National Vital Statistics Reports

Volume 65, Number 9 December 20, 2016

Using Literal Text From the Death Certificate to Enhance Mortality Statistics: Characterizing Drug Involvement in Deaths

by James P. Trinidad, M.P.H., M.S., U.S. Food and Drug Administration; Margaret Warner, Ph.D., Brigham A. Bastian, B.S., Arialdi M. Miniño, M.P.H., and Holly Hedegaard, M.D., M.S.P.H., National Center for Health Statistics

Abstract

Objectives—This report describes the development and use of a method for analyzing the literal text from death certificates to enhance national mortality statistics on drug-involved deaths. Drug-involved deaths include drug overdose deaths as well as other deaths where, according to death certificate literal text, drugs were associated with or contributed to the death.

Methods—The method uses final National Vital Statistics System–Mortality files linked to electronic files containing literal text information from death certificates. Software programs were designed to search the literal text from three fields of the death certificate (the cause of death from Part I, significant conditions contributing to the death from Part II, and a description of how the injury occurred from Box 43) to identify drug mentions as well as contextual information. The list of drug search terms was developed from existing drug classification systems as well as from manual review of the literal text. Literal text surrounding the identified drug search terms was analyzed to ascertain the context. Drugs mentioned in the death certificate literal text were assumed to be involved in the death unless contextual information suggested otherwise (e.g., "METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS INFECTION"). The literal text analysis method was assessed by comparing the results from application of the method with results based on ICD–10 codes, and by conducting a manual review of a sample of records.

Keywords: text analysis • drug-involved death • drug overdose • National Vital Statistics System

Introduction

Recent mortality trends in the United States show a substantial increase in the rate of drug overdose deaths. From 2000 to 2014, the mortality rate for drug overdose more than doubled from 6.2 to 14.7 per 100,000 population (1). To address this public health concern, many researchers use National Vital Statistics System mortality data (NVSS–M) to describe these trends and to monitor the populations most at risk (1–4).

The NVSS–M data are based on information from the death certificates filed in the 50 states and the District of Columbia. The data set includes cause-of-death, demographic, and geographic information extracted from death certificates for all decedents in the United States (5). The NVSS–M data are coded using a standardized classification system, the *International Classification of Diseases and Related Health Problems, Tenth Revision* (ICD–10) (6). While this classification system allows for consistency in identifying the underlying and contributory causes of death, there are limitations in the use of ICD–10-coded data to study drug-involved mortality. Specifically, in the ICD–10 classification system, only a few drugs (e.g., heroin, methadone, and cocaine) are assigned a unique classification code (T40.1, T40.3, and T40.5, respectively) under certain circumstances (e.g., when the death is an overdose). Most drugs, however, are assigned to broad categories (e.g., both oxycodone and morphine are categorized to T40.2, Poisoning: Other opioids) (7). The use of broad categories in ICD–10 makes it difficult to use ICD–10 coded data to monitor trends in deaths involving specific drugs that are not already uniquely classified in ICD–10.

Analysis of literal text has been used to enhance mortality statistics in investigations of sudden infant death syndrome, Creutzfeldt-Jakob disease, influenza and pneumonia, cancer, and drug poisonings (8–13). The literal text often includes information beyond the general classification captured in an ICD–10 code description. For example, researchers have examined the literal

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics
National Vital Statistics System

text to better understand the circumstances (e.g., unsafe sleep environments) contributing to sudden infant death syndrome (9,12). Literal text can also be analyzed to identify a specific subset of deaths coded to a broad ICD–10 classification. For example, reseachers have examined the literal text to identify deaths from Creutzfeldt-Jakob disease among decedents with an ICD–10 underlying cause of death of B94.8, Sequelae of other specified infectious diseases (13). Similarly, researchers have explored literal text analysis methods to better understand the contribution of specific drugs in drug-poisoning deaths, and found that the literal text data provided more information on specific drugs than the ICD–10-coded data (11). These previous literal text analyses involving information on specific drugs did not consider literal text information other than drug mentions, were limited by causes of death, and assessed only records from a single state. Further development and use of literal text analysis methodology provides an opportunity for an enhanced understanding of the national picture of drug involvement in deaths in the United States (11).

This report describes the collaborative efforts of the National Center for Health Statistics (NCHS) and the U.S. Food and Drug Administration (FDA) to develop and assess a method for using literal text from death certificates to identify specific drugs involved in deaths, that is, drug overdose deaths and deaths with other types of drug involvement. This report accompanies a study that highlights the specific drugs most frequently involved in drug overdose deaths from 2010 through 2014 (14).

Methods Development

Overview

The analysis method uses search terms to identify drugs mentioned in electronic death certificate literal text (i.e., the cause-of-death statements on the death certificate). Unless contextual information suggested otherwise, drugs mentioned in the death certificate literal text were assumed to be involved in the death. Therefore, the method also analyzes literal text surrounding the identified search terms to determine whether the drugs mentioned were not involved in death (e.g., "METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS INFECTION"). The processed data resulting from applying the method includes all identified drug mentions and contextual information on drug involvement.

The following sections describe the data source for the literal text analysis methodology; some issues considered during the methods development; an assessment of the quality of the literal text data; the approach used to optimize the efficiency of literal text analysis; the development of lists of terms and phrases that were used in the processing of literal text; the steps of the literal text analysis methodology; and the data produced by applying the literal text analysis methodology.

Data source

The literal text analysis methodology was developed using final NVSS–M data linked to literal text data. Both NVSS–M data and literal text data are derived from information on death certificates (5).

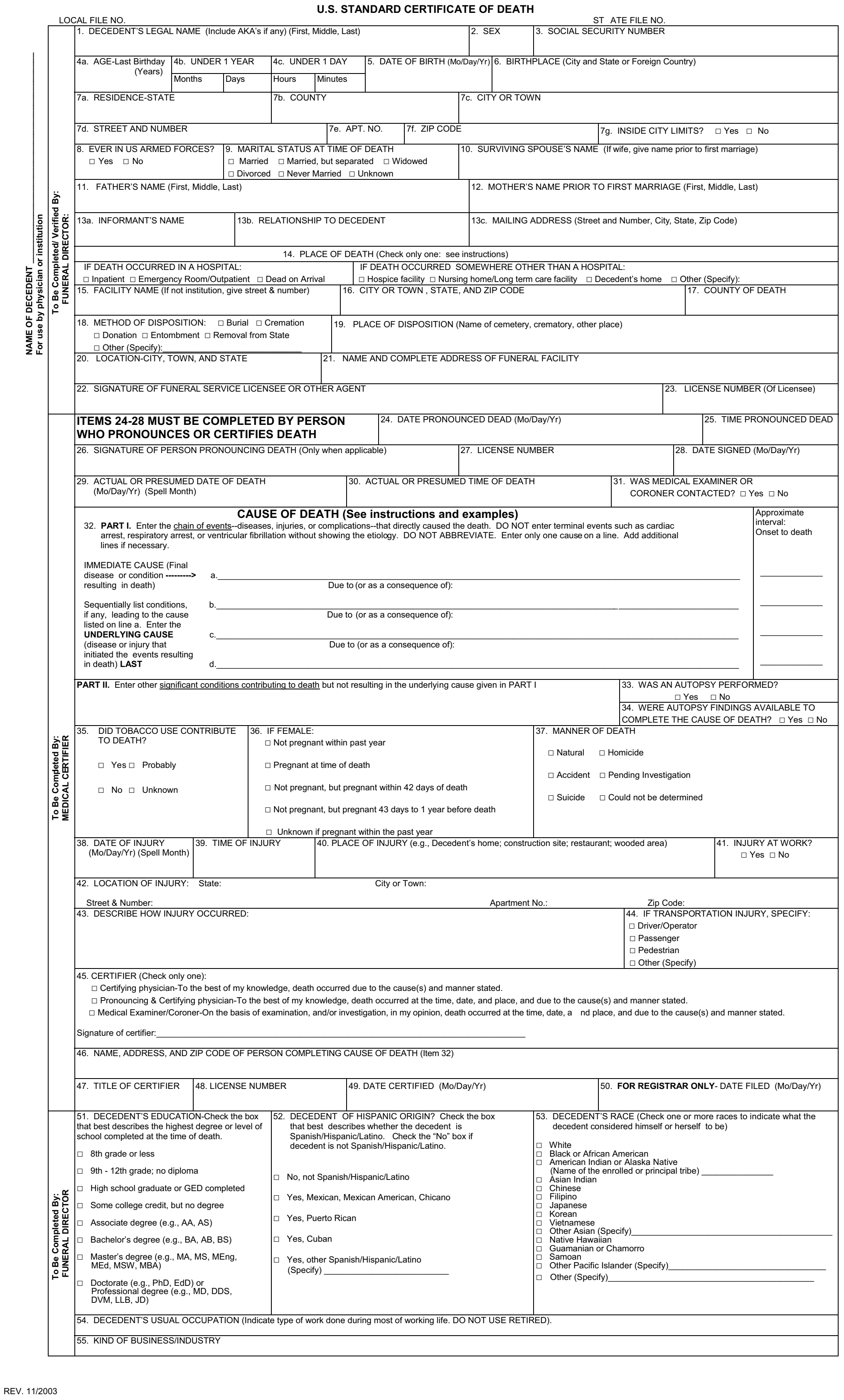
In NVSS–M, the coded causes of death are assigned based on information written in the cause-of-death section on the death certificate (Figure 1). The information written on the death certificate by the medical certifier on the cause, manner, circumstances, and other factors contributing to the death is referred to as the literal text fields. The literal text fields of the cause-of-death section on the U.S. Standard Certificate of Death (15,16) include:

- The chain of events leading to death (from Part I)
- Other significant conditions that contributed to the death (from Part II)
- How the injury occurred (in the case of deaths due to injuries [from Box 43])

NCHS uses a software program to code the literal text from the death certificate according to the rules of ICD–10 (17). These processes involve the identification of statements from death certificate literal text, such as "MYOCARDIAL INFARCTION" and "DIABETES MELLITUS." Some statements, such as "METHADONE INTOXICATION," refer to drug-involved mortality. The identified statements are translated into ICD–10 codes. For example, the identified statement "OXYCODONE POISONING" is coded to ICD–10 codes T40.2, Poisoning: other opioids, and X42, Accidental poisoning by and exposure to narcotics and psychodysleptics (hallucinogens), not elsewhere classified. Note that throughout this report, text from death certificates is indicated in quotes and uppercase letters.

