opioid and other substance use. The PHD Warehouse does not track substance use unless it comes to the attention of organizations that document it and contribute records to the warehouse. For example, participants shared how big data can be used to detect prescriptions and adverse events like mortality but it does not include measurement of illicit substance use. Participants also noted how blind spots can occur when substance use information is intentionally removed from source data.

Early life factors. Participants identified how experiences of childhood adversity, family history, and early education impact opioid use but are usually not captured in the PHD Warehouse. These omissions make it challenging to understand who develops substance use disorders and how they recover from it. Participants also noted how these data often do not offer the ability to link data on children with that of their parents or other family members, which makes it hard to understand the intergenerational effects of substance use disorders. Others wished for data that enabled examination of events over the life course with documentation of familial relationships and social contexts.

Homelessness, poverty. Researchers described the imprecise indicators of homelessness, which has made it possible to document it as a key predictor of mortality among people with opioid use disorder, but do not permit exploration of the directionality of opioid use and homelessness, or to understand the impact of homelesseness on entering housing or treatment outcomes. Participants also highlighted how poverty, as indicated by socioeconomic factors (income, employment history, education level, food insecurity), also function as important social determinants of opioid use, but are omitted from or captured poorly by the PHD Warehouse.

Contexts. Participants identified certain contexts that are known to shape opioid use but are not represented in the PHD Warehouse. For example, in the initial version of the PHD Warehouse, limited to no information was included on healthcare utilization within Veterans Affairs settings. Also omitted were details on interactions with the criminal justice system including jail and prison events, experiences with community supervision (e.g., probation, parole), and court records. Finally, participants highlighted how these data are limited to events that occur within a certain state.

Social support, patient perceptions. Participants expressed a need to include in the PHD Warehouse more indicators of social support, for example social networks, social services utilization, and self-reported feelings of social isolation or connectedness. Also needed are measures of patient perceptions of health status, healthcare, and other factors that influence health.

Privilege, discrimination, root causes of health inequity. Records on receipt of privately funded treatment for substance use is not included in the PHD Warehouse, meaning that "individuals with the most privilege" are usually excluded, a limitation that restricts the representativeness of results and curtails opportunities to conduct research using a health equity or social justice lens. Also, these data do not measure experiences of individual-level or structural discrimination. Finally, the nature of these data often causes research to focus on the person with opioid use disorder when instead focus should be shifted to understand the root causes of the opioid overdose epidemic.

4. Discussion

4.1. *Findings and implications*

Massachusetts is engaged in extraordinary use of individually-linked administrative big data on opioid use as routinely provided by health, criminal justice, social services, and vital statistics systems. We documented stakeholder perceptions of the strengths and limitations of using these administrative big data for research purposes. Participants highlighted the value of learning about the strengths to capitalize on advantages and of communicating the limitations to avoid inappropriate uses. Results can inform future research and aid other states and locales that seek to create similar resources.

Participants viewed the PHD Warehouse as a valuable research resource. A key perceived strength is the ability to use these data to detect new relationships, particularly among understudied sub-groups, that yield clinically significant results and challenge assumptions about the opioid overdose epidemic. Another advantage is that these data span institutional boundaries. Thus, by design, the PHD Warehouse creates opportunities to broaden understanding of risk and protective factors, treatment outcomes, and the effectiveness of interventions. Additionally, these data are suited to the creation of geographic-specific information and a focus on community-level health. The ability to conduct "real-world" research with these data was perceived to be less vulnerable to self-selection bias and implicit researcher bias. The PHD Warehouse contains the elements needed to examine complex phenomena such as the social determinants of health and the interrelationships between multiple related predictors and outcomes as they unfold and interact over time. Participants valued that research with these data is an opportunity to apply and advance innovations in methods and statistics. Finally, the big data offered by the PHD Warehouse are viewed as encompassing "all of us." Thus, it can support impactful findings that legitimize the scope and impacts of the opioid overdose epidemic, answer urgent questions, and influence public health policy and practice.

