

Mortality in Correctional Institutions (MCI) Disclosure Analysis
Prison Data on and after October 1, 2015

October 26, 2023

To prepare for a potential release of MCI datasets to the public, BJS conducted a disclosure analysis of data provided for state prisons on and after October 1, 2015, through the Mortality in Correctional Institutions (MCI) program.

The data provided contain many columns that can be used directly or in combination with each other or other publicly available data to determine a person's identity.

At the outset, BJS fully redacted the following columns which comprise the person's name and date of birth direct PII: Fname, Lname, Mname, DoByr, Dobday, and Dobmon.

Even with these columns redacted, other columns can be used in combination with other data to re-identify an individual in the MCI dataset.

Here are two scenarios:

Scenario 1

- Step 1: A user identifies that there is only one person of Race A and Sex B who died at facility C in year Y. From the file, they also learn the person's offenses, cause of death, and date of death.
- Step 2: The user uses the state's inmate search directory to obtain a list of all inmates of that race and sex at facility C. The tool returns a list of inmates.

Step 3: The user searches through the returned list, they find one person listed as released on the date of the death listed in the MCI dataset. They now know the person's name, N, in addition to other information supplied by the state department of corrections.
- Step 3: The user now can attribute all the information in the MCI dataset to person N.

Scenario 2

- Step 1: A person reads a news article about the recent death of a named individual at facility C. The news article gives some additional information including date of death and cause of death.
- Step 2: The person uses the state's inmate search directory to search for an inmate with name N at facility C. The directory returns a record for the inmate and includes date of release (by death) and other information supplied by the state department of corrections.
- Step 3: The person uses the information available from the search directory and the news article and links to a single record in the MCI dataset.

BJS applied a technique called “k-anonymity” to assess and reduce the identifiability of the data. K-anonymity “provides privacy protection by guaranteeing that each released record will relate to at least k individuals even if the records are directly linked to external information.”¹

To use this technique, BJS first needed to determine which columns to designate as “quasi-identifiers,” meaning that they could be used to identify someone if combined with other information, either on the MCI dataset or other available data.

BJS classified the following columns as quasi-identifiers:

Information	Column(s)
State	State
Release/Death Date	Deathmon, Deathday
Facility	Facname, Facloc
Race	Race, Hispanic, Race1, Race2, Race3, Race4, Race5, Race6, Othrace
Admission Date	Admitmon, Admitday , Admityr
Offense	Curroff1, Curroff2, Curroff3, Curroff4, Curroff5
Cause of Death	Cause, Illspec, Drugspec, Selfspec, Accspec, Suicspec, Homispec, Otherspec, Forcespec

BJS determined to apply this technique with k=2, which means every row in the protected dataset (where each row represents an individual decedent) will have at least one other row with the same combination of values for the quasi-identifiers. Therefore, if someone tries to reidentify someone in the dataset using some combination of quasi-identifiers, each decedent will have at least one additional decedent with the same variable values for the quasi-identifiers who could plausibly be thought to represent the decedent in question.

The following steps were taken to prepare the redacted dataset. All steps were conducted using software called RStudio. The k-anonymity algorithm used is part of the associated R package sdcMicro.²

- 1) Load **mci_state_prison_2010-2019_09152023**.
- 2) Subset the data to contain records where year is 2016 or greater OR year is 2015 and deathmon is 10 or greater.
- 3) Subset the data to keep the 56 columns that are part of the FOIA request.
- 4) Redact the information for all rows for fname, lname, mname, dobyr, dobmon, dobday.
- 5) For each of the quasi-identifiers listed above, redact any cell that is unique within its column.
- 6) Enforce K-anonymity, with the following parameters:
 - a. Quasi-identifiers as listed above.
 - b. k=2

¹ See <https://dataprivacylab.org/dataprivacy/projects/kanonymity/kanonymity2.pdf>

² See <https://www.jstatsoft.org/article/view/v067i04>

- c. Importance is null, meaning no columns are identified as priorities to not be suppressed. As such, the algorithm will minimize the number of cell suppressions, agnostic to any importance of a particular column.
 - d. All other parameters were unspecified and ran as the default setting.
- 7) Apply redactions on associated columns.
- a. If any of Race, Race1, Race2, Race3, Race4, Race5, Race6, Othrace, or Hispanic are redacted, redact Race1, Race2, Race3, Race4, Race5, Race6, Othrace and Hispanic.
 - b. If Cause is redacted, redact all of Illspec, Drugspec, Selfspec, Accspec, Suicspec, Homispec, Otherspec, Forcespec.
 - c. If State is redacted redact Facname and Facloc if they were not blank at the start.
 - d. If Facname is redacted, redact Facloc if wasn't blank at the start.
 - e. If Facloc is redacted, redact Facname if wasn't blank at the start.
- 8) For any cell that was redacted, insert the text "Redacted".
- 9) Add a column called FOIA_note that indicates "FOIA Exemptions (b)(6), (b)(7)(c)" for each row.
- 10) Randomize the order of the rows.
- 11) Output **MCI_FY2016-2019_k-anon-2023-10-26**.