ICD–10 codes reflect the conditions reported on the death certificate. During the coding process, the software program assigns ICD–10 codes to 1 underlying cause and up to 20 multiple causes of death. Records rejected by the software program are reviewed by trained nosologists, and ICD–10 codes are manually assigned. In general, nosologists manually code about one-fifth of the death records. For deaths with an underlying cause of drug overdose (deaths with an underlying cause code of X40–X44, X60–X64, X85, or Y10–Y14), about two-thirds are coded manually (18). Entity axis ICD–10 codes include the ICD–10 code and information on the placement of the coded condition on the death certificate.

NCHS maintains the coded NVSS–M final mortality file and the literal text data separately, and linkage between the NVSS–M and literal text data leverages the information from both data sets. To link the data, NVSS–M and literal text files were merged on year of death, state of occurrence, and death certificate number.



U.S. STANDARD CERTIFICATE OF DEATH

LOCAL FILE NO. ST ATE FILE NO.

NAME OF DECEDENT ____________ For use by physician or institution

To Be Completed/ Verified By: FUNERAL DIRECTOR:

1. DECEDENT'S LEGAL NAME (Include AKA's if any) (First, Middle, Last)
2. SEX
3. SOCIAL SECURITY NUMBER

4a. AGE-Last Birthday (Years)
4b. UNDER 1 YEAR Months Days
4c. UNDER 1 DAY Hours Minutes
5. DATE OF BIRTH (Mo/Day/Yr)
6. BIRTHPLACE (City and State or Foreign Country)

7a. RESIDENCE-STATE
7b. COUNTY
7c. CITY OR TOWN

7d. STREET AND NUMBER
7e. APT. NO.
7f. ZIP CODE
7g. INSIDE CITY LIMITS? □ Yes □ No

8. EVER IN US ARMED FORCES? □ Yes □ No
9. MARITAL STATUS AT TIME OF DEATH □ Married □ Married, but separated □ Widowed □ Divorced □ Never Married □ Unknown
10. SURVIVING SPOUSE'S NAME (If wife, give name prior to first marriage)

11. FATHER'S NAME (First, Middle, Last)
12. MOTHER'S NAME PRIOR TO FIRST MARRIAGE (First, Middle, Last)

13a. INFORMANT'S NAME
13b. RELATIONSHIP TO DECEDENT
13c. MAILING ADDRESS (Street and Number, City, State, Zip Code)

14. PLACE OF DEATH (Check only one: see instructions)

IF DEATH OCCURRED IN A HOSPITAL:
□ Inpatient □ Emergency Room/Outpatient □ Dead on Arrival

IF DEATH OCCURRED SOMEWHERE OTHER THAN A HOSPITAL:
□ Hospice facility □ Nursing home/Long term care facility □ Decedent's home □ Other (Specify):

15. FACILITY NAME (If not institution, give street & number)
16. CITY OR TOWN , STATE, AND ZIP CODE
17. COUNTY OF DEATH

18. METHOD OF DISPOSITION: □ Burial □ Cremation □ Donation □ Entombment □ Removal from State □ Other (Specify):___________________________
19. PLACE OF DISPOSITION (Name of cemetery, crematory, other place)

20. LOCATION-CITY, TOWN, AND STATE
21. NAME AND COMPLETE ADDRESS OF FUNERAL FACILITY

22. SIGNATURE OF FUNERAL SERVICE LICENSEE OR OTHER AGENT
23. LICENSE NUMBER (Of Licensee)

To Be Completed By: MEDICAL CERTIFIER

ITEMS 24-28 MUST BE COMPLETED BY PERSON WHO PRONOUNCES OR CERTIFIES DEATH
24. DATE PRONOUNCED DEAD (Mo/Day/Yr)
25. TIME PRONOUNCED DEAD

26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable)
27. LICENSE NUMBER
28. DATE SIGNED (Mo/Day/Yr)

29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month)
30. ACTUAL OR PRESUMED TIME OF DEATH
31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? □ Yes □ No

CAUSE OF DEATH (See instructions and examples)

32. **PART I.** Enter the chain of events--diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.

Approximate interval: Onset to death

IMMEDIATE CAUSE (Final disease or condition ---------> resulting in death) a.____ ____________
Due to (or as a consequence of):

Sequentially list conditions, if any, leading to the cause listed on line a. Enter the **UNDERLYING CAUSE** (disease or injury that initiated the events resulting in death) **LAST**

b.____ ____________
Due to (or as a consequence of):

c.____ ____________
Due to (or as a consequence of):

d.____ ____________

PART II. Enter other significant conditions contributing to death but not resulting in the underlying cause given in PART I

33. WAS AN AUTOPSY PERFORMED? □ Yes □ No
34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? □ Yes □ No

35. DID TOBACCO USE CONTRIBUTE TO DEATH?
□ Yes □ Probably
□ No □ Unknown

36. IF FEMALE:
□ Not pregnant within past year
□ Pregnant at time of death
□ Not pregnant, but pregnant within 42 days of death
□ Not pregnant, but pregnant 43 days to 1 year before death
□ Unknown if pregnant within the past year

37. MANNER OF DEATH
□ Natural □ Homicide
□ Accident □ Pending Investigation
□ Suicide □ Could not be determined

38. DATE OF INJURY (Mo/Day/Yr) (Spell Month)
39. TIME OF INJURY
40. PLACE OF INJURY (e.g., Decedent's home; construction site; restaurant; wooded area)
41. INJURY AT WORK? □ Yes □ No

42. LOCATION OF INJURY: State: City or Town:
Street & Number: Apartment No.: Zip Code:

43. DESCRIBE HOW INJURY OCCURRED:
44. IF TRANSPORTATION INJURY, SPECIFY:
□ Driver/Operator
□ Passenger
□ Pedestrian
□ Other (Specify)

45. CERTIFIER (Check only one):
□ Certifying physician-To the best of my knowledge, death occurred due to the cause(s) and manner stated.
□ Pronouncing & Certifying physician-To the best of my knowledge, death occurred at the time, date, and place, and due to the cause(s) and manner stated.
□ Medical Examiner/Coroner-On the basis of examination, and/or investigation, in my opinion, death occurred at the time, date, a nd place, and due to the cause(s) and manner stated.

Signature of certifier:____

46. NAME, ADDRESS, AND ZIP CODE OF PERSON COMPLETING CAUSE OF DEATH (Item 32)

47. TITLE OF CERTIFIER
48. LICENSE NUMBER
49. DATE CERTIFIED (Mo/Day/Yr)
50. **FOR REGISTRAR ONLY**- DATE FILED (Mo/Day/Yr)

To Be Completed By: FUNERAL DIRECTOR

51. DECEDENT'S EDUCATION-Check the box that best describes the highest degree or level of school completed at the time of death.
□ 8th grade or less
□ 9th - 12th grade; no diploma
□ High school graduate or GED completed
□ Some college credit, but no degree
□ Associate degree (e.g., AA, AS)
□ Bachelor's degree (e.g., BA, AB, BS)
□ Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)
□ Doctorate (e.g., PhD, EdD) or Professional degree (e.g., MD, DDS, DVM, LLB, JD)

52. DECEDENT OF HISPANIC ORIGIN? Check the box that best describes whether the decedent is Spanish/Hispanic/Latino. Check the "No" box if decedent is not Spanish/Hispanic/Latino.
□ No, not Spanish/Hispanic/Latino
□ Yes, Mexican, Mexican American, Chicano
□ Yes, Puerto Rican
□ Yes, Cuban
□ Yes, other Spanish/Hispanic/Latino (Specify) __________________________

53. DECEDENT'S RACE (Check one or more races to indicate what the decedent considered himself or herself to be)
□ White
□ Black or African American
□ American Indian or Alaska Native (Name of the enrolled or principal tribe) _______________
□ Asian Indian
□ Chinese
□ Filipino
□ Japanese
□ Korean
□ Vietnamese
□ Other Asian (Specify)_________________________________
□ Native Hawaiian
□ Guamanian or Chamorro
□ Samoan
□ Other Pacific Islander (Specify)_______________________________
□ Other (Specify)________________________________

54. DECEDENT'S USUAL OCCUPATION (Indicate type of work done during most of working life. DO NOT USE RETIRED).

55. KIND OF BUSINESS/INDUSTRY

REV. 11/2003

SOURCE: NCHS, National Vital Statistics System.