These perceived strengths were corroborated by published findings from PHD Warehouse data and other similar efforts. (Hallvik et al., 2021; Weiner et al., 2022) For example, studies using PHD Warehouse data have documented: (1) prevalence of opioid use disorder and variation in prevalence rates by certain factors such as time-period, characteristics of the population, and geographic location (Barocas et al., 2018; Schiff et al., 2018); (2) potentially inappropriate prescribing practices (Rose et al., 2018; Rose et al., 2011; Stopka et al., 2019); (3) availability, access, utilization, and outcomes of treatment with medications for opioid use disorder (MOUD) (Larochelle et al., 2018; Larochelle et al., 2019); and (4) vulnerable populations at higher risk for underutilization of MOUD or adverse health and social outcomes, such as military Veterans, (Jasuja et al., 2018) adolescents, (Chatterjee et al., 2019) and pregnant women. (Schiff et al., 2018) These and other findings have been used at the state level to develop an action plan of initiatives and enact changes to budgetary and legislative items to address the opioid overdose epidemic (Governor's Working Group on Opioid Addiction, 2017).

Our findings also document how oversimplified information and imprecise measures were key perceived limitations of the PHD Warehouse. Similar to other reports of the complex processes needed access administrative big data on substance use disorders, (Geissler et al., 2021) participants also noted challenges of data access and analysis, which underscored the need to plan for enough time, specialized expertise, mentorship, and institutional support to conduct this type of research. Also problematic are the substantial lag times from event to analysis and that data are made up of static historic records, a challenge that has been documented elsewhere. (Geissler et al., 2021) Advocates in particular described the drawbacks of static historic records and expressed interest in having these data made available in real time during clinical encounters. Such uses of PHD Warehouse data are explicitly prohibited to protect privacy and confidentiality. Results point to one of several ethical issues, explored in detail elsewhere, (Evans et al., 2020) regarding the need to identify how to use these data in ways that are valued by those who are most impacted by the opioid overdose epidemic.

Participants also identified several blind spots. Especially problematic is the limited and indirect measurement of opioid and other substance use and the missing or poorly defined operationalization of race and ethnicity. Participants explained how these data gaps, as documented in other studies, (Austin et al., 2018) can cause spurious or confounded results, mask upstream or root causes, and contribute to incomplete or biased understanding of the opioid overdose epidemic and how to address it. Importantly, the PHD Warehouse is not a defined dataset. Instead, it is updated at regular intervals to encompass data on more recent events. It is also being expanded upon now to include more data from the justice system and other data sources. Looking forward, the PHD warehouse would be strengthened by continued attention to blind spots and ways to address them. More broadly, the PHD Warehouse can be conceptualized as a public health innovation (Evans et al., 2019) that has the potential to generate information for health equity and optimal health. A critical implication is that as the PHD Warehouse moves through the different stages of development, it generates different capacities and concerns. This means, for example, that as the PHD Warehouse is further developed, it requires different resources, skills, and other inputs to operate, it is characterized by different strengths and limitations, and there is variation in its outcomes and impacts. Accordingly, some issues that are featured in this paper may be the result of the rapid development of the PHD Warehouse, and best understood as "growing pains" rather than as limitations of big data as a whole, and others could be mitigated for example through training and scientific literacy promotion for the public.

Finally, the Massachusetts PHD Warehouse was established in response to the opioid overdose epidemic, but it was intentionally designed for studying new and emergent public health issues such as maternal and child health and COVID-19. (Bharel et al., 2020) We expect that results are likely to be relevant to these other research areas. Thus, findings may generalize to these and other topics where big data are used to advance public health. Additionally, some of the lessons learned about the strengths of the PHD Warehouse are specific to this data resource, which offers future directions for other administrative big data sources. In particular, unprecedented research opportunities are created by that fact that the PHD Warehouse encompasses data on an entire population, not only on individuals with a certain health condition or treated by one healthcare delivery system, and these data can be used to track individuals from birth to death, enabling understanding of health over the life course. Another strength that is unique to the PHD Warehouse is that it is managed by the state department of public health, and not by a university or healthcare delivery system, which has the capacity to sustain these data as a research resource and to ensure that these data are used for the public good.

