Datasheet for the Dataset of Leading Causes of Death in Alberta*

Zixi Song   Yiyi Yao   Pu Yuan

March 11, 2024

We made a dataset to look at the main causes of mortality in Alberta from 2001 to 2022. This helps us see health patterns and figure out a better-fitting model. The data, put together by Service Albert, shows each year's death causes and total number of deaths. It is helpful for health experts and government workers to plan better health actions and studies.

0.0.1 Motivation

- **Purpose of Dataset Creation**: The dataset was assembled to present a comprehensive overview of mortality patterns in Alberta, listing the 30 most common causes of death annually. This information aids in understanding health trends and guiding public health interventions and policies. The dataset was compiled to provide structured and accessible mortality data, filling a gap where such detailed information was previously unavailable in a consolidated format.

- **Creators**: The dataset was compiled by Service Alberta, reflecting the collaborative efforts of various governmental health departments.

0.0.2 Composition

- **Data Instances**: The dataset comprises yearly records from 2001 to 2022, each record enumerating the instances of death categorized by the cause. Each instance correlates to a specific year and a cause of death.

- **Number of Instances**: The dataset encapsulates annual statistics, with the exact number of instances corresponding to the unique combinations of years and causes of death.

---

*Code and data are available at: https://github.com/Yaoee111/mortality_alberta.git.

- **Sampling**: The dataset is exhaustive, encompassing all recorded instances of death attributed to the specified causes within the stipulated timeframe in Alberta.

- **Data Type**: Each instance includes structured data with attributes such as the year, cause of death, and the total number of deaths for that cause.

- **Labels**: Each record is uniquely identified by the year and cause of death, with no missing information reported.

- **Explicit Relationships**: The dataset is self-contained with explicit year-to-cause mappings. No external relationships are described.

- **Data Splits**: Not applicable, as the dataset is a comprehensive collection rather than a sample requiring splits.

- **Errors/Redundancies**: Potential discrepancies may arise from the interpretation of cause-of-death classifications, though efforts are made to adhere to the International Classification of Diseases 10th Edition standards.

- **External Resources**: The dataset is self-contained and does not rely on external resources.

0.0.3 Collection Process

- **Data Collection**: The dataset aggregates officially recorded vital statistics related to mortality, sourced from governmental health databases.

- **Collection Mechanisms**: Standardized data collection protocols are employed, consistent with government practices for vital statistics. Validation is an ongoing process aligned with data entry and collection standards.

0.0.4 Preprocessing/Cleaning/Labeling

- **Preprocessing Done**: The data was subjected to cleaning and structuring to ensure accuracy and usability. Such preprocessing includes categorization based on the International Classification of Diseases.

- **Raw Data Availability**: Raw data may be available through official requests but is not directly linked in the dataset.

- **Preprocessing Software**: Data processing utilized government-sanctioned software tools and methodologies, specifics of which are maintained within the departmental records.

0.0.5 Uses

- **Current Uses**: The dataset serves as a reference for health researchers, policymakers, and public health officials to analyze mortality trends and craft health strategies in Alberta.

- **Potential Uses**: Future applications could extend to epidemiological modeling, healthcare resource allocation, and educational purposes in medical training.

- **Restrictions on Use**: The dataset is governed by the Open Government Licence - Alberta, which permits a wide range of uses provided that proper attribution is given to the source.

0.0.6 Distribution

- **Distribution Channels**: The dataset is available for download at the specified URL: Leading Causes of Death - Alberta.

- **Licence**: Distributed under the Open Government Licence - Alberta, which allows for free use, modification, and sharing of data.

0.0.7 Maintenance

- **Maintainers**: The dataset is maintained by Service Alberta, specifically the Registry Services division.

- **Contact Information**: Queries can be directed to Registry Services at sa.opendata-rs@gov.ab.ca.

- **Update Schedule**: The dataset is updated annually, with modifications and updates communicated through the dataset's official webpage.

- **Versioning**: Past versions of the dataset are superseded by updated releases, with change logs documented in the dataset metadata.

0.0.8 Ethical Considerations

- **Data Privacy**: The dataset does not include personally identifiable information, as it aggregates mortality data at a macro level.

- **Ethical Oversight**: As a compilation of publicly available vital statistics, ethical considerations primarily pertain to the accurate and respectful presentation of mortality data.

0.0.9 Disclaimer

- **Limitations**: The dataset is restricted to the geopolitical boundaries of Alberta and may not be indicative of trends in other regions.

- **Errata and Updates**: Users are advised to refer to the dataset's webpage for the latest information and to report any discrepancies for correction.

The information provided in this datasheet is intended to give potential users a clear understanding of the dataset's scope, limitations, and appropriate applications. It aims to facilitate responsible use of the data and to encourage its application in ways that can positively impact public health and policymaking in Alberta.