

**Office of the Secretary Patient-Centered Outcome Research Trust Fund Project HP-18-005
(IAA 750118PE090003)**

**National Center for Health Statistics
Centers for Disease Control and Prevention**

**Strengthening the Data Infrastructure for Outcomes Research on
Mortality Associated with Opioid Poisonings**

Final Report

March 13, 2023

Table of Contents

| | |
|--|----|
| Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings | 1 |
| Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings | 3 |
| Executive Summary..... | 3 |
| Background..... | 3 |
| Coding and Processing of Mortality Records..... | 4 |
| Quality of Drug Information on Death Certificates..... | 5 |
| Dissemination of Drug Information..... | 5 |
| Alignment of Mortality Infrastructure Improvements with the Needs of Researchers..... | 5 |
| Tasks, Objectives, and Deliverables | 6 |
| Major Accomplishments..... | 7 |
| Redesign of the mortality medical data system to incorporate specific drug information captured in the literal text..... | 7 |
| Extract supplemental drug information contain in the literal text fields and annually produce the NDI and the NVSS-M data files containing the supplemental information for those deaths involving drugs such as opioids..... | 8 |
| Promote standards-based interoperability between systems containing mortality information at the state level and improve the collection of drug information on death certificates | 9 |
| Re-architect the provisional data system for VSRR program to produce and release more in-depth information about drug overdose data (e.g., specific drugs, demographic information) on a monthly basis for public health surveillance..... | 9 |
| Align and ensure utility of mortality data infrastructure improvements with end-users’ needs, especially researchers, with studying drug overdose deaths involving opioids using the NVSS | 10 |
| Lessons Learned and Considerations for Future Work | 11 |

Strengthening the Data Infrastructure for Outcomes Research on Mortality Associated with Opioid Poisonings

FINAL REPORT

Executive Summary

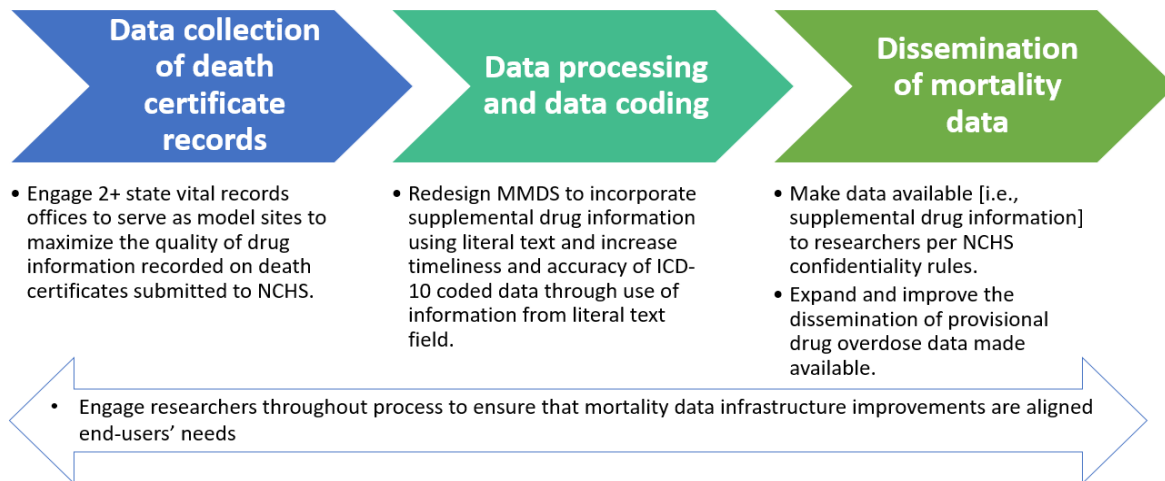
The purpose of this project was to strengthen the mortality data produced by the National Center for Health Statistics' (NCHS) National Vital Statistics System (NVSS) infrastructure for outcomes-based research on deaths associated with opioid poisoning. There are several distinct processes that were modified and enhanced. NCHS redesigned the Division of Vital Statistics (DVS) cause of death IT systems to electronically code the multiple cases of death listed as text strings on death certificates using natural language processes and machine learning algorithms. NCHS enhanced the incorporation of details about drugs named on death certificates onto supplemental data files for use by approved researchers by pulling these details for use on provisional data. This information is being obtained from the death certificate literal text fields and provides details on specific drugs that cannot be conveyed solely through ICD-10 coding. The timeliness of death record reporting from vital record offices for drug-involved deaths as well as the specificity of the drug information reported on death certificates has increased through the creation of HL7 Fast Healthcare Interoperability Resources (FHIR) standards-based interoperability between systems containing mortality information, with a focus on improving data exchange between medical examiners' and coroners' offices, state vital records offices, and the NVSS maintained by NCHS. The NCHS DVS Vital Statistics Rapid Release (VSRR) program has been restructured to produce and release more in-depth information regarding drug overdoses monthly for public health surveillance and research. These efforts enabled NCHS to ensure alignment and utility of the mortality infrastructure improvements with end-users' needs.