4.2. Limitations and strengths.

Qualitative research often relies on small sample sizes, which does not support broad generalizations but does provide in-depth information. (Creswell and Creswell, 2018; Curtis et al., 2000) Researchers and gatekeepers had more direct experience than advocates with the PHD Warehouse data collection, management, and analysis, which likely contributed to variation in responses to some prompts. Findings apply to static cross-sectoral administrative big data that is created by and for public health. Participants were prompted to discuss the PHD Warehouse, but it was not always possible to disentangle when participants were referring to the PHD Warehouse specifically or big data in general. We sought perspectives from different stakeholder groups and regarding big data on opioid use, thereby investigating understudied topics. The study was conducted in Massachusetts, which is leading the creation of big data for population health. In contrast to most research on big data, which has typically used a quantitative design, we used qualitative methods to explore experiences, thereby attaining knowledge of the factors that shape perspectives.

4.3. Conclusion

Using big data to conduct research on the opioid overdose epidemic offers advantages but is also subject to limitations which, if unrecognized, may undermine its utility. Findings can inform future research efforts and aid other states that are working to assemble administrative big data to guide public health practice and policy.

Ethics approval and consent to participate

All procedures were approved by the University of Massachusetts Institutional Review Board. Informed written consent was obtained from each participant prior to participation. All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication.

Not applicable.

Availability of data and materials

All data supporting the findings of this study are contained within the manuscript. Additional information on the study will be shared by the corresponding author on reasonable request.

Funding

The Greenwall Foundation funded this project, without any restriction and played no role in the conception, design, and reporting of findings. Dr. Evans is also supported by the National Institute on Drug Abuse (NIDA) UG3 DA0044830-02S1 and 1UG1DA050067-01, and the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Substance Abuse Treatment (CSAT) Grant No. 1H79T1081387-01.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors thank the study participants for taking part in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pmedr.2022.101847.

References

Anglin, M.D., Nosyk, B., Jaffe, A., Urada, D., Evans, E., 2013. Offender diversion into substance use disorder treatment: the economic impact of California's Proposition 36. Am. J. Public Health 103 (6), 1096–1102.

Austin, A.M., Bynum, J.P., Maust, D.T., Gottlieb, D.J., Meara, E., 2018. Long-term implications of a short-term policy: Redacting substance abuse data. Health Aff. 37 (6), 975–979.

Barocas JA, White LF, Wang J, Walley AY, LaRochelle MR, Bernson D, Land T, Morgan JR, Samet JH, Linas BP. 2011-2015. Estimated prevalence of opioid use disorder in Massachusetts, A Capture-Recapture Analysis. Am. J. Public Health. 2018 Dec;108 (12):1675-1681.

Bharel, M., Bernson, D., Averbach, A., 2020. Using data to guide action in response to the public health crisis of opioid overdoses. NEJM Catalyst. 1 (5) https://doi.org/10.1056/CAT.19.1118.

Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006; doi.org/10.1191/14780 88706 qp063 oa.

Braun, V., Clarke, V., 2014. What can "thematic analysis" offer health and wellbeing researchers? Int. J. Qual. Stud. Health Well-Being. 9 (1), 26152.

Case, A., Deaton, A., 2015. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. PNAS. https://doi.org/10.1073/pnas.1518393112.

Chatterjee, A., Larochelle, M.R., Xuan, Z., Wang, N., Bernson, D., Silverstein, M., Hadland, S.E., Land, T., Samet, J.H., Walley, A.Y., Bagley, S.M., 2019 Jan. Non-fatal opioid-related overdoses among adolescents in Massachusetts 2012–2014. Drug Alcohol Depend. 1 (194), 28–31.

Creswell, J.W., Creswell, J.D., 2018. Research design: qualitative, quantitative, and mixed methods approaches. Sage Publications, Thousand Oaks.

Curtis, S., Gesler, W., Smith, G., Washburn, S., 2000. Approaches to sampling and case selection in qualitative research: examples in the geography of health. Soc. Sci. Med. 50 (7-8), 1001–1014.

