



Panel A: proportions of all ICD-coded sarcoma-related deaths. Panel B: proportions recalculated among anatomically specified deaths only (sensitivity analysis), excluding C49.9 (unspecified).

Supplementary Figure S1. Five-year aggregated proportional distribution of sarcoma-related deaths by anatomical site, United States, 1999–2020. Two-panel stacked bar chart. Panel A shows the proportions of all ICD-10-coded sarcoma-related deaths. Panel B shows proportions recalculated among anatomically specified deaths only (sensitivity analysis), excluding unspecified site (C49.9).

Supplementary Table S1. ICD-10 Code List and Anatomical Site Mapping

All ICD-10 codes correspond to malignant neoplasms of bone, connective tissue, and soft tissue as the underlying cause of death. Code C49.8 was included in the query but was not returned in the CDC WONDER extract.

ICD-10 Code	Description	Assigned Site Category
C40.0	Malignant neoplasm of scapula and long bones, upper limb	Bone
C40.2	Malignant neoplasm of long bones, lower limb	Bone
C41.0	Malignant neoplasm of bones of skull and face	Bone
C41.1	Malignant neoplasm of mandible	Bone
C41.2	Malignant neoplasm of vertebral column	Bone
C41.3	Malignant neoplasm of ribs, sternum and clavicle	Bone
C41.4	Malignant neoplasm of pelvic bones, sacrum and coccyx	Bone
C41.9	Malignant neoplasm of bone and articular cartilage, unspecified	Bone
C48.0	Malignant neoplasm of retroperitoneum	Retroperitoneal
C48.1	Malignant neoplasm of specified parts of peritoneum	Retroperitoneal
C48.2	Malignant neoplasm of peritoneum, unspecified	Retroperitoneal
C49.0	Malignant neoplasm of connective and soft tissue of head, face and neck	Head & Neck
C49.1	Malignant neoplasm of connective and soft tissue of upper limb, including shoulder	Soft tissue - Extremity
C49.2	Malignant neoplasm of connective and soft tissue of lower limb, including hip	Soft tissue - Extremity
C49.3	Malignant neoplasm of connective and soft tissue of thorax	Soft tissue - Trunk
C49.4	Malignant neoplasm of connective and soft tissue of abdomen	Soft tissue - Trunk
C49.8	Malignant neoplasm of overlapping sites of connective and soft tissue	Not returned in extract (zero deaths)
C49.9	Malignant neoplasm of connective and soft tissue, unspecified	Unspecified

Abbreviation: ICD-10, International Classification of Diseases, 10th Revision.

Supplementary File S3

Reproducibility Supplement: CDC WONDER Query Settings, Data Extraction Details, and Analytic Code Documentation

Divergent Temporal Trends in Site-Specific Sarcoma Mortality in the United States, 1999–2020: A Population-Based Analysis

1. Data Source

All mortality data were obtained from the CDC WONDER (Wide-ranging Online Data for Epidemiologic Research) Underlying Cause of Death database, maintained by the Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS).

Parameter	Value
Database name	CDC WONDER Underlying Cause of Death
Database URL	https://wonder.cdc.gov/ucd-icd10.html
Cause-of-death coding	ICD-10 (International Classification of Diseases, 10th Revision)
Data extracted by	Muhanad Alzahrani (corresponding author)
Extraction date	April 2023
Study period	1999–2020 (22 years)
Geographic scope	United States (all 50 states and District of Columbia)
Population included	All ages, both sexes, all races/ethnicities

2. CDC WONDER Query Settings

The following grouping and filter options were applied in the CDC WONDER interface to extract the dataset used in this analysis:

Query Field	Setting Applied
Group results by	Underlying Cause of death (ICD-10 code); Year
Underlying cause of death	Selected ICD-10 codes (see Section 3 below)
Year/Month	Year; 1999–2020
Gender	All genders (not stratified)

Age groups	All ages (not stratified)
Race	All races (not stratified)
Hispanic origin	All origins (not stratified)
Urbanization	All (not stratified)
Export format	Tab-delimited text (.txt), then imported to R as CSV
Show totals	No
Show zero values	Yes — requested; see Section 4 for handling
Show suppressed values	Yes — requested; see Section 4 for handling
Precision	Standard (deaths, population, crude rate)

3. ICD-10 Code List and Site Mapping

The following ICD-10 codes were queried in CDC WONDER. All codes correspond to malignant neoplasms of bone, connective tissue, and soft tissue as the underlying cause of death. Code C49.8 (other specified connective and soft tissue) was included in the query but was not returned in the extract, indicating zero deaths were coded to this category in any year of the study period.

