

Visual Analytics for Data-Driven Understanding of the Substance Use Disorder Epidemic

INQUIRY: The Journal of Health Care Organization, Provision, and Financing
Volume 61: 1–13
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00469580241227020
journals.sagepub.com/home/inq



Danial Qaurooni, PhD¹ , Bruce W. Herr II, BSc¹ , Sarah Renee Zappone, MPH¹, Klaudia Wojciechowska, BA², Katy Börner, PhD¹, and Titus Schleyer, DMD & PhD¹

Abstract

The substance use disorder epidemic has emerged as a serious public health crisis, presenting complex challenges. Visual analytics offers a unique approach to address this complexity and facilitate effective interventions. This paper details the development of an innovative visual analytics dashboard, aimed at enhancing our understanding of the substance use disorder epidemic. By employing record linkage techniques, we integrate diverse data sources to provide a comprehensive view of the epidemic. Adherence to responsive, open, and user-centered design principles ensures the dashboard's usefulness and usability. Our approach to data and design encourages collaboration among various stakeholders, including researchers, politicians, and healthcare practitioners. Through illustrative outputs, we demonstrate how the dashboard can deepen our understanding of the epidemic, support intervention strategies, and evaluate the effectiveness of implemented measures. The paper concludes with a discussion of the dashboard's use cases and limitations.

Keywords

substance use disorder, epidemic, drug overdose, substance use detection, risk factors, mortality, public health, dashboard, visual analytics

What do we already know about this topic?

The substance epidemic is a serious public health crisis, and visual analytics can play a unique role in addressing this complex issue.

How does your research contribute to the field?

By using linked data and following the principles of good design, we develop a novel visual analytics dashboard that fosters collaboration among various stakeholders of this complex problem.

What are your research's implications towards theory, practice, or policy?

The dashboard can enhance understanding, aid intervention strategies, and assess the effectiveness of measures in combating the substance use disorder epidemic. This approach can be extended to other public health challenges, informing evidence-based decision-making and policy interventions.

Introduction

The drug overdose epidemic^{1,2} has cost the lives of over 700 000 people in the US, including 14 500 people in Indiana, over the last 10 years.¹ Effective solutions require close collaboration between experts that address the psychological, medical, and socioeconomic dimensions of the problem. Mental health and addiction, the overprescription of opioids, the availability of illicit opioids such as heroin and fentanyl, poverty, and unemployment are all contributing factors to the

¹Indiana University Bloomington, Bloomington, IN, USA

²Indiana University School of Medicine, Indianapolis, IN, USA

Received 30 June 2023; revised 15 December 2023; revised manuscript accepted 2 January 2024

Corresponding Author:

Danial Qaurooni, Cyberinfrastructure for Network Science Center, Luddy School of Informatics, Computing, and Engineering, Indiana University Bloomington, 700 N Woodlawn Ave, Bloomington, IN 47408, USA.

Email: dequeue@iu.edu



ongoing rise in overdose deaths,³⁻⁵ which has been exacerbated by the COVID-19 pandemic.⁶ Data analyses and visualizations can make unique contributions to the identification, communication, and implementation of effective interventions, ranging from prevention to medical treatment to long-term support.⁷ They can help discern unexpected demographic patterns, identify high-risk populations and areas, track the effectiveness of interventions, and orient research and policy.

This paper introduces a data visualization and analysis dashboard that has been designed to help decision-makers and stakeholders understand and address the drug overdose epidemic. The rest of the Introduction summarizes the challenges of data visualization in this area and reviews several large scale data analysis systems. This overview motivates our dashboard design, which is described in the Methods section. This is followed by a presentation of some illustrative visualizations in Results. The Discussion and Limitations section examines the outcomes from a critical distance, highlighting its strengths as well as shortcomings. The paper concludes with a summary of our work.

Socio-Technical Complexities of Visual Analytics Approaches

Biomedical informatics, often perceived as purely technical, is fundamentally a sociotechnical discipline.^{8,9} Its hybrid nature underscores the intricate interplay between technology and human factors in healthcare, which are often complex and dynamic. Such factors bear on the design of visual analytics systems in 2 main ways. First, the data such systems collect, aggregate, link, analyze, and display are imbued with human nuance and complexity. Second, to be usable and useful, their development has to take users into account at every step of the design cycle. In this section, we expand on the first set of factors. We focus on the design challenges in the Methods section.

Limitations related to data sources pose a significant challenge to the design of visual analytics systems. The quality, quantity, comprehensiveness, and timeliness of the data shapes downstream processes of design and development in subtle ways.⁷ For instance, one of the main sources of drug overdose fatality data are death certificates. Previous research has shown that as much as 25% of U.S. overdose deaths specified no drugs on the death certificate, resulting in the undercounting and underreporting of various drug classes.¹⁰ Much of this is due to state-by-state variations in death certification methods. Decentralized or hybrid medical examiner systems are less likely to report drugs on death certificates compared to centralized systems due to the fact that states vary in their approaches to certification. Coroners are typically elected individuals who do not need to hold a medical degree while medical examiners are forensic pathologists appointed by a government agency. Another contributing factor to the poor quality and comprehensiveness of the data stems from the prohibitive cost of post-mortem toxicology

testing. Many national labs, such as NMS Labs and Axis Forensic Toxicology, offer both “basic” and “expanded” toxicology panels, which can cost several hundred dollars per deceased person—often unaffordable given the limited budgets of many county coroners. This leads to gaps in death certificate completion and ultimately affects the quality and comprehensiveness of drug overdose data. Death certificates can also take several months to fully process, as they have to be initially completed by coroners or medical examiners, approved by a state’s Vital Records Agency, and analyzed by the National Center for Health Statistics before being published. As a result, the most current national drug overdose data released by the Centers for Disease Control and Prevention is almost always 6 months behind. Other data sources originate from electronic health records (EHRs), which are frequently real-time and contain information about health behaviors or laboratory test results that could indicate opioid misuse. However, EHRs still face quality problems and have limited generalizability due to biases inherent in particular healthcare recording processes.^{7,11}

A second set of challenges concerns the typically siloed nature of data collection. Data relevant to the substance use disorder epidemic is collected in many settings, such as healthcare events (eg, emergency department visits and inpatient stays); prescriptions, procedures, and interventions in healthcare institutions; emergency department runs through emergency medical services; arrests and incarcerations through the justice system; and information on housing, counseling, and needle exchanges through social service organizations. Although these sources of data describe a common phenomenon—that is, the substance use disorder epidemic and the people touched by it—they are rarely, if ever, linked to each other. Record linkage can help elucidate interactions with healthcare, social services, and criminal justice systems and related outcomes as they arise and influence each other over time.¹² For example, linking individual overdose rescue data with medical records can provide new insights about risk factors for overdose as well as more systematic information on the consequences of nonfatal overdoses.⁷ Despite such advantages, the existing infrastructure and protocols pose sociotechnical challenges for data linkage. On the technical side, different data sources often use different formats, structures, data types, and data scopes, which can make it difficult, even using advanced algorithms, to link them together.¹³ Beyond technical issues, data governance poses privacy and security concerns, ethical and legal implications, and data ownership and access issues. Although “big data” is anonymized and bound by other safeguards designed to prevent individual-level harms, these data could still be used in decision-making processes for which they were not intended. Examples include denying health insurance claims, the use of social welfare programs, or increasing criminal justice surveillance, prosecution, and incarceration.¹⁴ In addition, the anonymized individuals who are being examined in these data sets are often unaware

of their existence and are therefore excluded from decisions about their use. The result is that the individuals who often contribute to these data systems do not stand to benefit from them in a direct way. Other ethical considerations include the exacerbation of community-level health and social inequities. For instance, geographical hotspot maps of opioid overdoses draw attention to communities that are at particular risk of suffering from the epidemic—attention that could have negative economic impacts.