Background

NCHS DVS received support from the Department of Health and Human Services' Office of the Secretary Patient-Centered Outcomes Research Trust Fund (PCORTF) to modernize the infrastructure for capturing drug death data and enhancing research on opioid poisoning. This report summarizes the FY18 PCORTF project's goals, milestones, accomplishments, lessons learned, and results. The project began in 2018 and was completed in 2022, after a delay resulting from diversion of resources to respond to the COVID-19 pandemic.

Researchers and public health officials frequently rely on the cause of death information made available from the coding and processing of death certificate data for critical programmatic, policy, and research needs. NCHS has followed a multi-pronged approach for modernizing the capture of death certificate data (Figure 1).

Figure 1. Multi-Pronged Approach to Modernizing the Infrastructure for Capturing Drug Death Data and Enhancing Research on Opioid Poisoning using Death Certificates' Literal Text Fields



Coding and Processing of Mortality Records

The Mortality Medical Data System (MMDS), a subsystem of NCHS' National Vital Statistics System (NVSS), was used to electronically code all death certificate records in the United States at the time this project began in 2018. MMDS was developed in the 1980s and by 2018 required extensive technological updates to improve its antiquated and inefficient automated coding processes. MMDS used complex algorithms to assign underlying cause of death, multiple causes of death, and other fields using data inputs from the medical and demographic portions of the death certificate. Its ability to automatically code varied greatly by cause of death. Certificates that could not be electronically coded were submitted to a queue of records that required human coding, a process that added days to weeks to the processing time. Drug overdose deaths, which had become a national public health priority in the U.S. at the time of the project's initiation, were generally not electronically coded. In 2018, while 79% of all death certificate records were coded electronically only 33% of records with a drug overdose death were coded electronically. Timely and accurate reporting of these deaths is critical for surveillance and research efforts.

The literal text portion of the death certificate, where physicians may write in details of the cause and circumstance of a death, contains detailed information about drugs that caused or contributed to overdose death. To make these data more accessible for research purposes, NCHS and Food and Drug Administration had previously developed a suite of software programs that identify the specific drugs mentioned in the unstructured death certificate literal text, referred to as NVSS Drugs Mentioned with Involvement (NVSS-M DMI program). One aspect of this project was to work in collaboration with the National Library of Medicine (NLM) to enhance and integrate the NVSS-M DMI programs into routine MMDS processing of the mortality data. The enhanced programs would be used to help strengthen the MMDS International Classification of Diseases, Tenth Revision (ICD-10) coding process, identify

supplemental drug information found in the literal text field, and include the supplemental information on drug involved deaths in the National Death Index (NDI) and the NVSS-M files to be used for research.

Quality of Drug Information on Death Certificates

One goal of this project was to improve the quality of drug information being reported on death certificates for deaths associated with drug overdoses. In data year 2016, about 15% of death certificates indicating a drug overdose did not include the names of the specific drug involved. Previous research demonstrated that specificity on death certificates improves when physicians complete death certificates using their electronic medical records. In the case of drug overdose deaths, medical examiners and coroners (ME/Cs) are typically the certifying officials. Thus, providing ME/Cs the ability to integrate their case management systems with electronic death registration system (EDRS) was explored for its potential to improve the quality of data on drug overdose deaths and other deaths of interest to health researchers studying patient outcomes.

Dissemination of Drug Information

NCHS/DVS created the Vital Statistics Rapid Release (VSRR) Program to support more timely dissemination of vital statistics data. The VSRR supports surveillance of deaths associated with drug overdoses through: 1) quarterly provisional estimates of national death rates for leading causes of death including drug overdoses; and 2) monthly provisional counts of drug overdose deaths by state for a limited set of drugs identifiable from ICD-10 codes. The provisional monthly and quarterly releases provided timely information, but their utility was reduced by the limited demographic and geographic detail in the reports, as well as limited detail on specific drugs involved in the death. The provisional data system used to produce the drug mortality information from VSRR were to be enhanced to enable the inclusion of geographic (e.g., region and state) and demographic (e.g., age and sex) detail, and include information on a fuller range of drugs involved in the death. Further, the project was expected to enhance the system tools for tabulating and analyzing the provisional data to both simplify and automate the production of standard estimate releases, mainly via web-based dashboards (i.e., quarterly and monthly) and ad hoc VSRR reports, enabling NCHS to be more responsive to current and future needs of decision-makers for information.

