Text Analysis of Death Certificate Records to Ascertain Drugs Involved in Deaths in the National Vital Statistics System

Merianne Rose Spencer¹, Brigham A. Bastian², and Margaret Warner²

Division of Analysis and Epidemiology¹, Division of Vital Statistics²

National Center for Health Statistics

Government Advances in Statistical Programming (GASP) Workshop
Monday, September 23, 2019
Introduction: NCHS and NVSS

- National Center for Health Statistics (NCHS) provides statistical information that guides actions and policies to improve the health of the American people
- National Vital Statistics System (NVSS) encompasses the processing and coding of birth and death certificate records from 57 jurisdictions*
  - Cause-of-death (ICD-10) codes are assigned to death data received by jurisdictions at NCHS
  - Data are coded, processed and then disseminated for health statistics, surveillance and research

*50 states, New York City, District of Columbia and 5 US territories (American Samoa, Guam, Norther Marianas, Puerto Rico, and Virgin Islands)
Data Source: Death Certificate

Demographic information

Completed by the funeral director using information from the best qualified person: spouse, parent, child, another relative, or other person who has knowledge of the facts

Medical information

For natural causes, completed by attending physician, nurse practitioner, physician’s assistant

For sudden and unexplained deaths, completed by medical examiner, coroner, Justice of the Peace

Demographic information
U.S. Standard Death Certificate: Cause-of-Death Section

Referred to as the literal text:

- the information written by the medical certifier on the cause, manner, circumstances, and other factors contributing to the death
  - Part I – Chain of events that directly caused the death
  - Part II – Significant conditions contributing to death
  - Box 43 – Describe how the injury occurred
Drug Overdose Death Rates, United States, 1999-2017

T40.1 Heroin
T40.2 Natural and semisynthetic opioids
T40.3 Methadone
T40.4 Synthetic opioids other than methadone

Drug Overdose Death Rates, United States, 1999-2017

Deaths per 100,000 standard population

- T40.2 Natural and semisynthetic opioids
- T40.1 Heroin
- T40.3 Methadone
- T40.4 Synthetic opioids other than methadone

Examples of common drugs found on death certificates without specific ICD-10 code

**Opioids**
- Fentanyl
- Fentanyl analogs
- Oxycodone
- Hydrocodone
- Morphine
- Hydrocodone
- Tramadol

**Stimulants**
- Methamphetamine
- MDMA

**Benzodiazepines**
- Alprazolam
- Diazepam
- Clonazepam
Methods to Identify Specific Drugs on Death Records

- Collaboration between NCHS and FDA
- Developed methods to analyze literal text for mentions of specific drugs
  - Preprocessed text (e.g., remove stop words, special characters)
  - Examined and reviewed string terms and phrases
  - Categorized terms into categories (i.e., principal variants)
  - Considered context (e.g., “history of”, “insulin-dependent”)
- Referred to as the Drugs Mentioned with Involvement (DMI) methodology

1 Drugs Mentioned with Involvement (DMI) methodology: [https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_09.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_09.pdf)
### Top drugs involved in drug overdose deaths, 2011 and 2016

<table>
<thead>
<tr>
<th>Rank</th>
<th>Referent drug</th>
<th># of deaths 2011</th>
<th>Referent drug</th>
<th># of deaths 2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Oxycodone</td>
<td>5,587</td>
<td>Fentanyl</td>
<td>18,335</td>
</tr>
<tr>
<td>2</td>
<td>Cocaine</td>
<td>5,070</td>
<td>Heroin</td>
<td>15,961</td>
</tr>
<tr>
<td>3</td>
<td>Heroin</td>
<td>4,571</td>
<td>Cocaine</td>
<td>11,316</td>
</tr>
<tr>
<td>4</td>
<td>Methadone</td>
<td>4,545</td>
<td>Methamphetamine</td>
<td>6,762</td>
</tr>
<tr>
<td>5</td>
<td>Alprazolam</td>
<td>4,066</td>
<td>Alprazolam</td>
<td>6,209</td>
</tr>
<tr>
<td>6</td>
<td>Morphine</td>
<td>3,290</td>
<td>Oxycodone</td>
<td>6,199</td>
</tr>
<tr>
<td>7</td>
<td>Hydrocodone</td>
<td>3,206</td>
<td>Morphine</td>
<td>5,014</td>
</tr>
<tr>
<td>8</td>
<td>Methamphetamine</td>
<td>1,887</td>
<td>Methadone</td>
<td>3,493</td>
</tr>
<tr>
<td>9</td>
<td>Diazepam</td>
<td>1,698</td>
<td>Hydrocodone</td>
<td>3,199</td>
</tr>
<tr>
<td>10</td>
<td>Fentanyl</td>
<td>1,662</td>
<td>Diazepam</td>
<td>2,022</td>
</tr>
</tbody>
</table>

### Selected Results

**Age-adjusted rates for drug overdose deaths involving fentanyl, by sex: United States, 2011-2016**


Considerations and Next Steps

- Literal text analyses can enable researchers to report with more granularity, the specific drugs involved in deaths

- Challenges
  - Identification of specific drugs is dependent on testing, interpretation and reporting by death certifiers
  - Temporal changes in reporting of the specific drugs can impact trends analyses

- Future steps
  - Automate these processes (e.g., machine learning), while recognizing that we must also identify terms that appear a small number of times
Questions?

Merianne Rose T. Spencer, M.P.H.
Epidemiologist
Division of Analysis and Epidemiology
National Center for Health Statistics
301-458-4377 | kvd1@cdc.gov

Brigham A. Bastian, B.S.
Statistician (Demography)
Division of Vital Statistics
National Center for Health Statistics
301-458-4517 | xcb7@cdc.gov

Margaret Warner, Ph.D.
Injury Epidemiologist
Division of Vital Statistics
National Center for Health Statistics
301-458-4556 | mmw9@cdc.gov

The findings and conclusions of this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.