Outside of ethical, legal, and privacy considerations, there are also obstacles to data ownership and access. In most states, drug overdose data is housed within state health departments. Despite recent efforts to make data more accessible to the public, especially through public-facing data dashboards, challenges persist. The requirement that data be “cleaned up” and examined by state agencies prior to public consumption limits access to data. Acquisition of more comprehensive and raw data is often possible but faces a variety of practical obstacles and delays. A further challenge to data access is the fragmentation of data across multiple state agencies. In Indiana, for example, the Indiana Department of Health and the Indiana Management Performance Hubⁱⁱ house their data separately, creating obstacles in joint decision-making around the data.

Data-driven healthcare is heavily reliant on analytical methods, which introduces a third set of sociotechnical complications. Researchers need to understand what insights the data makes possible—a task that involves interpreting data variables and values. Despite being able to collect data, many health systems, public health departments, and other entities lack the infrastructure and resources to take this interpretation beyond simple descriptive reports. Such reports are limited in their ability to provide insight, as they often employ analyses that solely focus on the association between 2 variables (eg, time and overdose death, or geography and substance) and therefore fall short of addressing the intricacies of complex problems such as the substance use disorder epidemic. The challenge in such scenarios is to relate multiple variables and enable inferences that may incorporate an array of more advanced techniques such as multivariate analysis,¹⁵ machine learning,^{16,17} and dynamic visualizations.¹⁸ Only when this challenge is met can we expect data-driven analysis to provide a reasonably comprehensive and detailed understanding of the problem. Effective visual analytics requires overcoming such sociotechnical challenges in order to assemble high-quality data sources linked together and analyzed using advanced statistical, machine learning, and visualization techniques.

Review of Large-Scale Data Analysis Systems

Recently, several groups have attempted to address one or more of the discussed challenges by creating large-scale data analysis systems that link data sets from several sources. For

instance, the Massachusetts’ Public Health Data (PHD) Warehouse includes individual-level linked administrative data from more than 20 sources on all Massachusetts residents aged 11 and older with public or private health insurance.¹⁹ 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.²⁰ Findings generated thus far have informed public health efforts, resource allocation, community outreach, and intervention planning.¹⁹ These include new legislation that promotes responsible opioid prescribing, expands access to treatment, and calibrates anti-opioid enforcement measures. Another example is work done in the Kaiser system, which has brought together clinical, dispensing, and socioeconomic data sources to examine opioid-related outcomes across patients in multiple market segments, though the data are limited to Kaiser membership.²¹ Finally, in 2017, Governor Eric Holcomb of Indiana established the Management Performance Hub, an agency acting as the hub for data collection, analysis, and exchange for the Indiana state government. The goal of the agency was to implement data-driven policies within state government by creating a focal point for state-level data and providing technical assistance in partnership with state agencies. Another important area of work involves the effort to make linked data accessible to end users. Over the last several years, many state health departments have created “data dashboards” to better communicate substance use-related data to the public. In order to better gauge the gaps and weaknesses that currently exist in data dashboards, we reviewed substance use-focused dashboards from 43 different states (omitting instances where dashboards were either absent or under construction; see Appendix 1). The results show significant variability in 4 areas: the design of the dashboards, the data elements included in the dashboards, the programs used to build the dashboards (eg, Tableau vs Microsoft Power BI), and finally the degree of interactivity and manipulability by the end user. The list below notes several gaps among the dashboards examined:

- None include linked data that followed an individual over the course of their substance use disorder until death;
- very few highlight geographical patterns outside of general county data;
- many lack demographic information such as age and sex;
- many limit their data to overdose deaths, emergency department visits, hospitalizations, and prescription drug monitoring data; and
- interactivity remains unavailable or limited, leaving the end user unable to filter data, include or exclude variables, or manipulate the visualizations.

Table 1. Overview of Datasets.

Domain	Drug overdose deaths	Healthcare utilization	Emergency medical services	Justice involvement
Number and type of records	2331 people	~750 000 healthcare data points	2416 IEMS first responder runs	5515 jail data points
Unique individuals	2331	1751	612	496
Content	<ul style="list-style-type: none"> • Demographics, education, and employment • Date and address of death • Toxicology lab results • Selected medical diagnoses (mental health, substance use disorder, etc.) • Selected filled prescriptions, healthcare encounters, and procedures 	<ul style="list-style-type: none"> • Demographics • Healthcare encounters • Referrals • Diagnoses • Medications 	<ul style="list-style-type: none"> • Demographics • location of service • Chief complaint • Mechanism of injury 	<ul style="list-style-type: none"> • Demographics • Booking and release dates • Description of the offense • Jail admission information
Time period	2010-2018	2005-2018	2011-2018	2014-2018
Data provider	MCCO	RI/IHIE	IEMS	MCSO

Note. MCCO=Marion County Public Health Department; RI=Regenstrief Institute; IHIE=Indiana Health Information Exchange; IEMS=Indianapolis Emergency Medical Services; MCSO=Marion County Sheriff's Office.

These limitations suggest that existing dashboards are inadequate in addressing the needs of data end users, such as public health system leaders, policymakers, public, nonprofit, and prepaid healthcare systems, and other systems, such as the justice system.⁷ Elaborately created data visualizations are not useful unless they can be used to develop and support practical decisions and guide healthcare, public health, and law enforcement actions.

The above review highlights the sociotechnical complexities involved in collecting, linking, analyzing, and visualizing data in support of data-driven approaches. The work described below seeks to tackle the diverse challenges identified by building information resources that can serve as effective tools in combating the substance use disorder epidemic. The end goal of our project, Advanced Analytics for Indiana University's Addictions Grand Challenge, is to create visualizations and analytics to help viewers collaborate with each other and gain relevant insights from a rich set of linked data about deaths resulting from substance use. We have divided this goal into 2 objectives:

1. prepare, standardize and link several data sets to better understand trends and dynamics of the substance use disorder epidemic in Marion County, Indiana; and
2. develop visual analytics by following the principles of good design to make the dashboard usable and useful for a variety of stakeholders.

In the following section, we describe our methodology, introduce examples from the resulting visualization dashboard, discuss the possibilities and limitations of our work, and end with concluding remarks.