ICD-10 Code	Description	Assigned Site Category
C40.0	Malignant neoplasm of scapula and long bones, upper limb	Bone
C40.2	Malignant neoplasm of long bones, lower limb	Bone
C41.0	Malignant neoplasm of bones of skull and face	Bone
C41.1	Malignant neoplasm of mandible	Bone
C41.2	Malignant neoplasm of vertebral column	Bone
C41.3	Malignant neoplasm of ribs, sternum and clavicle	Bone
C41.4	Malignant neoplasm of pelvic bones, sacrum and coccyx	Bone
C41.9	Malignant neoplasm of bone and articular cartilage, unspecified	Bone
C48.0	Malignant neoplasm of retroperitoneum	Retroperitoneal

C48.1	Malignant neoplasm of specified parts of peritoneum	Retroperitoneal
C48.2	Malignant neoplasm of peritoneum, unspecified	Retroperitoneal
C49.0	Malignant neoplasm of connective and soft tissue of head, face and neck	Head & Neck
C49.1	Malignant neoplasm of connective and soft tissue of upper limb, including shoulder	Soft tissue – Extremity
C49.2	Malignant neoplasm of connective and soft tissue of lower limb, including hip	Soft tissue – Extremity
C49.3	Malignant neoplasm of connective and soft tissue of thorax	Soft tissue – Trunk
C49.4	Malignant neoplasm of connective and soft tissue of abdomen	Soft tissue – Trunk
C49.8	Malignant neoplasm of overlapping sites of connective and soft tissue	<i>Not returned in extract (zero deaths)</i>
C49.9	Malignant neoplasm of connective and soft tissue, unspecified	Unspecified

4. Handling of Suppressed and Zero-Count Cells

CDC WONDER suppresses death counts of 1–9 in any cell to protect privacy. The query was configured to display suppressed values. In this dataset, no suppressed cells were returned across any ICD-10 code, year, or combination thereof. All site-year cells had death counts of 10 or greater, or were returned as zero. Zero-count cells were retained in the dataset and included in denominators for proportional mortality calculations. No imputation was performed.

5. Software and Packages

Software / Package	Version	Purpose	Citation
R	4.3.3	All statistical analyses and visualisation	R Core Team (2024). R Foundation for Statistical Computing, Vienna, Austria.
dplyr	CRAN	Data manipulation and aggregation	Wickham H, et al. (2023). dplyr: A Grammar of Data Manipulation. R package.

ggplot2	CRAN	Figure generation	Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.
tidyr	CRAN	Data reshaping	Wickham H, Vaughan D, Girlich M (2023). tidyr: Tidy Messy Data. R package.
MASS	CRAN	Negative binomial regression (glm.nb)	Venables WN, Ripley BD (2002). Modern Applied Statistics with S. 4th ed. Springer.

6. Analytic Steps

The complete R analytic code (sarcoma_analysis_revised.R) is available from the corresponding author upon request. The code performs the following steps in order:

Step	Description	Detail
Step 1	Data import and cleaning	CSV export from CDC WONDER imported; 'Suppressed' cells coerced to NA (none present); population strings with commas converted to integer.
Step 2	ICD-10 site classification	All 17 ICD-10 codes mapped to 6 anatomical categories via a pre-defined look-up tibble (see Section 3). C49.8 absence confirmed.
Step 3	Aggregation	Deaths summed by year and anatomical site. Annual total deaths and U.S. population extracted. Proportional mortality and crude rates per 100,000 calculated.
Step 4	Overall trend model	Negative binomial regression of total annual deaths on calendar year (centered at 1999) with log(population) offset. IRR and 95% CI extracted.
Step 5	Site-specific trend models (primary analysis)	Separate negative binomial models fitted for each of the 5 specified anatomical sites and separately for the unspecified site. Year centered at 1999, log(population) offset. IRRs and 95% CIs extracted for each.
Step 6	Interaction model	Single negative binomial model with site \times year interaction (Bone as reference), specified sites only. Likelihood ratio test against additive model.
Step 7	Decomposition analysis	Absolute change in crude mortality rate per 100,000 between 1999 and 2020 calculated per site. Percentage contribution of each site to total change computed, with and without unspecified site in denominator.

Step 8	Sensitivity analysis 1 — exclude 2020	Steps 5 and 6 repeated on 1999–2019 data only.
Step 9	Sensitivity analysis 2 — specified-site proportions	Proportional mortality recalculated using only specified-site deaths as the denominator (C49.9 excluded). Mean, min, max proportions reported.
Step 10	Figure generation	All figures (Figures 1–3, Figure S1) generated using ggplot2 with consistent colour palette, Times New Roman font, theme_classic base theme, and saved at 300 DPI as PNG and TIFF.

7. R Session Information

The following session information is provided for full computational reproducibility:

```
R version 4.3.3 (2024-02-29) -- 'Angel Food Cake'
Platform: x86_64-pc-linux-gnu (64-bit)

Packages used:
dplyr >= 1.1.0
ggplot2 >= 3.4.0
tidyr >= 1.3.0
MASS >= 7.3-60 (glm.nb)
```

All analyses were run on Ubuntu 24 (x86_64). The analytic environment is fully reproducible from the publicly available CDC WONDER data source and the provided R code using the package versions listed above.