This describes the actions of /CHRP/NEOCARE/datawork/ohio\_death\_data/ohio\_death\_data.Rmd. This file processes the Ohio Death Index data and merges it with MH patient data in order to obtain the death data for MH patients.

* The 1999-2016 Ohio Death Index files were read in, stacked (dplyr::bind\_rows()), and processed:
  + The social security number column was standardized to only include the last four digits of the social security number.
  + Accent marks in first and last names were removed and coerced to uppercase so that only the 26 characters of the English alphabet would remain.
  + Rows were removed if the first and last names were both “REDACT”
* The already processed demographics table was read in from /NEOCARE/Corrected Data-MH Jan 2019/mh\_cohort\_age\_calculated/mh\_demogs.rds (see mh\_data\_prep.docx)
* The above two tables were merged on last four digits of social security number, first name, last name, and birth year. Matching was re-attempted on rows that did not match, using last name, first name, date of birth, and sex. A third round of matching was performed only on 2017 death records on given\_name, last\_name, and birth\_date since the ODH did not give us social security number nor sex.
* The resulting matches were written as /CHRP/NEOCARE/Corrected Data-MH Jan 2019/to\_ccf/2020-07-22/mh\_deaths.csv, which was transferred to the Cleveland Clinic via FileZilla.