Methods

Record Linkage and Refinement

Our team collaborated with 4 organizations to obtain 4 datasets that had not previously been linked to understand the substance use disorder epidemic. The organizations were the Marion County Coroner's Office (MCCO), Regenstrief Institute (RI)/Indiana Health Information Exchange (IHIE), Indianapolis Emergency Medical Services (IEMS), and Marion County Sheriff's Office (MCCO). This collaboration was novel because (1) the organizations had not previously contributed their data sets to a common research project and (2) the intent was to link records about *individuals* across all data sets. Table 1 provides an overview of the datasets and their relevant characteristics.ⁱⁱⁱ

The core dataset was a highly curated collection of data on 2331 individuals who died of an overdose in Marion County between 2010 and 2018. In order to help us obtain this information, the MCCO first provided a list of case numbers that were determined to be accidental fatal overdoses. A researcher then accessed the MCCO's database to search for each case and manually extract information from vital records and toxicology reports in order to compile the data into a single, composite file. Vital records included the entire death certificate for each deceased person. Demographic information included name, sex, race, date of birth, home address, marital status, education, and employment. Toxicology reports contained information on all substances that were detected in the deceased at the time of the decedent's death. The MCCO data was then linked to the other data using primarily the first name, last name, date of birth, and social security number (if available) of the deceased.²² In the case of IEMS and MCSO, the data contained only the first and last name, as well as the

date of birth. The RI/IHIE data was linked using these identifiers as well as social security numbers.

The RI/IHIE dataset deserves further comment due to its unique nature in terms of provenance and method of data collection. In contrast to the other data sets, which were collected directly and made available to the research team by a single organizational entity, IHIE data was collected through a large Health Information Exchange (HIE) and shared with the researchers via the Regenstrief Institute (RI). Since 2004, IHIE has overseen the Exchange to support patient care among its members while also allowing the Institute to provide data for research by a diverse range of partners. At the time of this writing, IHIE^{iv} has been involved in the collection and exchange of data with over 117 hospitals, 18486 practices, and 51688 providers in Indiana and elsewhere.^v Following a recent merger with 2 regional HIEs, IHIE has expanded its coverage of the Indiana Network for Patient Care (INPC) to all regions in Indiana and portions of southern Michigan. The information in the INPC comprises over 13 billion clinical data elements, such as encounters, lab tests, microbiology, pathology, radiology, cardiology, electrocardiograms, pharmacy data, and text reports. INPC data are aggregated from providers, payers, and public health agencies in real time, enabling functions such as clinical alerts that can influence patient care decisions at the point of care. Toward the end of 2022, the INPC contained at least one patient encounter for nearly two-thirds of the population in Indiana's 92 counties.

As Table 1 demonstrates, not all datasets covered the same time period. There are several reasons for this circumstance, such as changes in record-keeping processes and systems that housed the data. We included all data we could obtain but did not exclude individuals with incomplete data. To allow users to assess the impact of incomplete data, we designed a visualization that allows exploring the intersection of various datasets (see Appendix 2).

Dashboard Design

The design rationale for our visualization dashboard has been to make the interlinked data described above useful for research projects, policy development, and effective interventions.

In order to fulfill these goals, we have strived to follow the principles of responsive design, open design, and user-centered design.

Responsive design. From the use of scalable graphics and flexible grids to the clarity and concision of content, responsive design is primarily a question of detail, which makes it unsuitable for a written exposition. With this caveat, 2 aspects of the dashboard's responsiveness deserve specific mention.

First, in order to make the dashboard useful and usable to end users, we have interrogated data using a variety of visualization formats, including bar graphs, heat maps, timelines,

and geographical distribution maps. Each format draws attention to and allows inferences about certain facets of the complex dataset. For instance, bar graphs offer a comparative view of different data points and can be used to correlate and contrast data associated with various substances and prescriptions; heat maps provide a visual summary of changes in the intensity of substance use across time and age groups; timelines are good for illustrating trends and patterns over time; and geographical distribution maps provide insights into regional patterns and disparities. Combined, these visualizations can be used to glean insights that would have been challenging to discern in a more monolithic and/or less sophisticated presentation format.

Another key feature of the dashboard is its high degree of interactivity, which was made possible thanks to the incorporation of dynamic affordances into the design.²³ Affordances, which originate in the perceptual psychology of Gibson,²⁴ refer to the possibilities of use that an object offers. Thus, for example, the handle of a cup affords grasping, and its concavity affords carrying liquids. In the digital domain, designers can guide users in their interactions with digital objects through the careful design of affordances. Our visualizations offer many of the dynamic affordances that users have come to expect from digital objects. Users can turn various data filters on and off, select and de-select data variables, zoom and pan across maps, and slice time into custom intervals. The visualizations dynamically update themselves in response to these actions. This responsiveness makes for a more pleasant experience and encourages active user engagement, which has been shown to foster a deeper understanding of data and more nuanced interpretations of patterns.²⁵⁻²⁷

User-centered design. User acceptance and satisfaction is the ultimate goal of any design intervention. The primary audience for our visualizations are public health personnel at the local, regional, and federal levels, who are likely to have both the greatest interest in and use for the dashboard. A second group of users includes researchers, policymakers, and others with a stake in the substance use disorder epidemic. In evaluating the dashboard, our methodology was grounded in the principles of user-centered design,²⁸ which translated to 3 different practices:

1. At the outset of the project, 2 senior project researchers conducted individual, 30-min. semi-structured interviews with 8 researchers funded by IU's Addictions Grand Challenge. These interviews explored the different aspects of the projects in which the interviewees were involved, including stakeholders, interventions, outcome measurements, evaluation plans, and data and visualization needs. The interviews were subsequently analyzed and the insights were synthesized to inform the design of our project.
2. The research team designed the visualizations with occasional input from individual stakeholders.

Although the project was designed with more thorough stakeholder engagement, our plans were disrupted by the COVID-19 pandemic, which severely limited the availability of and access to public health personnel. The initial plan was thus scaled down to reflect these practical constraints. We conducted formal cognitive walkthroughs with a group of 8 stakeholders, encompassing researchers, policy-makers, and public health officials. Walkthrough sessions were virtual and consisted of facilitators sharing their screens, explaining the motivation behind the dashboard design, demonstrating the functionality of 5 distinct visualizations, and subsequently seeking participant queries and feedback. The insights and suggestions provided by the stakeholders were used to iterate on the dashboard design. We will return to this feedback in subsequent sections of the paper.

3. The dashboard was subjected to a heuristic review²⁹ by a human-computer interaction expert. Throughout this process, the expert investigated the user interface and visualizations, became familiar with the data, conducted a search operation using the term “Emergency Medical Services,” and scrutinized the results they received. The review aimed to identify potential concerns in 3 categories:
 - a. cosmetic issues pertaining to visual design that minimally affect the user experience;
 - b. minor problems causing some interactional difficulties and user confusion; and
 - c. major challenges that significantly disrupt the interaction flow, hindering the user’s progression

Our expert documented several concerns within the cosmetic category (eg, unconventional navigation menu links and naming inconsistencies), the minor category (eg, navigational difficulties between visualizations and readability complications), and the major category (eg, ambiguous user communication, limitations on user freedom, confusing links between interface elements, and so forth). A majority of these concerns have been addressed in subsequent iterations of the dashboard design, which we present below.