Alignment of Mortality Infrastructure Improvements with the Needs of Researchers

One of the key drivers of the modernization of the mortality data infrastructure is the necessity to meet the needs of researchers, especially as they relate to drug overdose deaths. Aligning researchers' interests and needs required bidirectional communication and engagements throughout the project to ensure that the work was relevant, feasible and meaningful with how the mortality data system is organized and governed within the United States.

Objectives, and Deliverables

This project had the following objectives and deliverables.

| Objectives | Deliverables |
|---|---|
| <p>Objective #1: Redesign of MMDS to incorporate specific drug information captured in the literal text.</p> | <ul style="list-style-type: none"> • Functional MMDS that can leverage literal text fields on death certificate data to provide new information on the NDI and NVSS-M on deaths involving opioids. • Publication(s) (e.g., National Vital Statistics Report, American Journal for Public Health, NCHS' website) that describe the new MMDS system and its added features, the new supplemental drug data on the morality files for researchers, and/or sample research questions that can be answered with the produced from the new MMDS system. |
| <p>Objective #2: Extract supplemental drug information contained in the literal text fields and annually produce the NDI and the NVSS-M data files containing the supplemental information for those deaths involving drugs such as opioids.</p> | <ul style="list-style-type: none"> • A published feasibility study leveraging the developed strategy and/or structured data elements to be included in the NDI and NVSS-M data files that may serve as a proof-of-concept for the types of research questions that can be answered from the redesigned system. |
| <p>Objective #3: Promote standards-based interoperability between systems containing mortality information at the state level and improve the collection of drug information on death certificates.</p> | <ul style="list-style-type: none"> • An open-source, scalable, secure, and piloted application programming interface (API) that exchanges relevant information for states to use to operationalize within their own respective jurisdictions. |
| <p>Objective #4: Re-architect the provisional data system for the VSRR program to produce and release more in-depth information about drug overdose data (e.g., specific</p> | <ul style="list-style-type: none"> • An enhanced VSRR system that produces monthly provisional estimates of drug overdose deaths that includes supplemental drug information generated from other sources, such as national provisional estimates of specific drugs beyond ICD-10 codes traditionally used to identify drug overdose deaths (e.g., fentanyl). |

| | |
|--|--|
| <p>drugs, demographic information) on a monthly basis for public health surveillance.</p> | <ul style="list-style-type: none"> • An NCHS publication on the new VSRR and the benefits of this new system for information on opioid and other causes of death. |
| <p>Objective #5: Align and ensure utility of mortality data infrastructure improvements with end-users' needs, especially researchers, with studying drug overdose deaths involving opioids using the NVSS.</p> | <ul style="list-style-type: none"> • Report on the expressed needs and interests of researchers and how the mortality infrastructure changes have addressed them, and on how researchers may more effectively use the improved infrastructure for drug overdose studies. • One or more presentations at national meetings and conferences to disseminate information about the project as well as elicit input from the research community on the project tasks throughout its development. • Report on comparative effectiveness research and patient-centered outcomes research on potential questions that could be answered by using existing birth and pregnancy-related data within state and/or federal government. Although opioids and other drugs can be used as the main focus for this purpose, the report can include other important conditions affecting birth outcomes. |

Major Accomplishments

Redesign of the MMDS to incorporate specific drug information captured in the literal text

The previous cause of death coding subsystem MMDS was replaced with a new system called MedCoder. MedCoder is a subsystem of NCHS' NVSS for coding causes of death indicated on the death certificates to ICD-10 codes. It is based on a fusion of new technologies such as machine learning and natural language processing with well-established technologies such as rules-based programming and regression testing. The processing is broken into two steps. First, all causes of death listed on the death certificate are identified and coded. Using this information, the underlying cause is then identified and coded. The multiple cause coding follows these steps.

The pre-processing component contains rules for analyzing the sequence of causes reported on the death certificate. This component is focused on the dictionary and natural language processing and handles spelling and terminology variants. The machine learning model is based on multiple years of coded mortality data where the primary focus is on high frequency causes of death. It produces ICD-10 multiple cause codes for high frequency causes based on the coding rules, as captured in the model. The post-processing component is based on coding rules to produce record-axis codes, which are a processed version of all coded entities reported on the death certificate.