Ettner, S.L., Huang, D., Evans, E., Rose Ash, D., Hardy, M., Jourabchi, M., Hser, Y.-I., 2006. Benefit-cost in the California Treatment Outcome Project: Does substance abuse treatment "pay for itself?". Health Serv. Res. 41 (1), 192–213.

Evans, E., Delorme, E., Harrington, C., 2019. The Massachusetts Public Health Data (PHD) Warehouse: Conceptual Framework for Action. Report submitted to the Massachusetts Department of Public Health (MDPH).

Evans, E., Delorme, E., Cyr, K., Goldstein, D., 2020. A qualitative study of big data and the opioid epidemic: Recommendations for data governance. BMC Medical Ethics. 48, 718–734.

Evans E, Delorme E, Harrington C. The Massachusetts Public Health Data (PHD) Warehouse: History, Current Operation, and Impacts. Report submitted to the Massachusetts Department of Public Health (MDPH), 2019.

Evans, E., Grella, C., Murphy, D., Hser, Y.I., 2010. Using administrative data for longitudinal substance abuse research. J. Behav. Health Serv. Res. https://doi.org/10.1007/s11414-008-9125-3.

Geissler, K.H., Evans, E., Johnson, J.K., Whitehill, J., 2021. A scoping review of data sources for the conduct of policy-relevant substance use research. Public Health Rep. https://doi.org/10.1177/00333549211038323.

Governor's Working Group on Opioid Addiction. Action items update November 14, 2017. Commonwealth of Massachusetts. https://www.mass.gov/lists/governors-opioid-addiction-working-group. 2017.

Hallvik, S.E., Dameshghi, N., El Ibrahimi, S., Hendricks, M.A., Hildebran, C., Bishop, C.J., Weiner, S.G., 2021. Linkage of public health and all payer claims data for population-level opioid research. Pharmacoepidemiol. Drug Saf. 30 (7), 927–933.

Hsing, A.W., Ioannidis, J.P.A., 2015. Nationwide population science: Lessons from the Taiwan national health insurance research database. JAMA Intern. Med. 175 (9), 1527.

Jasuja GK, Ameli O, Miller DR, Land T, Bernson D, Rose AJ, Berlowitz DR, Smelson DA. 2018. Overdose risk for veterans receiving opioids from multiple sources. Am. J. Manag. Care. Nov;24(11):536-540.

Kass, N.E., 2001. An ethics framework for public health. Am. J. Public Health 91 (11), 1776–1782.

KPMG Government Institute. The opioid epidemic: data driven case studies. 2017. Available from: https://institutes.kpmg.us/content/dam/institutes/en/government/pdfs/2017/opioid-data-case-studies.pdf. Accessed 13 April 2021.

Krebs, E., Enns, B., Evans, E., Urada, D., Anglin, M.D., Rawson, R.A., Hser, Y.-I., Nosyk, B., 2018. Cost-effectiveness of publicly funded treatment of opioid use disorder in California. Ann. Intern. Med. 168 (1), 10.

Larochelle MR, Bernson D, Land T, Stopka TJ, Wang N, Xuan Z, Bagley SM, Liebschutz JM, Walley AY. 2018. Medication for opioid use disorder after nonfatal opioid overdose and association with mortality: A Cohort Study. Ann. Intern. Med. Aug 7; 169(3):137-145.

Larochelle, M.R., Stopka, T.J., Xuan, Z., Liebschutz, J.M., Walley, A.Y., 2019 Mar 19. Medication for opioid use disorder after nonfatal opioid overdose and mortality. Ann. Intern. Med. 170 (6), 430–431.

Lee, H., Lee, J.R., Jung, H., Lee, J.Y., 2021. Power of universal health coverage in the era of COVID-19: A nationwide observational study. Lancet Reg. Health West Pac. https://doi.org/10.1016/j.lanwpc.2020.100088.