Figure 1. U.S. standard death certificate

Considerations in developing methods to process death certificate literal text

In developing the analysis methodology, several characteristics and limitations of the literal text needed to be considered.

Availability of literal text information–Deaths may have no literal text data or only literal text mentions regarding the status of the death investigation (e.g., mentions of "PENDING" or "UNDER INVESTIGATION"). For these deaths, there are no mentions of drugs in the literal text.

Syntax of literal text–The syntax of the death certificate literal text generally consists of a few words or simple phrases (e.g., "DRUG TOXICITY") rather than clauses or sentences (e.g., "DECEDENT DIED OF DRUG POISONING"). The literal text analysis methods were developed by imitating the software program and processes that extract and assign ICD–10 codes to the literal text information as described above. These processes identify statements in the text.

Four text fields in Part I–The text fields constituting Part I of the death certificate have an assumed interpretation: The cause of death listed in the first text field is due to (or a consequence of) the cause of death (if any) listed in the second text field, which is due to (or a consequence of) the cause of death (if any) listed in the third text field, which is due to (or a consequence of) the cause of death (if any) listed in the fourth text field. The first cause of death listed in this sequence is the immediate cause of death, and the cause of death on the lowest-used line in Part I is the underlying cause of death. The assumed interpretation works well for some deaths. However, the assumption does not work well for other deaths. For example, medical certifiers may list multiple causes of death on a single line, may list a single cause of death on multiple lines, or may not write the causes in the appropriate sequential order. To simplify analyses, the assumed interpretation in Part I was ignored, and the text fields constituting Part I were concatenated as a single text field.

Case, symbols, and numbers–Use of uppercase and lowercase characters, symbols, and numbers varies across deaths. Some death certificate literal text may be in uppercase only, others in lowercase only, and others in a mixture of uppercase and lowercase. Literal text may contain symbols, such as hyphens. While the names of some drugs have hyphens (e.g., "GAMMA-HYDROXYBUTYRIC ACID"), use of hyphens in drug names can vary across death certificates, which complicates the identification of mentions of these drugs in literal text. Drug names (particularly generic drug names) generally do not include numbers, although numbers may be informative in clarifying the extent of drug exposure (such as in the phrase "BLOOD LEVEL ≥ 20 MG/DL"). To simplify analyses, all text was converted to uppercase, and symbols and numbers were removed.

Specificity of drug information–The specificity of drug information varies across death certificates. Death certificates may have mentions of specific drugs in the literal text (e.g., "OXYCODONE" or "FENTANYL"), mentions of drug classes (e.g., "OPIOID"), or exposures not otherwise specified (NOS) (e.g., "DRUG," "CHEMICAL," or "POLYPHARMACY"). Death certificates may have a mixture of mentions of specific drugs, drug classes, and exposures NOS. When a specific drug is mentioned alongside mentions of drug classes or exposures NOS, the mentions are sometimes referential (e.g., heroin is assumed to be the opioid in the phrase "OPIOID (HEROIN) OVERDOSE").

Synonyms–A specific drug may be referenced by various terms that are synonymous. For example, acetaminophen (generic name), paracetamol (generic name), and APAP (abbreviation) all refer to the same drug. When referring to a single-ingredient product, Tylenol (brand name) is also synonymous with acetaminophen. Brand names can refer to products with one or more drug ingredients. Literal text can have plural forms of drug mentions (e.g., mentions of "DRUGS," the plural form of drug). Literal text can also include misspellings. While drug metabolites are not synonymous with the parent drug products, drug metabolites may appear in the literal text, and are assumed to be the same. For example, a literal text mention of a toxicological finding of desmethyldiazepam (metabolite) would indicate exposure to diazepam (the parent drug).

Contextual information–Mentions of drugs are often accompanied by contextual information, which are other words in the literal text that either describe the drug(s) or provide information on how it was involved in mortality, if at all. The words in proximity to the drug mentions provide more informative contextual information than words that are distant.

Contextual information can provide details on drug characteristics or characteristics of drug exposure, such as the number of drugs (e.g., "MULTIPLE DRUGS"), extent of drug exposure (e.g., "FATAL LEVEL OF DRUG" or "THERAPEUTIC AMOUNT OF DRUG"), drug formulation (e.g., "DRUG TABLET"), the type of drug (e.g., "ILLICIT DRUG" or "DRUGS WHICH WERE PRESCRIBED"), and possession or ownership of the drug (e.g., "HIS DRUG" or "HER DRUG"). These descriptions can be complex and use conjunctions, such as the word "AND" (e.g., "FATAL LEVEL OF PRESCRIPTION DRUGS ILLEGALLY OBTAINED" and "ILLEGAL AND PRESCRIPTION DRUGS").

Contextual information can also explicitly describe how the drug exposure was involved in the death (e.g., "HEROIN POISONING" and "ANAPHYLAXIS DUE TO ANTIBIOTIC") or other aspects of drug involvement. Other aspects of drug involvement include route of administration (e.g., "DRUG INJECTION"), medical history with drug exposure (e.g., "HISTORY OF DRUG ABUSE" or "THERAPEUTIC USE OF METHADONE"), and other complications with drug exposure (e.g., "DRUG-DRUG INTERACTION"). Contextual information can also indicate drug exposure, either explicitly (e.g., "USE OF DRUG") or implicitly (e.g., "DRUG BLOOD LEVEL 20 MG/DL").

The contextual information can also be used to determine whether the drug mentioned in the literal text was not involved in mortality. For example, the drug "METHICILLIN" in the phrase "METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS INFECTION" does not suggest drug involvement in mortality, but rather a type of bacterial infection. Similarly, the phrase "NOT DRUG RELATED" clearly indicates that a death did not involve drugs. This report distinguishes between a drug mention, a drug mentioned with involvement (DMI), and a DMI death.

- A drug mention is any mention of a drug, a drug class, or exposure NOS in the literal text fields.
- A DMI is defined as a mention of a drug, a drug class, or exposure NOS in the literal text fields, excluding mentions where the contextual information suggested that the drug was not involved in the death.
- A DMI death is defined as a death having at least one DMI.

Information in the literal text can be contextual in that it provides information about drug characteristics or characteristics of drug exposure (i.e., descriptors), or contextual in that it describes whether and how a drug was involved in mortality (i.e., contextual phrases). Although descriptors provide some detail about the drugs mentioned in the literal text, they provide little or no information about drug involvement and, therefore, for the purposes of developing the literal text methodology, are less important than contextual phrases.

Multiple drugs–Deaths may involve multiple drugs. Medical certifiers may list these drugs consecutively, but not necessarily in order of importance to the cause of death (e.g., alphabetical order). These sequential drug mentions may be written with conjunctions, such as the word "AND" in the phrase "METHICILLIN AND VANCOMYCIN." Other sequential drug mentions do not contain conjunctions, such as in the phrase "OVERDOSE (HEROIN, COCAINE)."

While keyword searches can be performed to identify drug mentions, keyword searches are not efficient in identifying the contextual information associated with each drug mention. This is because the same contextual information may relate to more than one drug. For example, an infection that is resistant to both methicillin and vancomycin is inferred in the phrase "METHICILLIN AND VANCOMYCIN RESISTANT INFECTION." In the example, a search for "METHICILLIN RESISTANT INFECTION" would not identify the mention of methicillin, and a search for "METHICILLIN" would fail to identify that methicillin was not involved in mortality.

Searching for statements that incorporate both drug mentions and contextual information (e.g., searching for the statement "METHICILLIN AND VANCOMYCIN RESISTANT INFECTION") is the most direct approach for simultaneously identifying drug mentions and associated contextual information. However, this approach would require a vast number of statements due to the large number of drugs that can be mentioned, variability in the order of the drug mentions, and variability in the contextual information. In summary, there is an inexhaustible variety of combinations of statements consisting of drug mentions and contextual information.

Assessment of the presence of uninformative literal text

The quality of the literal text and its potential utility in identifying drug mentions was assessed by determining the percentage of records with no information that could be used to assign the cause(s) of death. The literal text was considered uninformative if: 1) there was no text in any of the literal text fields (i.e., the fields were blank) or 2) the fields only contained descriptive words or phrases about the status of the investigation (e.g., mentions of "PENDING" or "UNDER INVESTIGATION"). In most cases, when all the literal text is uninformative, an underlying cause-of-death of ICD–10 code R99 (Other ill-defined and unspecified causes of mortality) is assigned. Figure 2 contains all the terms considered to be uninformative for the purposes of identifying drug mentions.

Among NVSS–M records merged with literal text data for year 2013 (the most recent year of data at the time of the assessment), a small minority (less than 1%) had blank or uninformative literal text (Table A). Most of these were assigned an underlying cause-of-death code R99 (Other ill-defined and unspecified causes of mortality). Therefore, a small minority (less than 1%) of all records have literal text fields and ICD–10 coding that provide no information on specific causes of death.

Exchangeability: Optimizing efficiency of processing literal text information

Manual review of the literal text revealed that drug mentions are exchangeable (i.e., conceptually similar) when contextual information is fixed. For example, the word "HEROIN" in the phrase "HEROIN OVERDOSE" could be replaced (i.e., exchanged) with the word "OPIOID," with no change in the broad interpretation of the literal text (i.e., the cause of death was a drug overdose). Combinations of drug mentions are also exchangeable. For example, "METHICILLIN AND VANCOMYCIN" is exchangeable with the word "ANTIBIOTIC" in the phrase "ANTIBIOTIC RESISTANT INFECTION." Descriptors are also exchangeable. For example, the word "RX" can replace the descriptor "MULTIPLE PRESCRIPTION" in the phrase "MULTIPLE PRESCRIPTION DRUGS."