Open design. Last but not least, the dashboard design is fully transparent about the relationship between data sources and data visualizations. Underneath each visualization page is an accordion section titled Data and Graphic Variable Extraction Information, which details how the visualization is grounded in the linked data set. It makes the following available to users:

- **the source code** used to extract data in Structured Query Language (SQL), a programming language used for managing relational databases,

- **the extracted data** as Comma-Separated Value (CSV) files, which can be opened in spreadsheet programs such as Microsoft Excel and Google Sheets, and
- **the visualization source code** as Vega Lite Specifications,^{vi} which provide a concise and convenient format for authoring common visualizations.

Ensuring easy user access to data sources provides 2 main benefits. First, it helps improve user trust in the visualizations. According to Park et al,³⁰ knowledge of data source makes a significant positive contribution to how users interpret visual information. Second, it contributes to data openness by adhering to the principles of FAIRness, that is, findability, accessibility, reusability, and interoperability of data.^{31,vii} FAIRness, however, does not entail lax access to data and accessibility should be granted under carefully defined conditions.^{viii}

Results

The interviews with 8 researchers funded by IU’s Addictions Grand Challenge conducted at the outset of the project provided a rich set of requirements for potential visualizations. Some examples follow:

1. Researchers emphasized the need to link data about individuals across different sources of data, including risk factors for substance use, prescriptions and mental health diagnoses, or justice involvement history and results of drug screens.
2. They underscored the value of capturing longitudinal data about individuals in order to “examine individual stories” about their trajectory through substance use disorder.
3. They highlighted the importance of designing flexible visualizations that can be used to identify emerging trends, for instance the appearance of new substances (or combinations thereof) in the data.
4. They insisted on the requirement that visualizations be able to focus on diverse populations, such as babies born to mothers suffering from substance use disorder, youths, and adults.

Findings from the cognitive walkthrough sessions enhanced our dashboard designs in different ways. The following list summarizes these findings and offers examples from each category:

- absence of clear action prompts for users, for example, ambiguity in instructions for zooming, selecting a specific range, or resetting the visualization;
- minor discrepancies across the visualizations, for example, age data, should consistently be displayed either from low to high or vice versa across all figures;

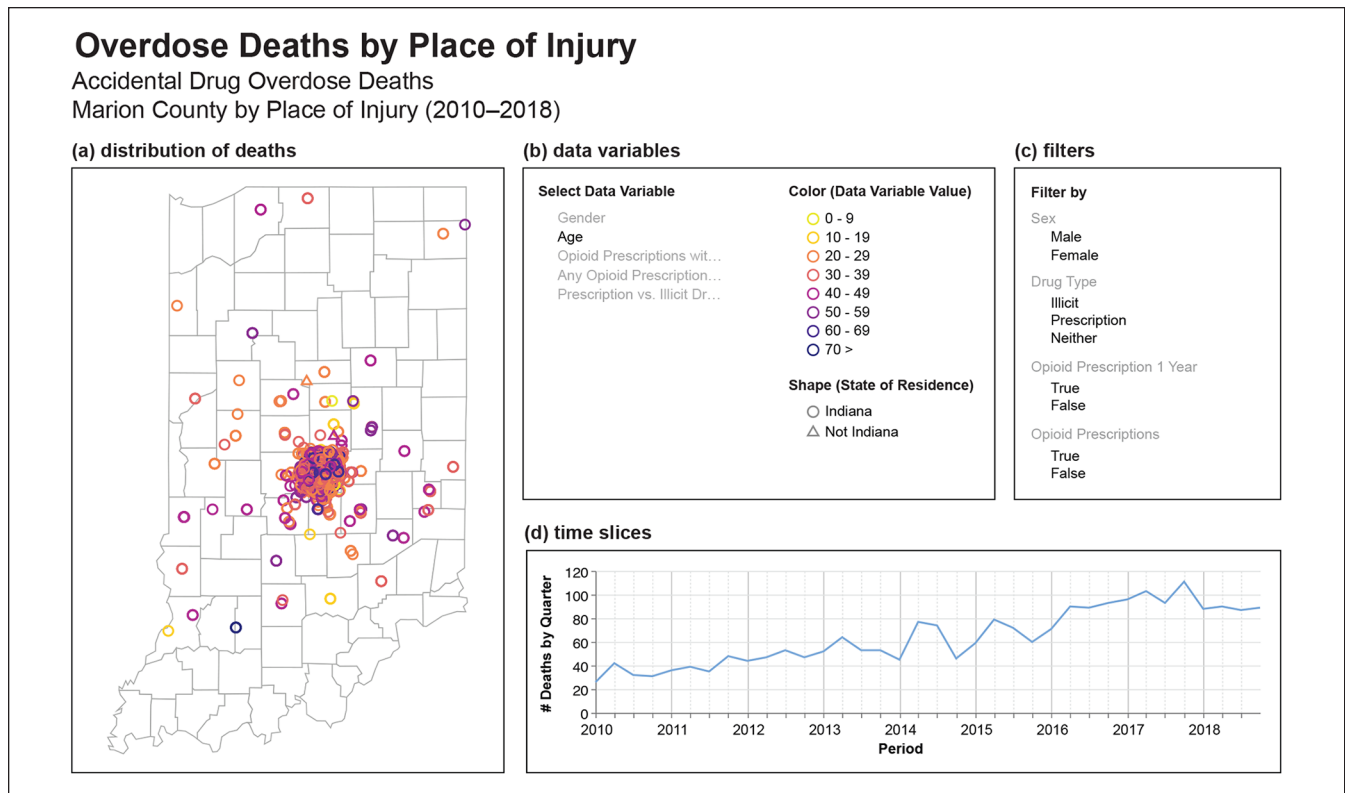


Figure 1. Visualization of the geographic context of accidental drug overdose deaths. This interactive visualization allows users to explore the relationship between the geographic distribution of deaths (a) and a range of variables (b), filters (c), and time slices (d). The panels have been rearranged to improve presentation clarity.

- recommendations for enhancing the clarity of information presentation, for example, modifications to the color palette used in figure legends; and
- calls for expanded functionality, for example, a heightened emphasis on justice data

The developed visualizations help viewers gain insights from a rich set of data about overdose deaths. In what follows, we illustrate the visual tools and analytics our dashboard offers using 3 examples. Based on the same dataset, each visualization illuminates a different facet of the data, enabling users to gain insights and potentially answering distinct questions regarding the substance use disorder epidemic.

The visualization in Figure 1 provides a dynamic distribution of overdose data, informing users about the geographic context of deaths and allowing them to ask questions such as, How are overdose deaths distributed across counties and neighborhoods? The map shown in panel (a) of Figure 1 affords zooming and panning to compare deaths between neighborhoods in Marion County and across the different counties of Indiana. Users can also explore how the distribution evolves across time using panel (d)'s interactive line chart. They can left-click and drag the mouse pointer to specify a customized window of time. As the user moves this time window back and forth in the chart, the map updates the

death distributions accordingly. The variables in panel (b) can be used to select one or several data variables of interest in order to compare changes and discern patterns. Some examples of these patterns include:

1. The map illustrates a differential distribution of deaths in Marion County and draws attention to “hotspots,” that is, areas of high overdose concentration. Although these hotspots move over time, they appear to cycle between a few locations.
2. Injury sites appear more concentrated in Marion County for the youngest (under 20) and oldest (over 50) people, and more dispersed throughout Indiana for the middle bracket (20-50). This makes intuitive sense, given the lower mobility of the youngest and oldest age groups.
3. Non-Indiana residents who overdose seem to be more concentrated in Marion County.