The underlying cause coding system that MedCoder employs is the rule-based subsystem ACME. ACME is a component of the former system, MMDS since no change was desired. See https://www.cdc.gov/nchs/nvss/mmds/about_mmds.htm for a description of ACME and how it works.

MedCoder was put in production on June 6, 2022. All deaths occurring in 2022 and later are being coded using this system. The system is documented here: <https://www.cdc.gov/nchs/nvss/medcoder.htm>.

Extract supplemental drug information contain in the literal text fields and annually produce the NDI and the NVSS-M data files containing the supplemental information for those deaths involving drugs such as opioids

NCHS worked with the NIH's NLM to create a structure and contents of a major drug database. This database was created using the contents from several curated sources: RxNorm ([RxNorm \(nih.gov\)](https://www.nlm.nih.gov/rxnorm/)) for pharmaceutically manufactured drugs and two different Drug Enforcement Agency sources for illicit drugs. A decision to include misspellings as encountered on death certificates kept the data base smaller than autogenerating them. Both synonyms and misspellings were linked to specific drug names so that the data could be summarized appropriately. A process for how to process negation terms and use context of text to code death certificate text was established (e.g., terms like antibiotic-resistant" does not identify antibiotics as a drug associated with the death). Finally, a process for how to maintain the database with newly added drugs from all the curated sources was also written for use by NCHS in the future.

Following the completion of the drug database by NLM, NCHSDVS's Information Technology Branch developed a system which would 1) compare the record literal text with the drug terms in the database, 2) suggest which literal terms in the text were drug and non-drug terms, 3) output these suggestions into lists that NCHS staff reviewed to assess the system's ability to categorize terms, and then 4) created a data visualization dashboard from which NCHS staff could review the quality and categorization of the terms identified in the system.

The new system outcome was compared with the standing system, to make sure the new system was of a high enough quality to deploy. The old system, while producing validated quality coding of drugs, was very time and labor intensive, and could not be run in an automated manner. This had meant that drugs were only evaluated after the final mortality file was complete, nearly a year after the end of the preceding calendar year, or up to almost 2 years after the initial deaths included in the file. Staff determined context and negation rules that had to be added to the system so that terms identified were actual drugs used as part of the cause of death. By the end of the project, the new system was considered of equivalent value.

The process was fully integrated into the cause of death production system as of 2022. This means that identification of drugs associated with deaths is being coded prior to cause of death being assigned so that the drug codes can be used to support ICD-10 code assignment for causes of death. A forensic toxicologist is supporting NCHS to review potential drug terms identified by the system, especially to identify misspellings and synonyms, to update the drug master list with new illicit drugs that are being identified in the US, and to address alignment of drug terms with ICD-10 codes.

The report entitled [National Vital Statistics Reports Volume 68, Number 3 March 21, 2019 Drug Overdose Deaths Involving Fentanyl, 2011-2016 \(cdc.gov\)](https://www.cdc.gov/nchs/data/ndi/20190321_drug-overdose-deaths-involving-fentanyl-2011-2016.pdf) documents the impact of fentanyl on drug overdose deaths. Mortality data previously was limited to ICD-10 codes for analytic work and fentanyl is coded as "synthetic opioids other than methadone." To analyze specific drugs such as fentanyl, NCHS is working towards creating a supplemental drug file with the specific drugs mentioned on the death records to be available as a restricted use data file.

Promote standards-based interoperability between systems containing mortality information at the state level and improve the collection of drug information on death certificates

Interoperability at the state level was a goal that required the development of a data standard, the development of technical support to be provided to states, and then the building, testing and refining processes to achieve success.

Vital records offices in six different vital records jurisdictions were funded to be part of the Implementers' Community that piloted various options to improve their business processes that could increase the timeliness of data submission to NCHS through patient-centered outcomes research (PCOR) funding.