Massachusetts Department of Public Health. An assessment of opioid-related deaths in Massachusetts (2013-2014). Massachusetts: 2016 Sep. Available from: https://www.mass.gov/doc/legislative-report-chapter-55-opioid-overdose-study-september-2016/download. Accessed 13 April 2021.

Massachusetts Department of Public Health. An assessment of fatal and non-fatal opioid overdoses in Massachusetts (2011-2015). Massachusetts: 2017 Aug. Available from: https://www.mass.gov/doc/legislative-report-chapter-55-opioid-overdose-study-august-2017/download. Accessed 13 April 2021.

Murphy SL, Xu JQ, Kochanek KD, Arias E. Mortality in the United States, 2017. Hyattsville, MD. National Center for Health Statistics; 2018. 8 p. Report No.: 328.

Rose AJ, McBain R, Schuler MS, LaRochelle MR, Ganz DA, Kilambi V, Stein BD, Bernson D, Chui KKH, Land T, Walley AY, Stopka TJ. 2019. Effect of age on opioid prescribing, overdose, and mortality in Massachusetts, 2011 to 2015. J. Am. Geriatr. Soc. Jan;67(1):128-132.

Rose AJ, Bernson D, Chui KKH, Land T, Walley AY, LaRochelle MR, Stein BD, Stopka TJ. 2018. Potentially inappropriate opioid prescribing, overdose, and mortality in Massachusetts, 2011-2015. J. Gen. Intern. Med. Sep;33(9):1512-1519.

Saloner B, Bachhuber M, Barry CL, Krawczyk N, Pasha O, Sen AP, et al. A blueprint for transforming opioid use disorder treatment in Delaware. Delaware Department of Health and Social Services: John Hopkins Bloomberg School of Public Health; 2018. Available from: https://dhss.delaware.gov/dhss/files/johnshopkinsrep.pdf. Accessed 13 April 2021.

Schiff DM, Nielsen T, Terplan M, Hood M, Bernson D, Diop H, Bharel M, Wilens TE, LaRochelle M, Walley AY, Land T. 2018. Fatal and nonfatal overdose among pregnant and postpartum women in Massachusetts. Obstet Gynecol. Aug;132(2): 466-474.

Scholl, L., Seth, P., Kariisa, M., Wilson, N., Baldwin, G., 2018. Drug and Opioid-Involved Overdose Deaths — United States, 2013–2017. MMWR Morb. Mortal. Wkly Rep. 67 (5152).

Smart R, Kase CA, Meyer A, Stein BD. Data sources and data-linking strategies to support research to address the opioid crisis: final report. US Department of Health and Human Services: RAND Corp; 2018. Report no.: EP-67716. Available from: https://www.rand.org/pubs/external_publications/EP67716.html. Accessed 13 April 2021.

Smart, R., Kase, C.A., Taylor, E.A., Lumsden, S., Smith, S.R., Stein, B.D., 2019. Strengths and weaknesses of existing data sources to support research to address the opioids crisis. Prev. Med. Rep. 6 (17), 101015.

Stopka TJ, Amaravadi H, Kaplan AR, Hoh R, Bernson D, Chui KKH, Land T, Walley AY, LaRochelle MR, Rose AJ. 2019. Opioid overdose deaths and potentially inappropriate opioid prescribing practices (PIP): a spatial epidemiological study. Int. J. Drug Policy. June; 68: 37-46.

Thygesen, L.C., Daasnes, C., Thaulow, I., Brønnum-Hansen, H., 2011. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. Scand J Public Health. 39 (7_suppl), 12–16.

Weiner SG, El Ibrahimi S, Hendricks MA, Hallvik SE, Hildebran C, Fischer MA, Weiss RD, Boyer EW, Kreiner PW, Wright DA, Flores DP, Ritter GA. 2022. Factors associated with opioid overdose after an initial opioid prescription. *JAMA Netw Open*. 4;5(1): e2145691.

Weissman, M.M., 2020. Big data begins in psychiatry. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2020.0954.

Winickoff, D.E., 2006. Genome and nation. Iceland's health sector database and its legacy. Innovations. 1 (2), 80–105.