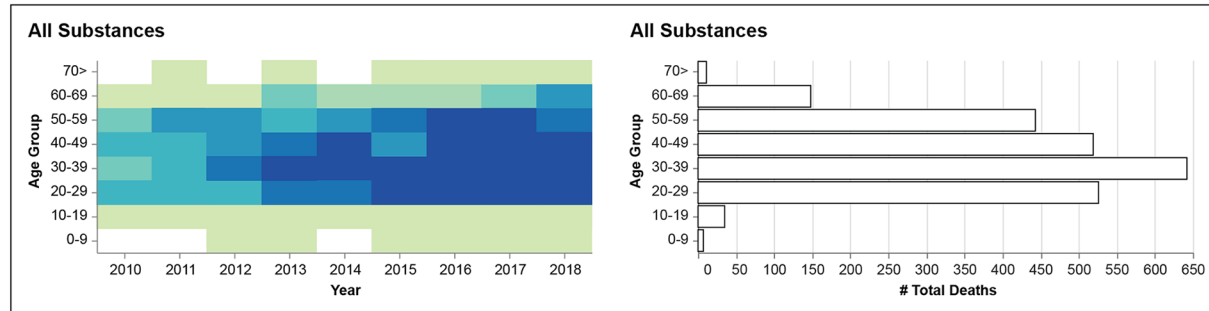
All of these are, of course, exploratory observations that must be further probed and substantiated through rigorous statistical analyses and machine learning techniques, highlighting once again the importance of making the chain of inference between datasets and visualizations fully transparent.

Overdose Deaths by Substance, Sex, and Age

Accidental Drug Overdose Deaths

Marion County by Substance, Sex, and Age (2010–2018)

(a)



(b)

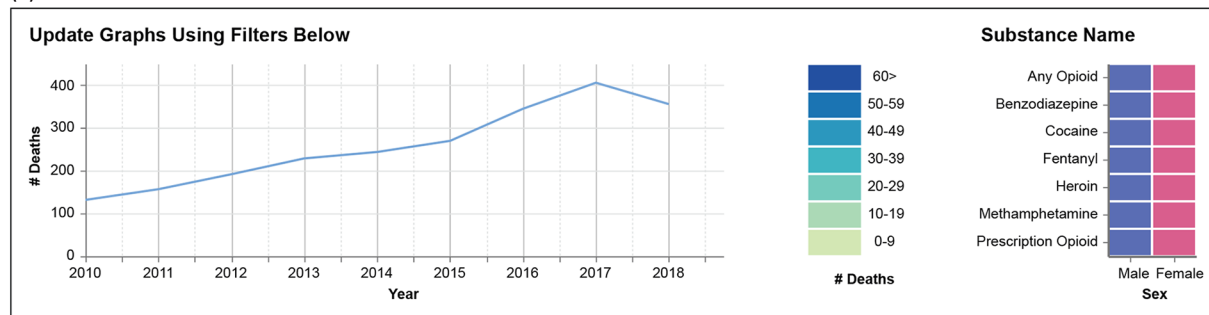


Figure 2. Visualization of drug overdose deaths by substance, gender, and age. The heat map in panel (a) at the top left depicts the distribution of deaths by age group and year, while the bar chart on the right sums each row of the heat map. Panel (b) allows the user to filter these 2 linked interactive visualizations by time, gender, and one or more substances. The panels have been rearranged to improve presentation clarity.

Figure 2 illustrates shifts in substance use, and invites the user to interrogate age, gender, and time patterns. Color gradients in heatmaps make them particularly effective tools for making patterns visible. Two kinds of patterns that can be more easily observed with heatmaps are clusters and thresholds. For example, the 30 to 39 age group shows the highest number of deaths for all substances between 2010 and 2018. Similarly, the 4 age groups that make up the range between 20 and 59 form a single age cluster with the highest death rate. One can make related observations about threshold phenomena. For example, the heatmap shows an ascending “staircase” of intensity beginning with the years 2012 and 2013, suggesting that something potentially significant may have happened around that time. A similar threshold phenomenon can be observed if we filter the results by substance (not shown in Figure 2). We can see a significant increase in the number of deaths associated with Fentanyl and methamphetamines starting around 2017, as well as a relative decline in the number of deaths associated with prescription opioids starting in 2017. This correlates with a change in Indiana law: Among other stipulations, Indiana Senate Bill 226³²

required state-licensed practitioners to limit the quantity of opioids that they prescribed. These clusters, thresholds, and staircases are higher-order visualization phenomena that can sensitize us to patterns that otherwise might be missed.

Figure 3 presents another view of the dataset that is more focused on aggregating individual data and highlighting differences between individual trajectories. Some people receive prescription opioids for many years before experiencing their first overdose, whereas others do not experience any overdose events until they die. However, the main advantage of this visualization over others is that it depicts life trajectories rather than statistical patterns. The visual correlations it draws allow the user to follow 4 types of events (ie, health encounters, opioid prescriptions, incarcerations, and overdoses) across the life of an individual up to its tragic end. The user can hover over each line to display detailed information about the person, including but not limited to time of first overdose, age at death, total number of prescriptions, overdoses, and incarcerations.

The health trajectories of Figure 3 supply the user with the elements of a sequential story line that begins with each