Technical partners supported by NCHS worked with these offices to extend and refine the HL7 Vital Records Death Reporting FHIR Implementation Guide ([VRDR FHIR IG](#)), the standard developed for this data flow. The first version of this standard was balloted in 2019, after which numerous corrections and modifications were made to be in alignment with the needs of the vital records community and NCHS. The VRDR IG release 2.0 was published as the standard for trial use in August 2022 and is the standard being used to implement FHIR-based interoperability between state vital records offices and the National Vital Statistics System at NCHS. A separate Implementation guide describing the transmission method expected using FHIR was also published, the Vital Records FHIR Messaging IG ([VITAL-RECORDS-FHIR-MESSAGING\Home - FHIR v4.0.1](#))

The main goal of the Implementers' Community was piloting the ability to implement data interoperability using FHIR. Representatives of the Implementers' Community attended multiple Connectathons, which are FHIR testing events, over the funding period. During these events, successful transmission of test records sent from jurisdictions' electronic death registration systems to NCHS and coded cause of death and coded race/ethnicity sent back to the jurisdictions for use was achieved. This testing both proved the capacity for jurisdictions and NCHS to use FHIR for our data flows as well as provided the testing needed to validate the FHIR VRDR IG. Furthermore, the data flow process from jurisdictions to NCHS using shared resources with The Association of Public Health Laboratories and CDC has been shown to be able to work effectively.

As NCHS began to scale up this work, their team developed a reference API. This serves an example implementation for vital records jurisdictions that handle submitting death records to NCHS using [VRDR FHIR Messages](#), provides reliable delivery of data (acknowledgements and retries), and retrieving message responses from the [NVSS API server](#). The implementation follows the FHIR Messaging IG. The reference implementation is developed for .NET using C# and leverages the [vrdr-dotnet](#) library to create and parse FHIR Messages. For more information, see [GitHub - nightingaleproject/Reference-Client: A reference implementation for a client application that uses the NVSS FHIR API](#)

Re-architect the provisional data system for VSRR program to produce and release more in-depth information about drug overdose data (e.g., specific drugs, demographic information) on a monthly basis for public health surveillance

The VSRR program supports surveillance of deaths associated with drug overdoses. In 2018, the provisional monthly and quarterly releases provided timely information, but their utility was reduced by

the limited demographic and geographic detail, as well as limited detail on specific drugs involved in the death. Through the funded project, NCHS's DVS restructured the provisional data system for the VSRR to produce and release more in-depth information regarding drug overdose data. This was done by first expanding the surveillance database contents to include detailed age, sex, and state. A statistical report on the timeliness of the surveillance data by sex, age, and state that was published documenting the time period by which 90% of death certificates were received by NCHS, the cut-point for allowing publication ([Vital Statistics Rapid Release, Number 009 \(June 2020\) \(cdc.gov\)](#)). This report noted that for 2017 death, an average of 70.0% of death records for deaths due to drug overdose were available 13 weeks after the date of death, and 93.3% by 26 weeks. This finding allowed VSRR reports to provide estimates on deaths due to drug overdose with a 26-week lag, a 12-week lag reduction.

Since 2018, the Vital Statistics Rapid Release program has expanded the quantity of data products, breadth of topics covered, and timeliness of data delivered. The program previously produced a quarterly dashboard on leading causes of death, a monthly drug overdose deaths counts dashboard, and a report on the timeliness of provisional death certificate data. The program has since added numerous additional data products including daily COVID-19 death counts, 20 weekly data files with demographic data for COVID-19 and overall deaths, numerous ad-hoc data files, quarterly county-level updates for drug overdose deaths, and various provisional reports. Provisional reports have included topics including life expectancy, suicide, COVID-19, and excess deaths. Starting in 2021, provisional data are also available through CDC's Wide-ranging ONline Data for Epidemiologic Research (also known as WONDER) Database. This allows users to query provisional data using custom demographic, geographic, and temporal parameters. Users can query any cause of death by International Classification of Diseases (ICD-10) cause of death code to search for specific causes of death.

Provisional drug overdose surveillance has also expanded since 2018. The monthly drug overdose death counts dashboard was initially limited to drug-specific data from 19 jurisdictions with a 6-month lag. Due to improvements in timeliness and data quality, the dashboard now includes drug-specific data for 48 jurisdictions, with a shortened reporting lag of 4 months. Since 2021, a county-level drug overdose death counts dashboard presents 12-month ending death counts for drug overdose by U.S. county, with quarterly data updates. The VSRR program also produced an exploratory analysis of Early Model-based Provisional Estimates of Drug Overdose Deaths using "nowcasting" methods to account for the normal lag between the occurrence and reporting of these deaths. (Rossen LM, Hedegaard H, Warner M, Ahmad FB, Sutton PD. Early provisional estimates of drug overdose, suicide, and transportation-related deaths: Nowcasting methods to account for reporting lags. Vital Statistics Rapid Release; no 11. Hyattsville, MD: National Center for Health Statistics. February 2021. <https://www.cdc.gov/nchs/data/vsrr/vsrr011-508.pdf>) The VSRR program continues to evaluate timeliness of data and explore methods to make data available to the public in near real time.