CAUSE UNDER INVESTIGATION, DEFERRED, PENDING, PENDING ADDITIONAL STUDIES, PENDING ADDITIONAL STUDY, PENDING AUTOPSY, PENDING AUTOPSY AND HISTOLOGY, PENDING AUTOPSY AND TOXICOLOGY, PENDING AUTOPSY HISTOLOGY, PENDING AUTOPSY TOXICOLOGY, PENDING AUTOPSY FINDING, PENDING AUTOPSY FINDINGS, PENDING AUTOPSY STUDIES, PENDING AUTOPSY STUDY, PENDING FURTHER INVESTIGATION, PENDING FURTHER STUDIES, PENDING FURTHER STUDY, PENDING HISTOLOGY, PENDING HISTOLOGY AND AUTOPSY, PENDING HISTOLOGY AND TOXICOLOGY, PENDING HISTOLOGY AUTOPSY, PENDING HISTOLOGY STUDIES, PENDING HISTOLOGY STUDY, PENDING HISTOLOGY TOXICOLOGY, PENDING LABORATORY STUDIES, PENDING LABORATORY STUDY, PENDING INVESTIGATION, PENDING TOXICOLOGY, PENDING TOXICOLOGY AND AUTOPSY, PENDING TOXICOLOGY AND HISTOLOGY, PENDING TOXICOLOGY AUTOPSY, PENDING TOXICOLOGY HISTOLOGY, PENDING TOXICOLOGY STUDIES, PENDING TOXICOLOGY STUDY, PENDING STUDIES, PENDING STUDY, PENDING TOX UNDER INVESTIGATION.

SOURCE: NCHS, National Vital Statistics System, death certificate literal text.

Figure 2. Literal text strings considered uninformative for assigning cause of death and identifying drug mentions

Table A. Deaths having no informative literal text on cause of death: U.S. residents, 2013

Characteristics	Number of deaths	Percent of deaths
All deaths	2,596,993	100.00
Deaths having no informative literal text	3,831	0.15
Deaths having no informative literal text and ICD–10 code R99 as underlying cause of death[1]	3,421	0.13

[1] The ICD–10 code R99 indicates Other ill-defined and unspecified causes of mortality.

NOTE: ICD–10 is the *International Classification of Diseases and Related Health Problems, Tenth Revision.*

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

The exchangeability of drug mentions enables the DMI programs to more efficiently process data on drug mentions and their associated contextual information. For example, replacing sequential drug mentions identified in the literal text (e.g., “METHICILLIN AND VANCOMYCIN” or “VANCOMYCIN AND METHICILLIN”) with the word “DRUG” greatly simplifies the processing steps for the DMI program.

A stepwise approach was used to enhance the DMI program efficiency in extracting information from the literal text. This stepwise approach leverages the exchangeability of drug mentions and the exchangeability of descriptors. In other words, contextual information on drug involvement can be most efficiently identified and processed using the computer algorithms when the variability in drug mentions and associated descriptors is reduced. This stepwise approach required the development of lists of drugs, descriptors, and joining phrases that link search terms or descriptors together, and the development of contextual phrases.

Developing a search term list for drugs

A list of search terms was developed to identify drug mentions. This list was developed using a two-phase approach. The final list of search terms included single words (e.g., “HEROIN”) and combinations of words (e.g., “CRACK COCAINE”) for specific drugs, drug classes, and drug exposures NOS.

In the first phase, the search term list was constructed from single-word generic names listed in the Substance Abuse and Mental Health Services Administration’s (SAMHSA) Drug Abuse Warning Network (DAWN) Drug Reference Vocabulary (DRV), published in 2012 (19). DAWN DRV is a drug vocabulary and classification system based on the Multum Lexicon database from Cerner Multum, Inc. Its structure is hierarchical with generic drugs categorized under higher-level groupings (e.g., drug class). For use with the DAWN data system, SAMHSA added substances that are misused and abused (e.g., illicit drugs and inhalants) that were not included in the Multum Lexicon database.

During this first phase of generating the search term lists, the following DAWN DRV categories were excluded: major substances of abuse, nutritional products, alternative medicines, medical gases, biologicals, immune globulins, immunostimulants, sterile irrigating solutions, and drugs unknown. Products in these categories had generic names that were difficult to condense into a single word denoting a drug product. The list also excluded combination products, nearly all of which could be identified by their components. The search term list that resulted from the first phase of development did not include names of drug classes or drug exposures NOS.

In the second phase, the search term list was expanded by adding terms for specific drugs not identified in the first phase, including illicit drugs; drug classes; drug exposures NOS; terms containing more than one word; brand names; and obvious, frequently occurring misspellings. Most of the search terms added during the second phase were identified through nonsystematic manual reviews and queries of the 2003–2014 literal text.

Methods development was focused on literal text data from 2007, the first year of literal text data that was available during methods development, and from 2013, the most recent year of data available at the time when assessments were conducted. Additional search terms for brand names of prescription drugs were identified using the Drugs@FDA website, and search terms for misspelled drugs were identified using FDA Adverse Event Reporting System data (20). A few search terms were also identified using other approaches, including comparison with ICD–10 codes.

The search term list that resulted from the second phase excluded foods and food additives (e.g., starch), excipients, gases (e.g., helium and carbon monoxide), airborne contaminants (e.g., soot), industrial chemicals (e.g., ethylene glycol), periodic table elements (e.g., lithium and iodine), and substances with unknown industrial or pharmaceutical applications. Although therapeutic uses of some of these substances is possible, these substances were not included because it proved difficult to determine whether the exposures to the substances were therapeutic, for misuse or abuse, or environmental.

Study team members trained in pharmacy and pharmacoepidemiology categorized search terms by various characteristics, including whether the terms referred to specific drugs, drug classes, or exposures NOS. Search terms were also classified by whether they represented generic drug names or other variants, such as brand names, common use or street names, abbreviations, metabolites, and misspellings. Most search terms were mapped to a single “principal variant,” the overarching label assigned to a drug, a drug class, or exposure NOS. In general, the principal variant was the generic drug name. Some search terms—mostly for combination drug products—were mapped to two or more principal variants. The use of principal variants made it possible to identify all deaths that involved the same drug.

The development of the search term list involved various efforts to create a comprehensive list of all drugs mentioned in literal text. Although many methods were used to develop the list, the list might not contain all possible search terms for all possible drugs. The assessments conducted during methods development were based on a June 2015 list of 2,865 search terms representing 1,649 principal variants (see Table I–1). This list was updated in November 2015 to include 3,116 search terms representing 1,643 principal variants (see Table I–2).

Developing lists of contextual information

Three lists of contextual information were developed using iterative manual reviews and queries of literal text for data years 2003 through 2014. The three lists consisted of descriptors, joining phrases, and contextual phrases.

The list of descriptors included a word or words that provide information on drug characteristics or characteristics of drug exposure, such as "MULTIPLE," "PRESCRIPTION," and "NON PRESCRIPTION." The list classified whether the descriptor should be identified before a drug mention, after a drug mention, or either before or after a drug mention (as would be the case for the descriptor "PRESCRIPTION" in the phrases "PRESCRIPTION DRUG" and "DRUG PRESCRIPTION"). The list also classified the descriptors by the characteristic(s) they aim to describe (e.g., "TABLET" and "TRANSDERMAL" describe type of drug formulation).

The list of joining phrases included words and asterisks that acted as conjunctions. For this list, each joining phrase was comprised of 1) two asterisks that indicate exchangeability of either drug mentions or descriptors and 2) potentially other words that indicate linkage. Examples of words that indicate linkage include "AND" and "AS WELL AS." Bookending these words were asterisks, as in the case of the joining phrases "* AND *" and "* AS WELL AS *." These asterisks were exchangeable with drug mentions or descriptors, as in the phrases "METHICILLIN AND VANCOMYCIN" and "ILLICIT AS WELL AS PRESCRIPTION." The simplest joining phrase was "* *," indicating two adjacent drug mentions or two adjacent descriptors.

The list of contextual phrases included words and asterisks that, altogether, describe drug involvement (if any). Examples of contextual phrases include "* TOXICITY" and "ABUSED *." Like the asterisks in joining phrases, asterisks in contextual phrases indicate exchangeability of mentions. However, while the asterisks in joining phrases refer to either drug mentions or descriptors, the asterisks in contextual phrases simultaneously refer to drug mentions, any associated descriptors, and joining phrases. In addition, while there are only two asterisks in joining phrases, contextual phrases may have one or more asterisks, as in the case of "ACCIDENTAL * TOXICITY WITH *," which could refer to the phrase "ACCIDENTAL DRUG TOXICITY WITH HEROIN AND OTHER ILLICIT DRUGS." The simplest contextual phrase was "*," indicating the mention of one or more drugs and associated descriptors, but no other contextual information.

Study team members classified the contextual phrases by various characteristics. The most important characteristic was whether the contextual phrase did not suggest drug involvement. Contextual phrases that suggested no drug involvement generally referred to health conditions or disease states. For example, when the word "INSULIN" replaces "*" in the contextual phrase "* DEPENDENT DIABETES," the resulting text refers to a health condition. Similarly, when the word "METHICILLIN" replaces "*" in the contextual phrase "* RESISTANT STAPHYLOCOCCUS AUREUS INFECTION," the resulting text refers to a type of bacterial infection. Other contextual phrases clearly indicated no drug involvement, which would be the case for the contextual phrase "NO * INVOLVED."