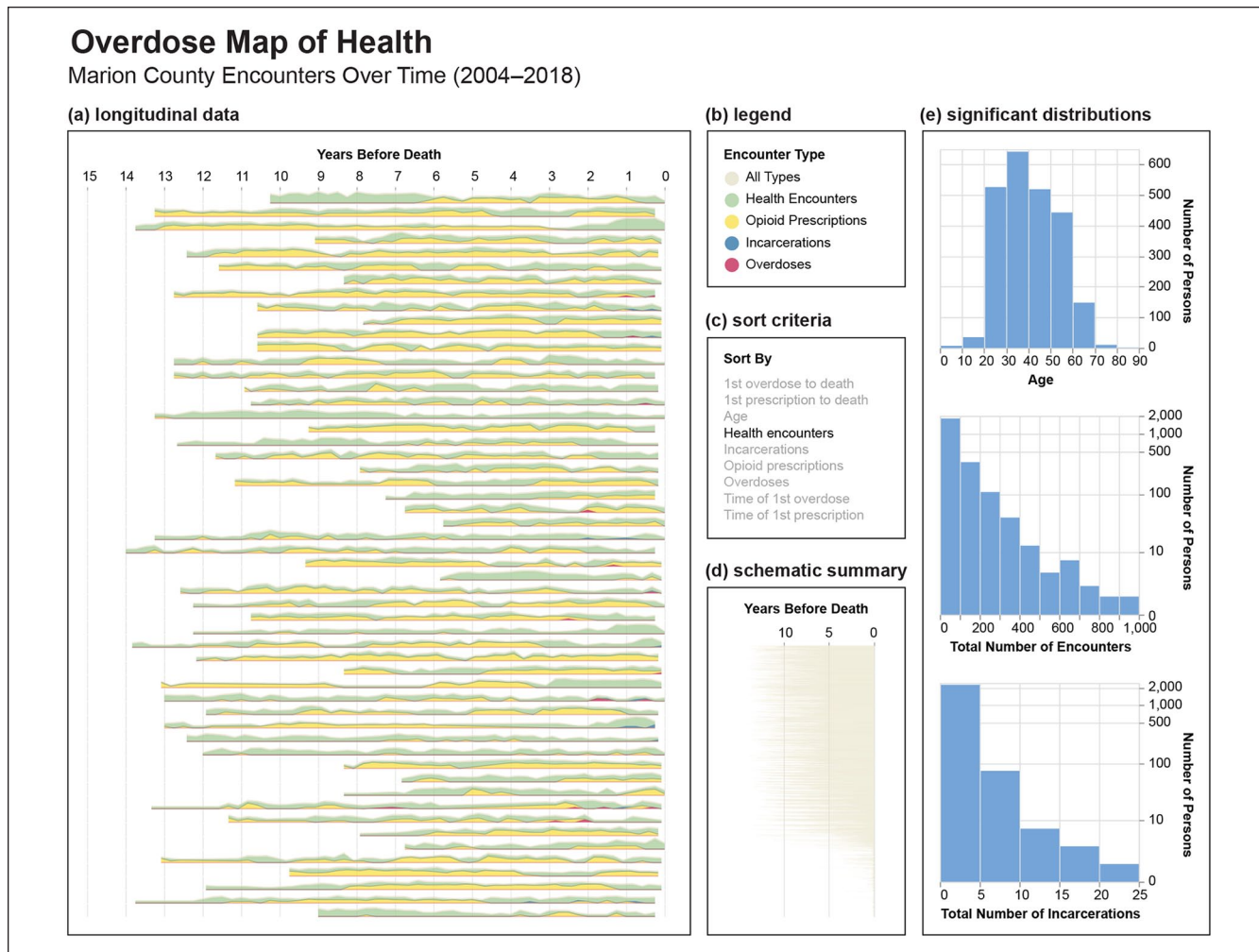


Figure 3. Map of Health: (a) longitudinal death, (b) legend, (c) sort criteria, (d) schematic summary, and (e) significant distributions. One of the 3 maps of health included in the dashboard, this visualization offers a unique perspective on the dataset. Each row combines data from the 4 original datasets in a way that is unique to the person in question. The data can be sorted according to different criteria. Panel (e) on the right shows distribution by age, total encounters, and total number of incarcerations. The panels have been rearranged to improve presentation clarity.

individual’s first encounter and ends with their death by accidental overdose. Of the 3 visualization formats we have discussed, this is the most narratologically rich. Narrative visualization is an emerging area of research that pairs data visualization with narrative techniques.^{33,34} This pairing introduces new opportunities for storytelling about or through data using novel interaction patterns.

Different data visualization techniques can be based on the same data set but follow different display logics. In our dashboard design, for example, Figure 1 highlights a logic of distribution, Figure 2 a logic of intensity, and Figure 3 a logic of stratification and layering. Understanding these display logics is important for both designers and users if the dashboard is to serve its purpose as a plank in the various platforms for research, policy development, and practical interventions with regard to the substance use crisis.^{ix}

Discussion and Limitations

It is vital to understand and appreciate the critical role large datasets can play in developing a cohesive, data-driven plan for understanding and addressing public health problems such as the substance use disorder epidemic. Consolidating diverse data sources in one place creates nuance, gives a more complete picture of individuals’ lives, and allows stakeholders and decision-makers to identify patterns that might be less apparent when examining data sources separately. Suppose, for example, that individuals who experienced fatal opioid overdoses also had a high frequency of specific encounters with healthcare providers or particular arrest records. Such correlations could be investigated for an increased risk of opioid overdose. As another example, linking deceased individuals to prescription drug histories

encourages a deeper analysis of correlations between prescribed medications and the onset of substance use disorders. Such patterns can inform targeted prevention strategies. Similar trends may remain siloed in separate datasets (eg, between geographic location, gender, health history, and interactions with emergency services and law enforcement).

Although record linkage is both necessary and insightful, it has not been devoid of complexity. A considerable amount of effort is required to procure data from a multitude of organizations. Diverse privacy requirements make data acquisition a complex task, occasionally even rendering it unfeasible. As an example, we were unable to gain direct access to data from Indiana's prescription drug management program, INSPECT.⁸ Overcoming such obstacles is a delicate sociotechnical dance to improve data accessibility without jeopardizing privacy and confidentiality. It goes without saying that higher quality datasets can decrease the burden of research and expedite the generation of timely insights into public health issues.

Mere data linkage is insufficient, however. For data to serve its public health purpose effectively, it should also be highly usable. The stakeholders involved in addressing the substance use disorder epidemic come from diverse backgrounds and tend to have varying levels of technical expertise, data and/or visualization literacy, or domain knowledge. By following the principles of user-centered, responsive, and open design, our dashboard makes it easier for them to comprehend and engage with the data, identify patterns and trends, and make informed, data-driven decisions. During our cognitive walkthroughs, feedback from several participants emphasized this point. An employee of the Indiana Department of Health succinctly captured this sentiment, commenting that the dashboard design is "so much better than anything we have available." Another participant envisioned the possibility of conducting comparative analyses with multiple dashboards in parallel, exploring, for instance, the correlations between the relative fall in the rate of opioid incidents and the relative rise in the rate of methamphetamine incidents.

Our work is subject to several limitations. First, while our core dataset covers the period of 2010 to 2018, we were not able to obtain data from all other sources for the same time period. Most notably, the justice involvement data only cover the period of 2014 to 2018 (see Table 1). The resulting "holes" in our data limit the conclusions that can be drawn. Second, while we used state-of-the-art linking approaches, probabilistic linking of data is subject to a certain degree of imprecision and is prone to both false positive and false negative matches. Despite our best efforts, failure to match and errors in matching are inevitable. Formal quantification of matching precision is possible but was beyond the scope of this project. Third, our cognitive walkthroughs demonstrated that complex interactivity and usability do not always go together. For instance, one of the most useful features of the first visualization (Figure 1) is the ability to zoom and pan across the geographical map. Yet the user-centered imperative to keep the interface simple and lean necessitated implementing these features using mouse controls, which hindered their discoverability. Like many of

the subtler features of the interface, our solution to this problem was to deploy a context-sensitive help menu. Whether this and many other information design choices have been sound ones awaits larger-scale user testing and broader adoption of our dashboard. Fourth, our current dashboard system is only accessible to authorized users due to privacy reasons. To address this limitation, we are working to create a public version that uses simulated data. Last, while this study's focus has been on data visualization, it is essential to note that linked datasets can offer valuable insights in other ways that are beyond the scope of the current work. These include, but are not limited to, analyses using statistical methods, machine learning techniques, and more sophisticated visualization narratives.