[Align and ensure utility of mortality data infrastructure improvements with end-users' needs, especially researchers, with studying drug overdose deaths involving opioids using the NVSS](#)

On November 29, 2018, the NCHS Board of Scientific Counselors (BSC) convened a meeting of the Patient Centered Outcome Research Trust Fund Drug Work Group at the NCHS headquarters in Hyattsville, Maryland. The BSC is charged with providing advice and making recommendations regarding the scientific and technical program goals and objectives of NCHS. The Work Group meeting gathered a

broad range of stakeholders to explore approaches to releasing more detailed drug-related information from death certificates to researchers and public health officials. Meeting participants considered and discussed what supplemental drug information should be extracted from death certificate literal text and how it should be classified. Topics of discussion included (1) gaps in the current drug classification system, (2) current innovations to extract and use information from the literal text, and (3) alternative classification schemes that would provide additional, useful information to end users. In July 2019, NCHS held a second meeting of the PCOR Workgroup of the NCHS BSC. The participants were asked to create recommendations on two topics: How to disseminate data files which include details on specific drug mentions and what future enhancements should be made for these drug files. The meeting generated several useful suggestions on both topics.

The BSC took the findings from the PCOR Work Group and developed a report and set of recommendations for NCHS after each of the two workshops. A summary of the recommendations are available in the minutes from the BSC meetings

(https://www.cdc.gov/nchs/data/bsc/bsc_minutes_january_2019.pdf and https://www.cdc.gov/nchs/data/bsc/bsc_minutes_september_2019.pdf)

DVS undertook two efforts to investigate how the opioid crisis was affecting birth outcomes in the US. The first effort was to identify if data collected on birth certificates could be used to accurately estimate the incidence of neonatal abstinence syndrome (NAS). Three states added this condition to their birth certificate. NCHS worked with those states to option data files linking hospital discharge records to birth certificates for infants identified at discharge with NAS. NCHS found that the states each included a different definition of NAS with their birth certificate instructions. Furthermore, there was variable reporting of NAS on birth certificates. Based on these findings, DVS did not move forward to recommend reporting NAS on birth certificates, but rather worked with the Council of State and Territorial Epidemiologists to create a NAS Standardized Case Definition ([Council of State and Territorial Epidemiologists \(ymaws.com\)](http://www.cste.org/epidemiology/standardized-case-definition))

The second effort looking at birth outcomes related to drug use was a National Vital Statistics Report entitled *Drug-involved Infant Deaths in the United States, 2015-2017* ([National Vital Statistics Reports Volume 70, Number 7 June 3, 2021 Drug-involved Infant Deaths in the United States, 2015–2017 \(cdc.gov\)](https://www.cdc.gov/nchs/data/brs/2021-07-03-Drug-involved-Infant-Deaths-in-the-United-States-2015-2017.pdf)) This report used the Drug Involved Mortality data files and the linked birth/infant death files from 2015-2017. Of the 442 drug-involved infant deaths identified in this time period, drugs were the underlying cause of death for 163 (37%) infants and a contributing cause of death for 279 (63%) infants.

Lessons Learned and Considerations for Future Work

The work conducted for this project provided an opportunity for learning. First, development of the new MedCoder system to automatically code cause of death required a much greater level of effort than expected. Machine learning required the development of a set of death records with which to train the system. While many causes of death can be simply translated into ICD-10 codes, there are nuances that human coders may interpret differently that had to be adjudicated and accurately coded for the machine learning to be conducted.

Secondly, development of a database with which to code drugs listed on death certificates proved much more difficult than anticipated. While there are several curated databases such as RXNorm by which to code pharmaceutically manufactured drugs, illicit drugs have no equivalent reference system. This

resulted in NCHS/DVS having to create its own system which references other systems but cannot be considered a definitive source of information, especially due to the need to include common misspellings of drugs and the quick pace of new illicit drugs development.

Finally, this project provided the launch for the use of HL7 FHIR in the exchange of mortality records. During this work, the initial HL7 FHIR Implementation Guide for Vital Records Death Reporting and the state and national APIs were created. The vital records office jurisdictions that worked in the initial experimentation phase worked through many different issues to create standards which other states are now following. The first set of states to use FHIR to send data to NCHS as their production system will be certified to do so in early 2023, with the expectation that the majority will follow by mid-2024.