In the United States, jails house those convicted of low-level offenses and those awaiting trial, but, despite none of their inmates being sentenced to the death penalty, many die in jails every year. While the Department of Justice (DOJ) tracks information on who dies in jails, unlike other public records, they do not publish any of it, justifying this a motivator for jails to report complete and accurate information without fear of punishment (Eisler et al). In order to access this public record, one must go through a lengthy information request process.

This process makes it next to impossible to advocate for or legislate jail death reforms because there is no database of publicly-accessible jail death records. To fill this information gap, and because of an independent investigation into jail deaths, Reuters journalists decided to compile and disseminate a dataset that tracks all jail deaths from 2008-2019 from every large jail in America (defined as those with 750+ inmates), making sure to have at least ten jails per state for a nationally representative dataset (Smith). Ultimately, they ended up collecting information from 523 jails via 1,500+ public records requests (Smith). The dataset is publicly available on Reuters’ website, linked to their investigation, “Dying Inside,” published October 16, 2020.

The data is compiled into one national dataset, but is also available divided by state. The datasets themselves are CSV files, but there are also PDFs outlining the key findings nationally and by state. The dataset CSV files come in a downloadable folder with a spreadsheet of details about the jails cataloged, a spreadsheet of deaths cataloged, and a spreadsheet that serves as a guide to navigate both. Importantly, while this dataset is linked to Reuters’ investigative report into jail deaths, one does not have to click through the report in order to access the dataset, giving the public a chance to look at the data for themselves without the bias of Reuters’ report.

On the jail information spreadsheet, each jail is assigned an ID number so it can be matched with each jail death, and the following information is collected for each jail: state/county Federal Information Processing System (FIPS) code, state code, state name, notes about the state’s jails (when appropriate), the jail’s county, the jail’s name, notes about the jail (when appropriate), total number of deaths from 2008-2019, total deaths in this window divided by suicides, acute drug/alcohol abuse, illness/ natural cause, homicides, accidents, and other deaths, the average daily inmate population for each year, and the jail’s medical providers.

The jail death spreadsheet also begins with a corresponding ID number to ensure the deceased can be matched to their jail. The following information is collected for each death: jail state code, jail county, jail name, year of inmate death, date of death, full name of inmate, inmate last name, inmate first name, inmate middle name or initial, inmate’s suffix, date incarcerated, cause of death by category (accident, suicide, homicide, illness, acute drug/alcohol problem, or other), additional details about the death (when available), inmate’s date of birth, year of birth, age, race (categorized as white, black, Hispanic, Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, or other), additional information about race, inmate gender, whether the inmate was convicted or unconvicted, and a description of any information extracted from the media, rather than the DOJ directly. Not every death has information corresponding with each category, so, at times, Reuters turned to the media for more details.

While there is no reason to believe that Reuters would be motivated to manipulate the data around jail deaths, the necessity of the last category reveals a major problem with this dataset, namely the questionable accuracy of information provided to the DOJ that they then shared with Reuters. “Dying Inside” discusses tactics jails use to downplay and obfuscate details around jail deaths that reflect poorly on them. Inaccuracies stem from issues like some jails not reporting all deaths, jails changing the cause of death if it reflects poorly on them, and jail heads releasing inmates from on hospital deathbeds so that their deaths do not count towards total jail deaths (Eisler et al). Local officials running jails can face serious consequences if they are discovered with disproportionately high death rates, avoidable medical deaths, or are discovered to have otherwise caused or facilitated an undue death, so bias exists at the very beginning of the data collection process.

Assuming Reuters got all available information without bias from the DOJ, there are still likely huge inaccuracies in the dataset because of the selection bias, underreporting, inaccurate reporting, and general pro-jail bias in the data collection process. This sort of bias, however, is unavoidable in this sort of dataset unless there is a way to have a neutral third-party track and collect jail death data. The DOJ’s choice to not publish this information likely helps decrease inaccuracies in the data, but even the DOJ can impose sanctions on or push for investigations into wrongful jail deaths, so, bias would be hard to eliminate without a complete carte blanche for jails when deaths occur. While this data is inherently biased, in Reuters’ phase of data collection, there does not appear to be any bias.

Another issue with this dataset is the gaps in information about jail deaths, again linked to underreporting either by the DOJ or by jails themselves. While Reuters was unable to uncover information for each of their selected categories, using these records and the media, they did assemble all publicly-accessible information. Gaps do still exist, but their methodology suggests that all information that could be collected was collected.

Still, with known gaps and biases in available information, any findings from this dataset must be accurately contextualized before being used. For example, some jails never document whether or not inmates are convicted at their time of death, so the raw number of known unconvicted inmate deaths could be pulled from this dataset, but should be contextualized as the minimum number of unconvicted inmate deaths, recognizing the real total is probably higher. Further, any attempts to compare data about a specific attribute to the rest of the dataset at-large, like trying to find the percentage of unconvicted inmates among all jail deaths, need to be heavily contextualized, and may need to be redesigned. While it would be fair to pull out a percentage of those unconvicted within the population designated with a convicted/unconvicted status, it would not be accurate to take the percentage of unconvicted people from the whole dataset with or without a designation, as the conviction status of many is unknown.

While this dataset is incomplete and, in some cases, inaccurate, with the proper context, it is still an extremely useful tool for understanding who is dying and jail and why. Ultimately, Reuters can publish clean, thoughtfully collected data, but because they are just compiling a dataset, and not collecting the raw data itself, any conclusions drawn from it must be properly contextualized before publication. All data has an origin, so even information from reputable sources like Reuters needs to be considered critically and contextualized properly before usage.

Works Cited

*Dying Inside: All States in Survey*. Reuters, 16 Oct. 2020, https://graphics.thomsonreuters.com/data/jails/Allstatesinsurvey.zip.

Eisler, Peter, et al. “Dying Inside: Death Sentence.” *Reuters Investigates*, Reuters, 16 Oct. 2020, www.reuters.com/investigates/special-report/usa-jails-deaths/.

Smith, Grant. “Dying Inside: Data and Key Findings.” *Reuters Investigates*, Reuters, 16 Oct. 2020, www.reuters.com/investigates/special-report/usa-jails-graphic/.

*I pledge my honor that this assignment represents my own work in accordance with University standards. - Alison Fortenberry*