Conclusion

This project has been an attempt to "push the envelope" in aggregating and visualizing data to help foster collaboration among stakeholders and generate useful insights for addressing the substance use disorder epidemic. We have successfully integrated 4 different data sets about healthcare encounters, opioid prescriptions, incarcerations, and overdoses for 2331 individuals who died from drug overdoses between 2010 and 2018 in Marion County, IN. The linked dataset has been the basis for the development and implementation of visualizations beyond the standard dashboard designs for the substance use disorder epidemic and similar public health applications. Key to our success has been a high-quality, comprehensive dataset about overdose deaths in Marion County, which was developed through a multi-year effort. The principles of responsive, user-centered, and open design have served as our design guidelines. The resulting visualizations, including those profiled in this paper, use different display logics to shed light on the same data from different perspectives. They also provide a high degree of interactivity through a simple and highly usable interface. Further work is needed to expand and iterate on our visualization dashboard design, in particular when it comes to its ability to help public health personnel make practical decisions about the substance use disorder epidemic.

Appendix I

List of state dashboards that were reviewed:

1. Alabama
2. Alaska
3. Arizona
4. Arkansas
5. California
6. Colorado
7. Connecticut
8. Delaware
9. Florida
10. Georgia
11. Hawaii

- 12. Idaho
- 13. Illinois
- 14. Indiana
- 15. Kansas
- 16. Kentucky
- 17. Louisiana
- 18. Maine
- 19. Massachusetts
- 20. Michigan
- 21. Minnesota
- 22. Missouri
- 23. Montana
- 24. Nevada
- 25. New Hampshire
- 26. New Jersey
- 27. New Mexico
- 28. New York
- 29. North Carolina
- 30. Ohio
- 31. Oregon
- 32. Pennsylvania
- 33. Rhode Island
- 34. South Carolina
- 35. South Dakota
- 36. Tennessee
- 37. Texas
- 38. Utah
- 39. Virginia
- 40. Washington State
- 41. West Virginia
- 42. Wisconsin
- 43. Wyoming

Appendix 2

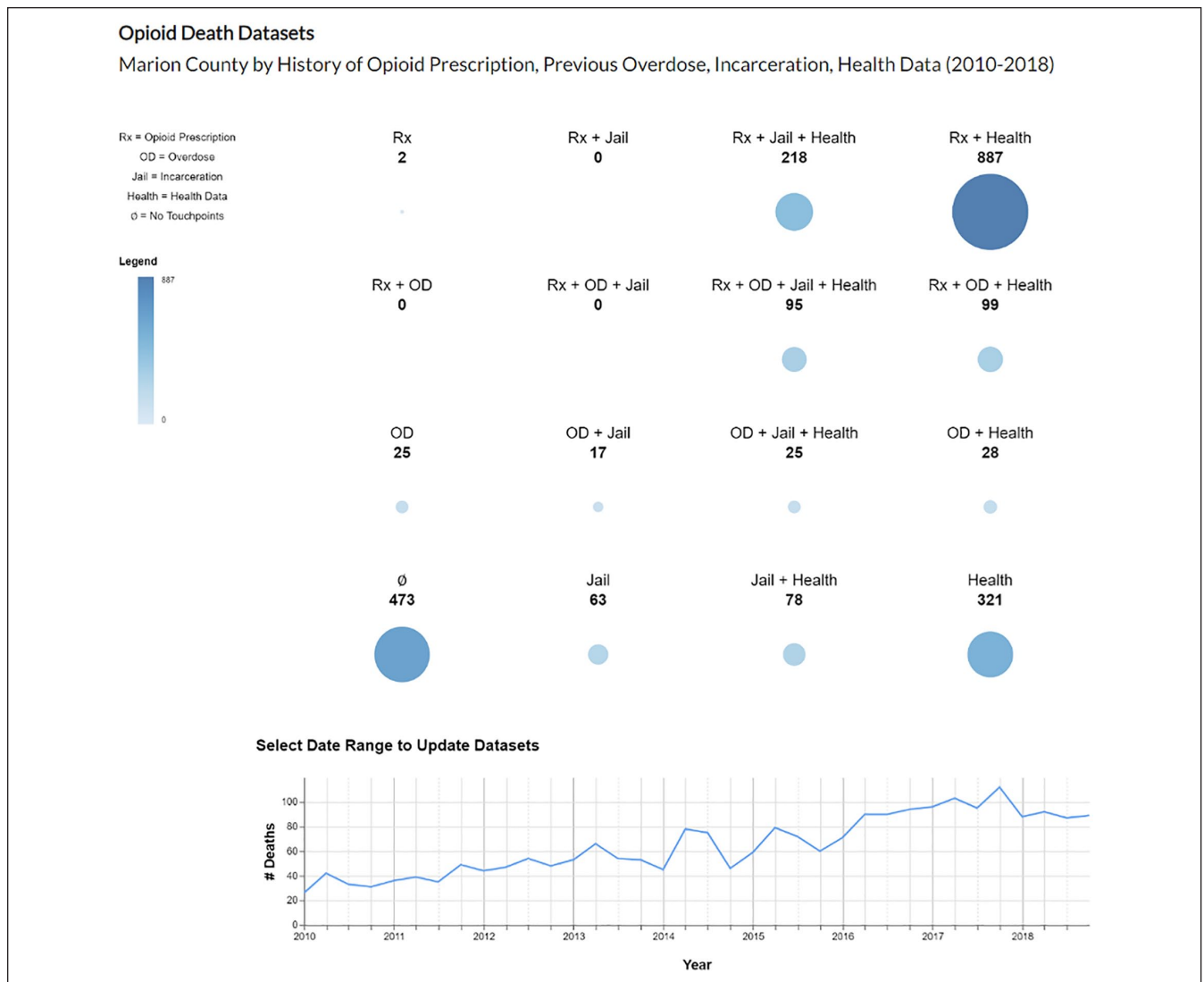


Figure A1. Opioid death datasets. This visualization shows the number of individuals contained in the intersection of the Opioid Death Dataset with other source datasets (History of Opioid Prescription, Previous Overdose, Incarceration, Health Data in Marion County). The visualization at the bottom allows the user to choose a custom time range for in-depth analysis.

Acknowledgments

We are grateful for the help of many individuals in this project, including Brad Ray, Philip Huynh, Kosali Simon, Sumedha Gupta, the members of our technical teams, our colleagues funded by the IU Addictions Grand Challenge, and the group of stakeholders who worked with us on the design and evaluation of the visualizations. Finally, we thank two anonymous reviewers for their encouragements and constructive critiques.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was funded by Indiana University Addictions Grand Challenge and the Indiana University Health and the Indiana Clinical and Translational Sciences Institute, grant ULI TR002529 from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Science Award.

