# STUDY OBJECTIVES

Historically, detailed occupation has not been systematically collected and reported by any known nationwide central cancer registry, including those in the United States (US). As of 2020, detailed occupation has been systematically recorded and made available in US death records, providing an unprecedented opportunity to investigate cancer mortality byoccupation**.1** The purpose of this study is to describe cancer mortality rates in the US by occupation, sex, and cancer site.

# METHODS

## Study design and data sources

The cross-sectional design enables estimation of average annual, age-adjusted cancer mortality rates by occupation and sex for deaths occurring 2020 through 2023 among working-age adults; those 20 to 64 years old at death. We will calculate mortality rates using deaths reported in National Center for Health Statistics (NCHS) data and population at risk from American Community Survey (ACS) historical data gathered from the Integrated Public Use Microdata Series system (IPUMS).

### Numerator

Cancer deaths will be obtained from the NCHS and National Vital Statistics System’s (NVSS) public use mortality data, the Mortality Multiple Causes of Death (MMCD), from 2020 to 2023. Documentation for recording occupation information in the mortality file is available.2

**Occupation**

Decedent occupation was coded using different schemes based on year of death; 2018 Census Occupation codes among 2023 deaths, and 2010 Census Occupation codes among 2020-2022 deaths. To achieve compatibility between occupation codes among 2023 and 2020-2022 deaths 3 several occupation codes will be re-coded (combined or split) based on the 2018-2010 crosswalk4 provided by the Census Bureau. Those occupations that combined two or more 2010 Census Occupation codes into new and fewer 2018 Census Occupation Codes will be assigned relative percent weights (e.g., instead of 1 death among 1 occupation in 2023, a fraction of a death among multiple occupations in 2023) based on the number of employed individuals on the most recent year that all the 2010 codes were available within population denominator data from IPUMS-ACS (Appendix 1, part 1). Several decedent occupations will be further re-coded to align with IPUMS population denominators. Thus, final decedent occupation will be based on recodes from the 2018-2010 Census Occupation code crosswalk and harmonization of decedent 2010 Census Occupation codes with IPUMS-ACS 2010 Census Occupation codes (Appendix 1).

The resulting 2010 Census Occupation codes are a hierarchical coding system5 of jobs into:

* 459 occupations
* 23 major groups
* Further aggregation into 16 and 5 groups

**Cancer death**

A licensed medical certifier completed date of death and cause of death (recoded into international classification of diseases, 10th revision (ICD-10) codes) for MMCD decedent information6. These ICD-10 codes have been classified into 22 groups, which will be used to characterize cancer sites/types in this study. Each of the 22 cancer site-/type-specific deaths can also be identified using UCR 113-codes (Appendix 2).

**Selection Criteria**

To represent the target population of cancer deaths among working-age civilians living in the U.S., the following selection criteria will be applied to each dataset:

* Ages 20 to 64
* Resident status of “Residents”, “Intrastate non-residents”, or “Interstate non-residents”
* Natural manner of death
* Cause of death indicating malignant neoplasms (ICD-10 C00-C99)
* Known civilian occupation codes (omitted: ‘Unknown occupation code’ 9999, military codes 9800-9830, missing occupations, or non-paid occupations [Own Home/Homemaker; Volunteer not for pay; retired; student attending a high school or college; can not work, never worked, disabled, patient or inmate])

Cancer deaths overall and by site will be summed across 2020-2023 in preparation of calculating average annual rates for each age, sex, and occupation.

### Denominator

For each age-sex- and occupation strata generated in the numerator, corresponding age-sex- and occupation-specific denominators will be obtained from annual ACS-IPUMS.7,8

**Historical cohort**

Age-, sex-, occupation, and year-specific cohorts will be generated based on historical employment data from 1976-2023 to account for occupational exposure variation due to recording of decedent’s ‘usual’ occupation, cancer latency from occupation-related exposure, different exposure duration based on years working in occupation, and temporal changes in employment by occupation, sex, and age. Within each calendar year (2020-2023), each age-, sex-, and occupation-specific cohorts will be measured as the average number of people employed in an occupation across a working lifetime, where working lifetime occurs between 20 and 64 years of age for all relevant sex-, age-, year-, and occupation strata. Therefore, the following calculation of single age (20-64 years)- and year (2020-2023)-specific denominators for each sex and occupation will be used as the denominator:

Because ACS data prior to 2006 were assessed decennially (i.e., 2000, 1990, 1980, 1970), counts for years when ACS was not administered were generated by linear interpolation using values of the nearest survey years that the occupation was assessed and in existence.