The drugs mentioned in the death certificate literal text were assumed to be involved in the death unless contextual information suggested otherwise.

Contextual phrases that described similar ideas (such as "* TOXICITY" and "TOXICITY FROM *") were classified under a common category. Some phrases were classified under more than one category; for example, "TOXICITY FROM * INJECTION" was classified under the category for toxicity and the category for injection.

The assessments conducted during methods development were based on 527 descriptors, 22 joining phrases, and 1,641 contextual phrases that were listed as of June 2015. These lists were updated in November 2015.

Identifying mentions of drugs and ascribing context

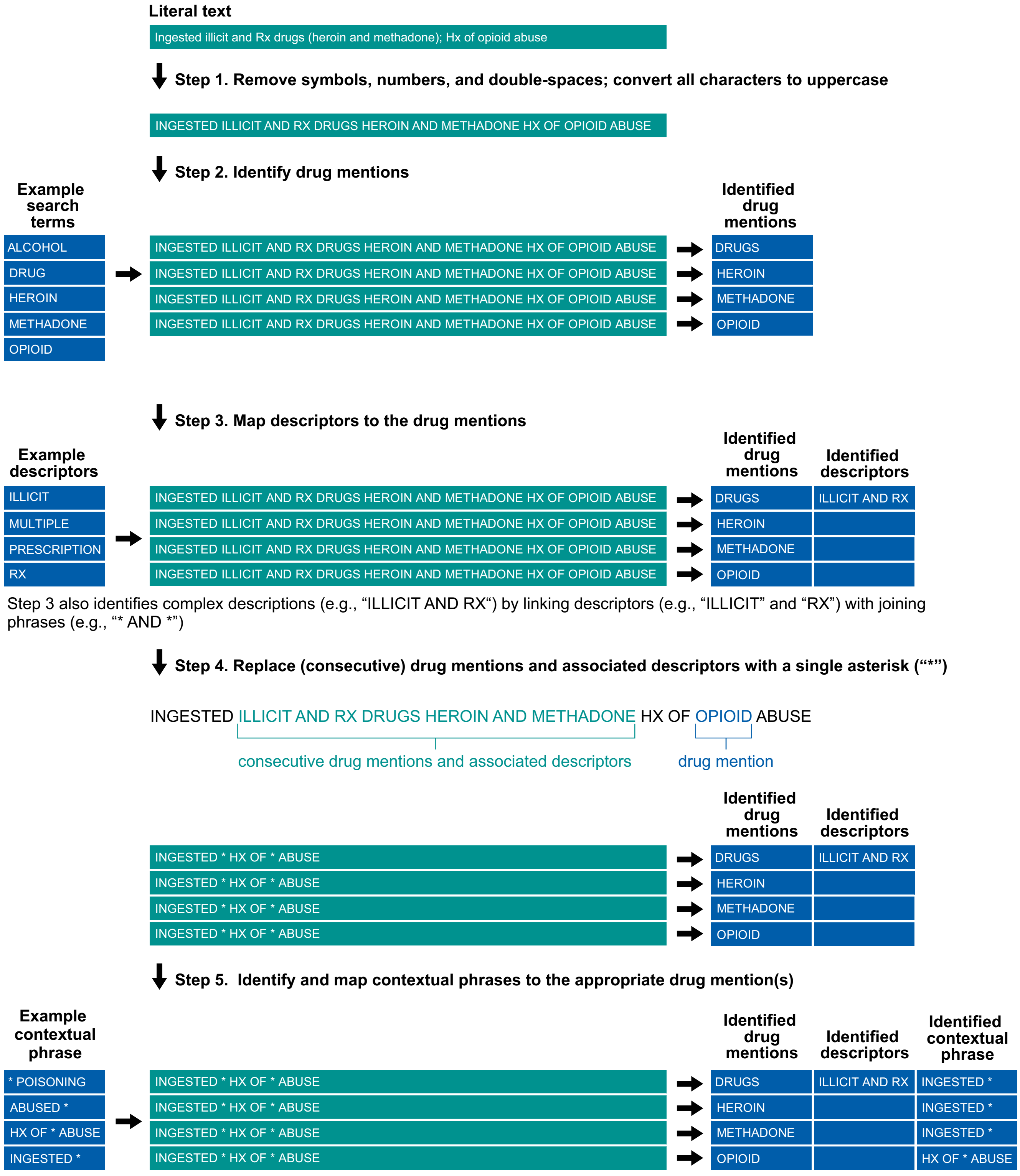
Using SAS Version 9.3 (21), a suite of software programs (referred to as the DMI programs) was developed to automate the identification of drug mentions in the literal text and to determine possible involvement of the drug in the death based on contextual information.

Figure 3 provides an example of the application of the DMI program logic to the following death certificate literal text: "INGESTED ILLICIT AND RX DRUGS (HEROIN AND METHADONE); HX OF OPIOID ABUSE." Leveraging the exchangeability of drug mentions and the exchangeability of descriptors, the DMI programs use five steps to identify drug mentions and ascribe context to each drug mention (Figure 3).

The first step prepares the literal text, resulting in text that does not have symbols, numbers, and double spaces, and is formatted in uppercase letters.

The second step uses the list of search terms to identify drug mentions in the literal text. During this step, a new record is generated for every search term (i.e., drug mention) identified in the literal text. The DMI programs also identify simple plural forms (i.e., search term plus the letter "S"). In the example in Figure 3, the DMI programs generate four records for the mentions of "DRUGS," "HEROIN," "METHADONE," and "OPIOID."

Using the list of descriptors, the DMI programs iteratively identify descriptors for each drug mention in the third step. In the first iteration, the DMI programs identify and map descriptors (such as "RX") to adjacent drug mentions (such as the mention of "DRUGS"), resulting in a drug mention with a simple description (e.g., "RX DRUGS"). Subsequent iterations use the list of joining phrases and list of descriptors to form more complex descriptions. In the example, the DMI programs link the descriptor "ILLICIT" and the descriptor "RX" with the



Literal text

Ingested illicit and Rx drugs (heroin and methadone); Hx of opioid abuse

Step 1. Remove symbols, numbers, and double-spaces; convert all characters to uppercase

INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE

Step 2. Identify drug mentions

Example search terms		Identified drug mentions
ALCOHOL	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	DRUGS
DRUG	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	HEROIN
HEROIN	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	METHADONE
METHADONE	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	OPIOID
OPIOID		

Step 3. Map descriptors to the drug mentions

Example descriptors		Identified drug mentions	Identified descriptors
ILLICIT	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	DRUGS	ILLICIT AND RX
MULTIPLE	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	HEROIN	
PRESCRIPTION	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	METHADONE	
RX	INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE	OPIOID	

Step 3 also identifies complex descriptions (e.g., “ILLICIT AND RX“) by linking descriptors (e.g., “ILLICIT” and “RX”) with joining phrases (e.g., “* AND *”)

Step 4. Replace (consecutive) drug mentions and associated descriptors with a single asterisk (“*”)

INGESTED ILLICIT AND RX DRUGS HEROIN AND METHADONE HX OF OPIOID ABUSE

	Identified drug mentions	Identified descriptors
INGESTED * HX OF * ABUSE	DRUGS	ILLICIT AND RX
INGESTED * HX OF * ABUSE	HEROIN	
INGESTED * HX OF * ABUSE	METHADONE	
INGESTED * HX OF * ABUSE	OPIOID	

Step 5. Identify and map contextual phrases to the appropriate drug mention(s)

Example contextual phrase		Identified drug mentions	Identified descriptors	Identified contextual phrase
* POISONING	INGESTED * HX OF * ABUSE	DRUGS	ILLICIT AND RX	INGESTED *
ABUSED *	INGESTED * HX OF * ABUSE	HEROIN		INGESTED *
HX OF * ABUSE	INGESTED * HX OF * ABUSE	METHADONE		INGESTED *
INGESTED *	INGESTED * HX OF * ABUSE	OPIOID		HX OF * ABUSE

NOTE: In this example, the DMI (drug mentioned with involvement) programs identify three drug mentions (“DRUGS,” “HEROIN,” “METHADONE”) in the literal text and map these drug mentions to one contextual phrase (“INGESTED *”). The DMI programs also identify one drug mention ("OPIOID") and map this drug mention to one contextual phrase (“HX OF * ABUSE”).
SOURCE: NCHS, Division of Vital Statistics.

Figure 3. Example of the application of the DMI program logic to the literal text

joining phrase "* AND *" to form the more complex description "ILLICIT AND RX." The resultant drug mention and associated descriptors are subsequently more complex (e.g., "ILLICIT AND RX DRUGS").

The fourth step replaces drug mentions and associated descriptors with a single asterisk "*" and also replaces consecutive drug mentions and associated descriptors with a single asterisk "*." This step also uses joining phrases to determine whether drug mentions are consecutive. For example, using the joining phrase "* AND *," the mention of "HEROIN" and the mention of "METHADONE" are consecutive in the text "HEROIN AND METHADONE." Similarly, with the joining phrase "* *," the mention of "ILLICIT AND PRESCRIPTION DRUGS" and "HEROIN" are consecutive mentions. In the example, the mention of "OPIOID" is not listed consecutively with other drug mentions.