ORCID iDs

Danial Qaurooni  <https://orcid.org/0000-0001-6190-073X>

Bruce W. Herr  <https://orcid.org/0000-0002-6703-7647>

Notes

- i. Data available at the National Center for Health Statistics at <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm> and at Indiana Department of Health's Office of Data Analytics: https://gis.in.gov/apps/isdh/meta/stats_layers.htm, respectively.
- ii. <https://www.in.gov/mph/>
- iii. Since the data sets only included information about deceased individuals, the Regenrief Institute Privacy and Compliance Officer determined that the project was not human subjects research. However, since the visualizations included individually identifiable data (such as the location of death), the officer stipulated that they needed to be made available only to individually authorized users, not the public.
- iv. Accessible at: <https://www.ihie.org/>
- v. Accessible at: <https://www.ihie.org/participant-list/>
- vi. Accessible at: <https://vega.github.io/vega-lite/docs/spec.html>
- vii. As an aside and in line with FAIRness principles, all visualizations can also be exported in multiple graphic formats, including Portable Network Graphic (PNG) and Scalable Vector Graphic (SVG).
- viii. Per the direction of the Regenrief Institute's Privacy and Compliance Officer (see footnote 3), the visualizations are only accessible to users who have been authenticated and have access permissions granted through their Indiana University or Google accounts. This level of access control is necessary. Although the visualizations do not contain any personally identifying information, the data is fairly granular and enough contextual knowledge allows for the identification of individuals who have died from a drug

overdose. While the confidentiality provisions of the Health Insurance Portability and Accountability Act of 1996 do not apply to deceased individuals, both Indiana University and the Regenrief Institute decided not to make the dashboard public.

- ix. Our platform includes one other visualization, which we were unable to cover due to space limitations. Age Group And Gender of Accidental Drug Overdose allows viewers to relate drug overdoses to Marion County's population pyramid. This visualization offers another perspective that complements the 4 (Figures 1-3 and Appendix 2) discussed here.
- x. Accessible on the website of Indiana State's government at: <https://www.in.gov/pla/inspect/>

References

1. Biancuzzi H, Dal Mas F, Brescia V, et al. Opioid Misuse: A review of the main issues, challenges, and strategies. *Int J Environ Res Public Health*. 2022;19(18):11754. doi:10.3390/ijerph191811754
2. Krausz RM, Westenberg JN, Ziafat K. The opioid overdose crisis as a global health challenge. *Curr Opin Psychiatry*. 2021;34(4):405-412. doi:10.1097/YCO.0000000000000712
3. Battaglia M, Groenewald CB, Campbell F, et al. We need to talk: the urgent conversation on chronic pain, mental health, prescribing patterns and the opioid crisis. *J Psychopharmacol*. 2023;37(5):437-448. doi:10.1177/02698811221144635
4. Fink DS, Keyes KM, Branas C, et al. Understanding the differential effect of local socio-economic conditions on the relation between prescription opioid supply and drug overdose deaths in US counties. *Addiction*. 2023;118(6):1072-1082. doi:10.1111/add.16123
5. Manchikanti L, Singh VM, Staats PS, et al. Fourth wave of opioid (Illicit drug) overdose deaths and diminishing access to prescription opioids and interventional techniques: cause and effect. *Pain Physician*. 2022;25(2):97-124.
6. Simha S, Ahmed Y, Brummett CM, et al. Impact of the COVID-19 pandemic on opioid overdose and other adverse events in the USA and Canada: a systematic review. *Reg Anesth Pain Med*. 2023;48(1):37-43.
7. Blanco C, Wall MM, Olfson M. Data needs and models for the opioid epidemic. *Mol Psychiatry*. 2022;27(2):787-792. doi:10.1038/s41380-021-01356-y
8. Braude RM. People and organizational issues in health informatics. *J Am Med Inform Assoc*. 1997;4(2):150-151.
9. Marquard J. Human factors and organizational issues in health informatics: Innovations and opportunities. *Yearb Med Inform*. 2021;30(1):91-099.
10. Warner M, Paulozzi LJ, Nolte KB, Davis GG, Nelson LS. State variation in certifying manner of death and drugs involved in drug intoxication deaths. *Acad Forensic Pathol*. 2013;3(2):231-237.
11. Boyd AD, Gonzalez-Guarda R, Lawrence K, et al. Potential bias and lack of generalizability in electronic health record data: reflections on health equity from the National Institutes of Health Pragmatic Trials Collaboratory. *J Am Med Inform Assoc*. 2023;30:1561-1566. doi:10.1093/jamia/ocad115
12. 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;28:1.

13. Ferris LM, Weiner JP, Saloner B, Kharrazi H. Comparing person-level matching algorithms to identify risk across disparate datasets among patients with a controlled substance prescription: retrospective analysis. *Jamia Open*. 2022;5(1):8. doi:10.1093/jamiaopen/ooac020
14. 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.
15. Kline D, Hepler SA. Estimating the burden of the opioid epidemic for adults and adolescents in Ohio counties. *Biometrics*. 2021;77(2):765-775. doi:10.1111/biom.13295
16. Allen B, Neill DB, Schell RC, et al. Translating predictive analytics for public health practice: A case study of overdose prevention in Rhode Island. *Am J Epidemiol*. 2023;192:1659-1668. doi:10.1093/aje/kwad119
17. Bharat C, Hickman M, Barbieri S, Degenhardt L. Big data and predictive modelling for the opioid crisis: existing research and future potential. *Lancet Digit Heal*. 2021;3(6):e397-e407. doi:10.1016/S2589-7500(21)00058-3
18. Thayer JG, Ferro DF, Miller JM, et al. Human-centered development of an electronic health record-embedded, interactive information visualization in the emergency department using fast healthcare interoperability resources. *J Am Med Inform Assoc*. 2021;28(7):1401-1410. doi:10.1093/jamia/ocab016
19. 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):4.
20. 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 Rep*. 2022;137(5):944-954.
21. Mosen DM, Rosales AG, Mummadi R, Hu W, Brooks N. Demographic, clinical, and prescribing characteristics associated with future opioid use in an opioid-naive population in an integrated health system. *Perm J*. 2020;24:64-69.
22. Grannis S, Overhage J, McDonald C. Analysis of identifier performance using a deterministic linkage algorithm. *Proc AMIA Symp*. 2002:305-309.
23. Norman DA. *The Psychology of Everyday Things*. Basic Books; 1988.
24. Gibson JJ. *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press; 2014.
25. Cano MB, Perry P, Ashman R, Waite K. The influence of image interactivity upon user engagement when using mobile touch screens. *Comput Human Behav*. 2017;77:406-412.
26. Sundar S. Social psychology of interactivity in human-website interaction. Published online 2009.
27. Sundar SS, Bellur S, Oh J, Jia H, Kim HS. Theoretical importance of contingency in human-computer interaction: Effects of message interactivity on user engagement. *Commun Res*. 2016;43(5):595-625.
28. Still B, Crane K. *Fundamentals of User-Centered Design: A Practical Approach*. CRC Press; 2017.
29. Tory M, Moller T. Evaluating visualizations: do expert reviews work? *IEEE Comput Graph Appl*. 2005;25(5):8-11.
30. Park S, Bekemeier B, Flaxman A, Schultz M. Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. *Inform Health Soc Care*. 2022;47(2):175-193.
31. Wilkinson MD, Dumontier M, Aalbersberg IJ, J. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3(1):1-9.
32. Merritt J, Ruckelshaus J, Charbonneau E, Houchin E, Mrvan F. Senate Bill 226: Prescribing and dispensing of opioids. Published online 2017. <https://iga.in.gov/legislative/2017/bills/senate/226/details>
33. Lee B, Riche NH, Isenberg P, Carpendale S. More than telling a story: transforming data into visually shared stories. *IEEE Comput Graph Appl*. 2015;35(5):84-90.
34. Segel E, Heer J. Narrative visualization: telling stories with data. *IEEE Trans Vis Comput Graph*. 2010;16(6):1139-1148.