* For example, a sex and occupation-specific 40-year-old age cohort for those who died in 2020 is defined as the average of number of 40-year-olds in 2020 (i.e., oldest age within working lifetime), number of 39-year-olds in 2019, back tracked to the number of 20-year-olds in 2000 (i.e., youngest age within working lifetime).

**Selection criteria**

Following the calculation for adjusted persons at risk above, only known civilian occupation codes will be included in the denominator; excluding the following occupations:

* ‘Unknown occupation code’ 9999
* Military codes 9800-9830
* Missing occupations

**Sex-specific correction**

To account for changes in population and employed persons across the historical cohort and yield contemporarily comparable rates, each stratum-specific adjusted persons at risk calculation above will be multiplied by a sex-specific correction, weightm or weightf, to match the sex-stratified employed working-age population estimated from ACS 2020-2023. The ACS 2020-2023 average population estimation will include only employed working-age civilians who live in states within a Vital Statistics jurisdiction in each respective year. Thus the following states will be excluded from calculating the correction factor:

* 2020 exclusion: Iowa, Arizona, North Carolina, Rhode Island, and Washington, D.C.
* 2021 exclusion: Rhode Island, and Washington, D.C.
* 2022-2023 exclusion: none

The weight for males will be the number of 20-64-year-old employed males in the ACS summed between 2020 and 2023 divided by the annually (2020-2023) summed historical cohort of 20-64-year-old males. Similarly, the weight for females will be the number of working age employed females in the ACS summed between 2020 and 2023 divided by the annually (2020-2023) summed historical cohort of working age females.

The mathematical product between the sex-specific correction factor and stratum-specific observed denominator will be the final denominator used for each age-, sex, and occupation-specific strata (below).

## Statistical analysis

### Age-adjusted mortality rates

The primary outcome will be age-adjusted cancer mortality rates calculated for each sex and occupation. We will use the direct adjustment method where, first, deaths (numerators)and employed (denominators) will be separately summed over the years of 2020-2023. Age-adjusted death rates were calculated using standardized, five-year age groups (20-24, 25-29, …, 60-64) from the Surveillance Epidemiology and End Results (SEER) program. Age-adjusted mortality rates will be calculated as the weighted sum of the age group-specific rates and presented per 100,000 persons.

Sex- and occupation-specific mortality rate =

Age strata weights will use the 2000 SEER standard million proportions rescaled to weights of age groups between 20 and 64 years so that rates are comparable to age-adjusted rates of other populations (i.e., ensuring that age group weights sum to 1.0).9

Crude mortality rates will also be presented based on the quotient of sex- and occupation-specific deaths sex- and occupation-specific employed and presented as per 100,000 persons.

The confidence intervals for age-adjusted mortality rates will be calculated using the Tiwari method based on modified gamma distribution intervals, as implemented in US-wide cancer surveillance systems.10

Any occupation with less than 10 cancer deaths totaled across all four years will be considered unstable and combined into “not elsewhere classified” or “all other” categories within the occupation’s major category. Deaths by the 22 cancer sites will be identified using ICD-10 disease codes (Table S1), and site-specific age-adjusted mortality rates will be calculated by detailed occupation and 23 major occupation categories. We will calculate additional measures in an attempt to understand mortality trends: ranks (1=highest rate) of all stable rates by detailed occupation and the three occupation categories (see ‘Occupation’ above), forest plots of highest 50 and lowest 50 overall cancer mortality rates and 95% CIs by sex and occupation with summaries of cancer deaths and population at risk, crude rates for comparison to age-adjusted rates, and two interactive forest plots of cancer mortality rates, ranks, and 95% CIs (Figures S1a and S1b). Each interactive forest plot will be organized by sex and detailed occupation, but one will display cancer mortality rates by occupation categories (selection 1) and cancer site/type (selection 2), and the other will display mortality rates for all cancer sites/types by detailed occupation (selection 1). The former will convey trends within sex- and cancer site-specific across occupations (Figure S1a), while the latter will summarize cancer site-specific burden for a detailed occupation (Figure S1b). All tables of cancer mortality rates will include average annual deaths as an additional measure of population burden.

### Output

Despite similar coding, data used for the numerators and denominators were collected through separate agencies and denominator data were sampled while numerators are considered a complete counting of deaths. As such, instances of deaths recorded to each sex-age-year- strata for specific occupations that correspond to zero estimated population were therefore collapsed into the respective ‘not elsewhere classified (NEC)’ occupation category. The corresponding NEC category for any occupation within a 23 major 2010 Census Occupation grouping that had more than one NEC category will be determined by manual review of each occupation and typically was the NEC category that followed the occupation in numeric sequence. Detailed occupations with less than 10 total deaths are considered statistically unstable and will be collapsed into the respective NEC category using the same procedure described above. Detailed occupations will also be aggregated by the three occupation levels to report higher level occupation mortality rates.