Using the list of contextual phrases, the fifth step identifies and maps contextual phrases to the appropriate drug mention(s), that is, the drug mentions that were replaced in step 4. In the example, the mentions of "DRUGS," "HEROIN," and "METHADONE" are mapped to the contextual phrase "INGESTED *," while the mention of "OPIOID" is mapped to the contextual phrase "HX OF * ABUSE."

Each search term is mapped to only one contextual phrase. To optimize the mapping procedures, contextual phrases with asterisks located between other words (e.g., "HX OF * ABUSE") are mapped before contextual phrases with asterisks located at the end of the contextual phrase (e.g., "INGESTED *").

Data produced by applying the literal text analysis methodology

Application of the literal text analysis methodology results in a data set of decedents, drug mentions, and contextual information associated with each drug mention (Figure 3). The drug mentions are categorized by principal variant and whether the drug mentions refer to specific drugs, drug classes, or exposures NOS. The contextual phrases are categorized by indication of involvement of drugs in death. When the processed literal text data are linked with NVSS–M data, the ICD–10 underlying and multiple cause-of-death codes, demographic information, geographic information, and other information in the multiple cause-of-death file are also available.

In this report, the literal text analysis methodology was applied to NVSS–M data linked with literal text for year 2013 as an example. For this analysis, mentions of alcohols, tobacco, and nicotine were excluded, as they are involved in many deaths that do not involve other drugs.

Table B shows the number of U.S. resident deaths with drug mentions and DMIs based on the 2013 literal text. Of the approximately 2.6 million deaths in 2013, 114,621 had at least one drug, alcohol, tobacco, or nicotine mention. The number of deaths with a drug mention was 72,518. The number of deaths with at least one drug mention and no contextual information indicating that the drug was not involved in the death (DMI) was 65,062. Among these deaths, there were 150,342 DMIs, for an average of 2.3 DMIs per death.

Table C shows the level of specificity of the drug mentions (i.e., whether the drug mention was a specific drug, a drug class, or an exposure NOS) for the 65,062 DMI deaths in 2013. The majority of DMIs referred to a specific drug (58%). Most of the specific drug mentions were generic names (82,895 DMIs,

Table B. Deaths with drug mentions and mentions of drug involvement: U.S. residents, 2013

Characteristics	Number of deaths	Number of mentions
Deaths among U.S. residents	2,596,993	…
Deaths with at least one drug, alcohol, tobacco, or nicotine mention	114,621	216,361
Deaths with at least one drug mention	72,518	158,104
Deaths with at least one DMI (drug mentioned with involvement in death)	65,062	150,342

… Category not applicable.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Table C. Number and percentage of DMIs, by level of specificity of the drug mention: U.S. residents, 2013

Type of DMI	Number	Percent
All DMIs	150,342	100.0
Specific drug	87,764	58.4
Drug class	8,979	6.0
Exposure not otherwise specified[1]	53,599	35.7

[1]Category includes nonspecific references to drugs (e.g., mention of "POLYPHARMACY" or "DRUG").

NOTES: Mentions of alcohols, tobacco, and nicotine were excluded from the analyses. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with literal text data.

Table D. Number and percentage of DMI deaths, by level of specificity of the DMI: U.S. residents, 2013

Type of DMI	Number	Percent
All DMI deaths	65,062	100.0
Deaths with mention of at least one specific drug	45,035	69.2
Deaths with mention of a drug class only	4,560	7.0
Deaths without mention of a drug class or specific drug[1]	15,467	23.8

[1] Category includes DMI deaths with mentions of nonspecific drug references (e.g., mention of "POLYPHARMACY" or "DRUG").

NOTES: Mentions of alcohols, tobacco, and nicotine were excluded from the analyses. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with literal text data.

or 95%), while the remainder of the specific drug mentions were other variants, such as brand names and misspellings. Slightly more than one-third of all DMIs (36%) were nonspecific references to drugs.

Similarly, DMI deaths can be categorized by the highest level of specificity of the drugs involved. Table D shows the number of DMI deaths from 2013. Of the 65,062 DMI deaths, 69% had mentions of at least one specific drug, while 7% had mentions of a drug class but not a specific drug. For 24% of the deaths, only nonspecific drug references were found (i.e., neither a drug class nor specific drug were mentioned).

Assessments of the literal text analysis methodology

Two assessments examined the performance of the literal text analysis methodology in identifying DMIs and DMI deaths. The assessments were conducted using NVSS–M data linked with literal text for year 2013.

The first assessment examined the agreement between data produced by the DMI programs and ICD–10 coded data for three selected drugs. In the ICD–10 classification system, there are a few T codes, F codes, and R codes that identify deaths with poisonings, mental and behavioral disorders, and toxicological findings related to specific drugs, respectively. These specific drugs include cocaine, heroin, and methadone. The ICD–10 rules for assigning codes in mortality can be found elsewhere (17). Comparisons were made between the numbers of DMI deaths identified by the DMI programs and the numbers of deaths identified as having one of the specific T, F, or R codes for cocaine, heroin, or methadone (Table E). Considering the differences between the DMI definition (i.e., a drug was mentioned and there was no contextual information indicating that the drug was not involved in the death) and the ICD–10 definitions and rules for assigning T, F, and R codes, there was high agreement (greater than 90%) between the DMI programs and the ICD–10 codes in the identification of deaths involving cocaine, heroin, and methadone (Table E).

The second assessment examined the accuracy of the DMI programs in identifying DMIs and DMI deaths. This assessment was based on two subsets of mortality records that were likely to have DMIs: 1) deaths selected by the application of the DMI programs to mortality records and 2) deaths with no uninformative literal text fields and selected using ICD–10 entity axis codes that likely pertained to a drug-involved mortality. These codes included ICD–10 codes referring to mental or behavioral disorders due to psychoactive substance use, poisonings, adverse effects due to drugs and alcohol, and ICD–10 codes whose title or definition explicitly indicated drug involvement (e.g., P04.4 Fetus and newborn affected by maternal use of drugs of addiction) (Figure 4). ICD–10 codes that only indicated alcohol, tobacco, or nicotine involvement were excluded from the list of selected ICD–10 codes. In summary, the codes used in the analysis included those typically used to identify drug overdose deaths and those that indicated other drug involvement (e.g., anaphylaxis) (2,22).

From the pool of mortality records identified by either of the two selection methods, a simple random sample of 2,000 records was taken and manually reviewed to determine whether drug mentions in the literal text (if any) met the definition of a DMI

Table E. Agreement between DMI programs and selected ICD–10 codes: U.S. residents, 2013

Referent drug	ICD–10 code(s) that apply to referent drug[1]	Deaths with DMI of referent drug[2] [A]	Deaths with ICD–10 code(s) that apply to referent drug[3] [B]	Deaths with either DMI of referent drug or ICD–10 code(s) that apply to referent drug [C]	Deaths with both referent drug mention and ICD–10 code(s) that apply to referent drug [D]	*D/A* x 100	*D/B* x 100	*D/C* x 100
Cocaine	T40.5, F14.–, R78.2	7,324	7,176	7,361	7,139	97.4	99.5	97.0
Heroin	T40.1	8,924	8,360	8,968	8,316	93.2	99.5	92.7
Methadone	T40.3	4,005	3,737	4,029	3,713	92.7	99.4	9.2

[1] ICD–10 codes used in this analysis were entity axis codes.

[2] The DMI programs identify deaths with mention of the referent drug in the literal text fields, excluding mentions where the contextual information suggested that the drug was not involved in the death.

[3] The listed T codes, F codes, and R codes identify deaths due to poisonings, mental and behavioral disorders, and toxicological findings related to the referent drug, respectively.

NOTES: DMI is a drug mentioned with involvement in the death. ICD–10 is *International Classification of Diseases and Related Health Problems, Tenth Revision*.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

A80.0, D52.1, D59.0, D59.2, D61.1, D64.2, D68.3, E03.2, E06.4, E16.0, E23.1, E24.2, E27.3, E66.1, F11–F16, F19, F55, G21.1, G24.0, G25.1, G25.4, G25.6, G44.4, G62.0, G72.0, H26.3, H40.6, I42.7, I95.2, J70.2, J70.3, J70.4, K85.3, L10.5, L23.3, L24.4, L25.1, L27[.0–.1], L27[.8–.9], L43.2, L56[.0–.1], L64.0, M10.2, M32.0, M34.2, M80.4, M81.4, M83.5, M87.1, N14[.0–.2], O35.5, P04[.0–.1], P04.4, P04[.8–.9], P58.4, P93, P96[.1–.2], Q86[.1–.2], R50.2, R78[.1–.6], R78[.8–.9], R82.5, R83[.2–.3], R84[.2–.3], R85[.2–.3], R86[.2–.3], R87[.2–.3], R89[.2–.3], T36, T37, T38, T39, T39[.1–.4, .8–.9], T42, T43–T50, T57[.8–.9], T65[.5, .8–.9], T88[.0–.1, .6–.7], T96, T97, X4T400–X44, X49, X60–X64, X69, X85, X89–X90, Y10–Y14, Y19, Y40–Y47, Y49–Y59, Y88.0, Z03.6, Z72.2, Z91.0, Z92[.1–.2]

SOURCE: *International Classification of Diseases and Related Health Problems, Tenth Revision* (ICD–10).

Figure 4. ICD–10 entity axis codes likely pertaining to a drug-involved mortality

and whether the sampled record reflected a true DMI death. The results from the manual review served as the "gold standard."

The performance of the DMI programs to identify DMIs and DMI deaths was quantified using the following measures: true positives, false positives, false negatives, true negatives (only calculated for deaths, not drug mentions), and positive predictive values (PPVs). Each drug mention was categorized as either a true-positive mention (identified by both the DMI programs and by manual review), a false positive mention (identified by the DMI program but not the manual review), or a false negative mention (identified by manual review but not the DMI program). Reasons that a mention was categorized as false positive or false negative were described.

Similarly, each death was categorized as either a true-positive death, false positive death, or false negative death. True-negative deaths were identified only by ICD–10 codes, but were not categorized as DMI deaths according to manual review. PPVs quantified the percentage of DMIs or DMI deaths correctly identified as such by the DMI programs (i.e., true positives/[true positives + false positives]). Measures of sensitivity and specificity could not be calculated because the selected records were not randomly sampled from all mortality records.

From the application of the DMI programs, 65,062 deaths were identified as possible DMI deaths in 2013. From selection based on ICD–10 codes (Figure 4), 61,282 deaths were identified as likely pertaining to drug involvement. Combined, the two methods identified 69,493 unique deaths with possible drug involvement. A majority of these deaths (56,851 or 81.8%) was identified by both methods, while a minority of these deaths (4,431 or 6.4%) was only identified using ICD–10 codes. The remaining deaths (8,211 or 11.8%) were identified by the DMI programs only.

The 2,000 randomly sampled deaths included 1,808 deaths identified using ICD–10 codes, of which 1,691 deaths were also identified by the DMI programs. The remaining 192 deaths in the sample were only identified by the DMI programs.

The DMI programs identified DMIs with high accuracy (Table F). According to manual review of literal text, 4,357 (97%) of the 4,487 mentions identified by the DIM programs were true-positive mentions, while the remaining 130 mentions (3%) were categorized as false positive. The DMI programs failed to identify 52 mentions of drugs involved in mortality. Some deaths may have a mixture of true-positive, false-positive, and false-negative mentions in their literal text.

The DMI programs also identified DMI deaths with high accuracy (Table G). According to manual review of literal text for deaths identified using ICD–10 codes, 1,804 of the 1,883 deaths (96%) identified by the DMI programs were true-positive deaths, while the remaining 79 deaths (4%) were categorized as false positive. The DMI programs did not identify 100 deaths that did not have drug involvement (true-negative deaths), but failed to identify 17 deaths that did have drug involvement (false-negative deaths). All 117 of these deaths were identified using ICD–10 codes.

The false-positive and false-negative mentions fell into nine categories (Table H). In a few instances, the DMI programs identified more text than should have been identified. For example, the DMI programs identified a mention of "PAIN NARCOTIC" instead of "NARCOTIC" in the literal text "BACK PAIN NARCOTIC DEPENDENT." In contrast, the DMI programs sometimes identified one or more search terms that were nested in a longer drug name, resulting in false-positive and false-negative mentions. The DMI programs also identified false-positive mentions for other reasons, including: search terms were not drugs, search terms were used to describe health conditions and disease states, or contextual information indicated no drug involvement. Manual review of literal text also identified other reasons for false-negative mentions: drugs mentioned in the literal text were not search terms, or a drug mention was not separated by a space from other words in literal text.

The findings from the assessment were used to update and improve the lists of search terms and contextual information.

Table F. DMI programs' ability to identify DMIs among a random sample of 2,000 deaths having one or more ICD–10 entity axis codes or identified using the DMI programs: U.S. residents, 2013

Evaluation	DMIs identified from the manual review		
	Yes	No	Total
DMIs identified by the DMI programs	4,357	130	4,487
DMIs not identified by the DMI programs	52	…	…

… Category not applicable.

NOTES: See Figure 4 for list of entity axis codes. Positive predictive value calculated as: 4,357 mentions/4,487 mentions = 97.1%. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Table G. DMI programs' ability to identify DMI deaths among a random sample of 2,000 deaths having one or more ICD–10 entity axis codes or identified using the DMI programs: U.S. residents, 2013

Evaluation	DMI deaths identified from the manual review		
	Yes	No	Total
DMI deaths identified by the DMI programs	1,804	79	1,883
DMI deaths not identified by the DMI programs	17	100	117

NOTES: See Figure 4 for list of entity axis codes. Positive predictive value calculated as: 1,804 deaths/1,883 deaths = 95.8%. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Table H. Reasons for false-positive and false-negative mentions in the assessment of the DMI programs to identify DMIs and DMI deaths

Reason for false-positive or false-negative mention	Example	Result of assessment
Search term was not a drug	DMI program identified "DIFLUOROETHANE," which is not a drug	Identified a false-positive mention
Search term used to describe health condition or disease state	DMI program identified "FOLIC ACID" in text "FOLIC ACID DEFICIENCY," or identified "PCP," referring to pneumocystis pneumonia	Identified a false-positive mention
Drug mention nested in an identified search term	DMI program identified "PAIN NARCOTIC" instead of "NARCOTIC" in text "BACK PAIN NARCOTIC DEPENDENT"	Identified a false-positive mention and a false-negative mention
Search term was nested in a longer drug name	DMI program identified "DRUG" in text "NON STEROIDAL ANTIINFLAMMATORY DRUG"	Identified a false-positive mention and a false-negative mention
Context adjacent to search term indicated no drug involvement	DMI program identified "DRUG" in text "NO DRUG INVOLVEMENT"	Identified a false-positive mention
Drug was not a search term	"CONTRAST DYE" was not identified because it was not a search term	Identified a false-negative mention
Drug name was not separated from other words in literal text	DMI program failed to identify "ALPRAZOLAM" in text "OPIOID ANDALPRAZOLAM OVERDOSE"	Identified a false-negative mention

NOTES: A false-positive mention indicates that the drug was identified by the DMI programs but not during the manual review. A false-negative mention indicates that the drug was identified during the manual review but not by the DMI programs. DMI is drug mentioned with involvement.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Discussion

New method for identifying drug involvement in death

The application of the literal text analysis methodology described in this report can be used to enhance mortality statistics by facilitating the identification of specific drugs involved in drug overdose deaths and deaths with other drug involvement. ICD–10 (6), which has historically been used to classify the drugs involved in the deaths in NVSS–M, is limited in that the vast majority of drugs are classified into broad categories. For example, oxycodone, hydrocodone, and morphine are all classified to T40.2 (Poisoning: Other opioids) (7). There are a few notable exceptions, such as heroin (T40.1), methadone (T40.3), and cocaine (T40.5), which are separately coded in the case of a drug overdose death. In contrast, the methods described in this report allow for the identification of drugs that are not uniquely identified in ICD–10.

The identification of specific drugs provides flexibility in analyses. Specific drugs can be categorized according to classification schemes different than those of the ICD–10 categories. Identifying specific drugs also allows comparisons between drugs within a particular class. In addition, identifying specific drugs allows for more detailed analysis on deaths involving multiple drugs that are classified to the same or even different categories.

The literal text analysis methodology was developed to extract information on the specific drugs involved in deaths from the nonstructured literal text data obtained from death certificates. Utility of the methodology depends on the quality and quantity of information in literal text. The methodology will not identify a drug mention among deaths whose literal text only states "POISONING" or "OVERDOSE," but does not have any reference to drugs. Many issues were considered when designing this methodology, including the unstructured nature of the data, the number of drugs mentioned, the contextual information describing the drug involvement, and the efficiency of the programs to extract information on the drugs involved. Ultimately, the methods that were developed imitate, to some degree, the current processes used to identify statements in death certificates for eventual translation into ICD–10 codes. With the search terms, descriptors, and contextual phrases identified, it is possible to approximately construct literal text statements related to drug-involved mortality.

Development and maintenance of the DMI lists and programs

The importance of creating comprehensive lists to be used by the DMI programs cannot be overstated. The DMI programs are a series of steps that identify drug mentions, descriptors of those drug mentions, and other contextual information. Each step of the processing of literal text requires lists: search terms, descriptors, joining phrases, or contextual phrases. Incomplete lists used by the DMI programs may result in failure in any of the processing steps, which would result in a failure to associate drug mentions with the appropriate contextual information.

The development of the lists used by the DMI programs requires an understanding of the drugs of interest as well as iterative manual reviews of literal text, and this development process is time intensive. The high percentage of agreement between the DMI programs and the manual review suggests that the lists used by the DMI programs were generally comprehensive. However, even with the careful development of these lists, the DMI programs found a few mentions that did not refer to drugs or failed to identify DMIs. For example, the DMI programs found false-positive mentions of "PCP" that referred to pneumocystis pneumonia, but the programs failed to find mentions of "CONTRAST DYE," which was a drug class that was not a search term. Incompleteness in the list of contextual phrases also yielded false-positive DMIs (e.g., identification of "FOLIC ACID" in the text "FOLIC ACID DEFICIENCY"). These false-positive and false-negative mentions demonstrate the importance of careful development of these lists. Updating and refining the lists used by the DMI programs will help resolve these issues for future investigations.

This report found that a little over one-third of deaths involving drugs did not include information on the death certificate about the specific drug(s) involved. This finding from the literal text analysis is consistent with other analyses of the ICD–10 coded data (23). Efforts are underway in many states to improve the specificity of drugs listed on death certificates (24,25). It is possible that search terms for certain drugs rarely seen in drug overdose deaths were not included despite the multiple avenues taken to develop the list of search terms.