All analyses will be conducted using SAS version 9 and figures, graphs, including forest plots will be generated using R Studio 2024.9.0.

# Appendix

Appendix 1. Occupation recodes applied to NCHS data



Please see standalone excel above. Crosswalk between 2018 and 2010 will first be conducted to create 2010 OCC. The resulting 2010 OCC will then be aggregated into an appropriate *SOC2010* variable recode.

Appendix 2. Cancer site-specific codes identified by UCR113.

|  |  |
| --- | --- |
| Site/ICD-10 | UCR113 recode |
| Malignant neoplasms of lip, oral cavity and pharynx (C00-C14) | 20 |
| Malignant neoplasm of esophagus (C15) | 21 |
| Malignant neoplasm of stomach (C16) | 22 |
| Malignant neoplasms of colon, rectum and anus (C18-C21) | 23 |
| Malignant neoplasms of liver and intrahepatic bile ducts (C22) | 24 |
| Malignant neoplasm of pancreas (C25) | 25 |
| Malignant neoplasm of larynx (C32) | 26 |
| Malignant neoplasms of trachea, bronchus and lung (C33-C34) | 27 |
| Malignant melanoma of skin (C43) | 28 |
| Malignant neoplasm of breast (C50) | 29 |
| Malignant neoplasm of cervix uteri (C53) | 30 |
| Malignant neoplasms of corpus uteri and uterus, part unspecified (C54-C55) | 31 |
| Malignant neoplasm of ovary (C56) | 32 |
| Malignant neoplasm of prostate (C61) | 33 |
| Malignant neoplasms of kidney and renal pelvis (C64-C65) | 34 |
| Malignant neoplasm of bladder (C67) | 35 |
| Malignant neoplasms of meninges, brain and other parts of central nervous system (C70-C72) | 36 |
| Malignant neoplasms of lymphoid, hematopoietic and related tissue (C81-C96) | 37 |
| Hodgkin's disease (C81) | 38 |
| Non-Hodgkin's lymphoma (C82-C85) | 39 |
| Leukemia (C91-C95) | 40 |
| Multiple myeloma and immunoproliferative neoplasms (C88, C90) | 41 |
| All other and unspecified malignant neoplasms (C17, C23-C24, C26-C31, C37-C41, C44-C49, C51-C52, C57-C60, C62-C63, C66, C68-C69, C73-C80, C96-C97) | 42,43 |

Appendix 3. Occupation Suppression Rules Into “NEC”



Please see standalone excel above. For each strata where the occupation has <10 in annual average deaths, the occupation will be aggregated into its respective “other” or “not otherwise classified” category. This variable is indicated as *occ\_nec*.

REFERENCES

1. National Center for Health Statistics. Vital Statistics Online Data Portal. 2024; <https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple>. Accessed February 4, 2025.

2. Steege A, Billock R, Miniño A. Industry and Occupation (I&O) data as applicable to mortality vital statistics, 2020: History, Background, and Control Tables. In: CDC.

3. Mortality Multiple Cause-of-Death. *Public Use Data File Documentation* <https://www.cdc.gov/nchs/nvss/mortality_public_use_data.htm>, 2025.

4. 2018 Census Occupation Code List with Crosswalk. *Industry and Occupation Code Lists & Crosswalks* <https://www.census.gov/topics/employment/industry-occupation/guidance/code-lists.html>.

5. United States Department of Commerce, Census Bureau. About Occupation. <https://www.census.gov/topics/employment/industry-occupation/about/occupation.html>. Accessed January 14, 2025.

6. Centers for Disease C, Prevention. Instructions for Classifying Multiple Causes of Death, 2022. In: National Center for Health Statistics; 2022.

7. Steven Ruggles SF, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. Integrated Public Use Microdata Series: Version 15.0. In: Minnesota MUo, ed2024.

8. Bureau USC, Department of Commerce E, Statistics A. 2010 American Community Survey 1-Year Estimates; using American FactFinder; <<http://factfinder2.census.gov>>. In*.* Vol 20142014.

9. National Cancer I. Surveillance Epidemiology and End Results. In*.* Vol 2013: Cancer Statistics Branch Surveillance Research Program. Division of Cancer Control and Population Sciences. National Cancer Institute; 2010.

10. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res.* 2006;15(6):547-569.