Future directions

Data from the literal text could potentially be used to detect emerging trends in drug-involved mortality. For instance, the methods used in this report could be modified to identify deaths involving newly approved prescription drugs, new illicit drugs, and other health threats. Furthermore, the software programs used to mine the literal text could be modified to help identify emergent trends in drug-involved mortality, even before the annual mortality statistical files are finalized. With the rise of synthetic drugs, such as the fentanyl analogs (26), this may be necessary in the future. In order to detect emerging trends, periodically updating the text search capabilities is critical to surveillance of drug overdose deaths.

The amount of information that can be extracted from the literal text is a function of the level of detail that certifiers provide. There are general references that provide guidance on filling out death certificates that describe the importance of details (27,28). In addition to these general references, there is guidance for certifying drug overdose deaths, which stresses the importance of including the specific drugs involved (24). Because of the importance of including the specific drugs on death certificates for public health purposes, there are recommendations to help epidemiologists develop partnerships to help improve specificity of drugs on the death certificates (25).

Currently, the literal text analysis methodology focuses on using the contextual information to identify the mentions of drugs involved in the death. In the future, additional analysis of the contextual information may be informative. For instance, the method could be used to explore the route of administration (e.g., inhalation, injection, or transdermal), specific drug effects (e.g., anaphylaxis), and antibiotic resistance.

Conclusion

This report details a new method that was developed to extract information from the National Vital Statistics System death certificate literal text to improve national monitoring of drug-involved mortality. The literal text analysis method described in this report leverages existing information on the death certificates for statistical monitoring of drug-involved mortality deaths. Assessments conducted during the methods development process demonstrate that these methods have high accuracy in identifying the drugs mentioned and involved in mortality as well as the corresponding deaths. These methods could be applied to analyze mortality data for causes of death classified to broad ICD categories or for emerging causes of death with no ICD code assigned. Although the methods are limited by the level of drug-specific detail provided in the death certificate literal text, these methods are an enhancement to current ICD–10-coded mortality data.

References

1. Rudd RA, Aleshire N, Zibbell JE, Gladden RM. Increases in drug and opioid overdose deaths—United States, 2000–2014. MMWR Morb Mortal Wkly Rep 64(50–51):1378–82. 2016. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6450a3.htm.
2. Hedegaard H, Chen LH, Warner M. Drug-poisoning deaths involving heroin: United States, 2000–2013. NCHS data brief, no 190. Hyattsville, MD: National Center for Health Statistics. 2015. Available from: http://www.cdc.gov/nchs/data/databriefs/db190.pdf.
3. Chen LH, Hedegaard H, Warner M. Drug-poisoning deaths involving opioid analgesics: United States, 1999–2011. NCHS data brief, no 166. Hyattsville, MD: National Center for Health Statistics. 2014. Available from: http://www.cdc.gov/nchs/data/databriefs/db166.pdf.
4. Centers for Disease Control and Prevention. Vital signs: Overdoses of prescription opioid pain relievers and other drugs among women—United States, 1999–2010. MMWR Morb Mortal Wkly Rep 62(26):537–42. 2013. Available from: http://www.cdc.gov/mmWr/preview/mmwrhtml/mm6226a3.htm.
5. Kochanek KD, Murphy SL, Xu JQ, Tejada-Vera B. Deaths: Final data for 2014. National vital statistics reports; vol 65 no 4. Hyattsville, MD: National Center for Health Statistics. 2016. Available from: http://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_04.pdf.
6. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). Volume 1. 1992.
7. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). Volume 3, Section III: Table of drugs and chemicals. 2011.
8. Davis J, Sabel J, Wright D, Slavova S. Epi tool to analyze overdose death data. Council of State and Territorial Epidemiologists blog post. 2015. Available from: http://www.cste.org/blogpost/1084057/211072/Epi-Tool-to-Analyze-Overdose-Death-Data.
9. Kim SY, Shapiro-Mendoza CK, Chu SY, Camperlengo LT, Anderson RN. Differentiating cause-of-death terminology for deaths coded as sudden infant death syndrome, accidental suffocation, and unknown cause: An investigation using US death certificates, 2003–2004. J Forensic Sci 57(2):364–9. 2012.
10. Koopman B, Zuccon G, Nguyen A, Bergheim A, Grayson N. Automatic ICD–10 classification of cancers from free-text death certificates. Int J Med Inform 84(11):956–65. 2015.
11. Ossiander EM. Using textual cause-of-death data to study drug poisoning deaths. Am J Epidemiol 179(7):884–94. 2014.
12. Shapiro-Mendoza CK, Kim SY, Chu SY, Kahn E, Anderson RN. Using death certificates to characterize sudden infant death syndrome (SIDS): Opportunities and limitations. J Pediatr 156(1):38–43. 2010.
13. Holman RC, Belay ED, Christensen KY, Maddox RA, Miniño AM, Folkema AM, et al. Human prion diseases in the United States. PLoS One 5(1):e8521. 2010.
14. Warner M, Trinidad JP, Bastian BA, et al. Drugs most frequently involved in drug overdose deaths: United States, 2010–2014. National vital statistics reports; vol 65 no 10. Hyattsville, MD: National Center for Health Statistics. 2016.
15. National Center for Health Statistics. Report of the panel to evaluate the U.S. standard certificates. 2000. Available from: https://www.cdc.gov/nchs/data/dvs/panelreport_acc.pdf.
16. National Center for Health Statistics. U.S. Standard Certificate of Death. 2003 revision. Available from: http://www.cdc.gov/nchs/data/dvs/death11-03final-acc.pdf.
17. National Center for Health Statistics. ICD–10 mortality manual parts 2a, 2b, and 2s. Available from: http://www.cdc.gov/nchs/nvss/instruction_manuals.htm.
18. Rowe M. Personal correspondence. 2016.
19. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Drug Abuse Warning Network methodology report, 2010 update. 2010.
20. Food and Drug Administration. FDA Adverse Event Reporting System. Available from: http://www.fda.gov/Drugs/InformationOnDrugs/ucm135151.htm.
21. SAS Institute Inc. SAS (Release 9.3) [computer software]. 2012.
22. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). 2nd ed. Geneva, Switzerland. 2004.
23. 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 3(2)231–7. 2013.
24. Davis GG, National Association of Medical Examiners and American College of Medical Toxicology Expert Panel on Evaluating and Reporting Opiod Deaths. Complete republication: National Association of Medical Examiners position paper: Recommendations for the investigation, diagnosis, and certification of deaths related to opioid drugs. J Med Toxicol (10)1:100–6. 2014.
25. Sabel J, Poel A, Tuazon E, Paone D, Slavova S, Bunn T, et al. Recommendations and lessons learned for improved reporting of drug overdose deaths on death certificates. Council of State and Territorial Epidemiologists. 2016. Available from: http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PDFs/PDFs2/4_25_2016_FINAL-Drug_Overdos.pdf.
26. Gladden RM, Martinez P, Seth P. Fentanyl law enforcement submissions and increases in synthetic opioid-involved overdose deaths—27 states, 2013–2014. MMWR Morb Mortal Wkly Rep 65(33)837–43. 2016. Available from: http://www.cdc.gov/mmwr/volumes/65/wr/mm6533a2.htm.
27. National Center for Health Statistics. Medical examiners' and coroners' handbook on death registration and fetal death reporting. 2003. Available from: http://www.cdc.gov/nchs/data/misc/hb_me.pdf.
28. Hanzlick R. Cause of death and the death certificate: Important information for physicians, coroners, medical examiners, and the public. North Field, IL: College of American Pathologists. 2006.

U.S. DEPARTMENT OF
HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551
Hyattsville, MD 20782–2064

For more NCHS NVSRs, visit:
http://www.cdc.gov/nchs/products/nvsr.htm.

National Vital Statistics Reports, Vol. 65, No. 9, December 20, 2016

Contents

Abstract 1
Introduction 1
Methods Development 2
- Overview 2
- Data source 2
- Considerations in developing methods to process death certificate literal text 4
- Assessment of the presence of uninformative literal text 5
- Exchangeability: Optimizing efficiency of processing literal text information 5
- Developing a search term list for drugs 6
- Developing lists of contextual information 7
- Identifying mentions of drugs and ascribing context 7
- Data produced by applying the literal text analysis methodology 9
- Assessments of the literal text analysis methodology 10

Discussion 12
- New method for identifying drug involvement in death 12
- Development and maintenance of the DMI lists and programs 13
- Future directions 13

Conclusion 13
References 14

Acknowledgments

This joint project between the National Center for Health Statistics and the the Office of Surveillance and Epidemiology within the Center for Drug Evaluation and Research of the U.S. Food and Drug Administration (FDA) utilized the information on deaths involving drugs captured in the National Vital Statistics System–Mortality linked with death certificate literal text. The authors would like to acknowledge the contributions of Dr. Louis An, a pharmacist with the Office of Surveillance and Epidemiology who assisted in the verification of mappings from the individual drug search terms to the principal variants.

Suggested citation

Trinidad JP, Warner M, Bastian BA, et al. Using literal text from the death certificate to enhance mortality statistics: Characterizing drug involvement in deaths. National vital statistics reports; vol 65 no 9. Hyattsville, MD: National Center for Health Statistics. 2016.

For e-mail updates on NCHS publication releases, subscribe online at: http://www.cdc.gov/nchs/govdelivery.htm.
For questions or general information about NCHS: Tel: 1–800–CDC–INFO (1–800–232–4636) • TTY: 1–888–232–6348
Internet: http://www.cdc.gov/nchs • Online request form: http://www.cdc.gov/info
DHHS Publication No. 2017–1120 • CS272542