

Occupational Differences in Cancer Prevalence Among U.S. Adults: Evidence from NHIS 2023

Shiyong Wu^{1, 2, *}

¹ Department of Mathematics, Oklahoma State University, Stillwater, OK 74078, USA

² Department of Agribusiness, Oklahoma State University, Stillwater, OK 74078, USA

* Corresponding Author Email: alyssa.wu@okstate.edu

Abstract. This analysis examines how lifetime cancer prevalence differs across major occupation groups among U.S. adults, using the 2023 National Health Interview Survey (NHIS). All estimates apply the final adult sample weight (WTFA_A) so that results represent the national population. The outcome was whether a respondent had ever been told by a health professional that they had cancer (CANEV_A). Occupation (EMDOCCUPN2_A), sex, and age served as the core predictors. Weighted prevalence and 95% confidence intervals were reported overall and by demographic subgroups. Adjusted contrasts came from survey-weighted logistic regression (binomial link) across 23 Standard Occupational Classification (SOC) major groups, with the largest weighted group used as the reference. Four prespecified robustness checks were used to gauge sensitivity: pooling very small occupation groups ($n < 100$), replacing the continuous age term with age-band indicators (18–44, 45–64, 65+), estimating sex-stratified models, and excluding the Armed Forces. Cancer prevalence was higher among legal and community/social service workers and lower among farming workers and members of the Armed Forces. These contrasts remained stable across the robustness checks.

Keywords: Cancer Prevalence; Occupation; Weighted GLM; Sensitivity Analysis; Epidemiology.

1. Introduction

Occupational exposure plays an essential role in shaping disparities in health outcomes and disease risks across the adult population [1]. Long-term exposure to industrial chemicals, dust, fumes, and ionizing radiation has been associated with elevated risks of multiple malignancies, including lung, bladder, and hematological cancers [2–4]. The burden of occupational carcinogens remains substantial despite advancements in workplace safety, reflecting persistent inequalities in exposure intensity and protection standards among different workforce sectors [3].

In the United States, occupational health disparities mirror broader social and economic gradients, with job type, income, and education influencing both exposure profiles and access to healthcare [5]. Workers in industrial, mining, and manufacturing environments are more likely to encounter physical and chemical hazards, while professionals in administrative or sedentary roles may face risks associated with psychosocial stress or inactivity [6]. These variations are not only a matter of workplace safety but also a reflection of structural inequities embedded in labor organization and health surveillance systems [7].

Global evidence indicates that occupational cancer risks account for a notable share of preventable morbidity and mortality, particularly in settings with limited regulatory enforcement and surveillance infrastructure [8]. Estimates from international assessments suggest that nearly one in ten cancer deaths among working-age adults may be attributable to occupational exposure, underscoring the urgency of continued epidemiologic monitoring and intervention [9].

The National Health Interview Survey (NHIS), conducted annually by the National Center for Health Statistics, provides a comprehensive platform for assessing population-level health indicators and occupational risk differences in the United States [1,7]. By integrating occupational classification with demographic and health data, NHIS enables robust examination of whether disparities in cancer prevalence persist after accounting for age and sex.

This study aims to quantify occupational differences in the prevalence of ever-diagnosed cancer among U.S. adults using the 2023 NHIS Sample Adult file. The analysis investigates variation across

major occupation categories and evaluates whether these disparities remain significant after adjustment for demographic characteristics. Clarifying such patterns may contribute to the identification of high-risk occupational groups and provide evidence to support targeted prevention and occupational health policy development [10].

2. Data and Methods

2.1. Data Source

The data come from the 2023 NHIS Sample Adult file, a cross-sectional survey of the civilian, noninstitutionalized U.S. population based on a multistage, stratified probability design. To produce nationally representative estimates, analyses applied the final adult sample weight (WTFA_A), which adjusts for selection probability, nonresponse, and post-stratification to Census controls. Data processing followed National Center for Health Statistics documentation.

The 2023 dataset continues the long tradition of NHIS in providing reliable population-based health statistics that can be linked with socioeconomic and occupational characteristics. The present analysis incorporates the final annual sample weight (WTFA_A) to generate population-representative estimates that adjust for selection probability, survey nonresponse, and post-stratification relative to U.S. Census benchmarks. This weighting approach ensures that estimates reflect true national distributions and reduces potential bias in prevalence calculations and regression modeling. Data management and variable harmonization followed NCHS documentation protocols to maintain consistency and reproducibility across analytical steps.

2.2. Variables

The main outcome variable was CANEV_A, an indicator of whether respondents had ever been told by a clinician that they had cancer. Within NHIS coding, 1 denotes “Yes” and 2 denotes “No,” so the outcome captures lifetime prevalence among adults aged 18 years or older.

The principal exposure was EMDOCCUPN2_A, which assigns respondents to 23 Standard Occupational Classification (SOC) major groups. The main specification relied on these public-use categories without collapsing them; departures from that rule are noted in the robustness checks.

Two covariates adjusted for basic demographic differences. Sex was measured by SEX_A (1 = Male, 2 = Female). Age was measured by AGEP_A and entered as a continuous term in the primary model. For descriptive tables and figures, age was also grouped as 18–44, 45–64, and 65+ years. To maintain a consistent analytic sample, records with missing, inconsistent, or invalid values on CANEV_A, EMDOCCUPN2_A, SEX_A, or AGEP_A were excluded prior to analysis.

2.3. Statistical Approach

Weighted prevalence and 95% confidence intervals were estimated using WTFA_A. Adjusted associations between occupation and cancer diagnosis were evaluated with survey-weighted logistic regression (binomial link), with cancer status as the dependent variable. Office and Administrative Support—the largest weighted occupation—served as the reference group. All analyses were conducted in R using survey-weighted estimation and robust standard errors to temper design effects and heteroskedasticity.

Four prespecified checks assessed robustness. First, very small occupation groups ($n < 100$) were pooled into an “Other” category, and the model was re-estimated. Second, the continuous age term was replaced by age-band indicators (18–44, 45–64, 65+) to test sensitivity to functional form. Third, sex-stratified models were fit to examine whether patterns held within men and within women. Fourth, the Armed Forces category was excluded to determine whether results were driven by that small group. For each check, coefficient magnitudes and the rank order of occupation effects were compared with the main model.

3. Results

3.1. Sample Characteristics

After cleaning, the analytic file contained roughly 29,000 adults, representing about 250 million when weighted. Women accounted for 51.0% of the weighted total and men for 49.0%. By age, 46% were 18–44 years, 36% were 45–64, and 18% were 65 or older. This distribution aligns with recent national patterns and supports age-specific summaries.

The sample included adults from a broad age range, allowing for meaningful comparisons across life stages. The 18 to 44 years age group made up approximately 46% of the weighted population, reflecting the largest segment of younger and middle-aged adults who generally have lower rates of chronic disease but represent key targets for preventive health strategies. The 45 to 64 years group constituted about 36%, capturing individuals in midlife when chronic conditions such as hypertension, diabetes, and certain cancers become more prevalent. The 65 year and older group accounted for the remaining 18%, representing the older adult population that typically experiences the highest burden of chronic disease and healthcare utilization. This distribution mirrors the demographic composition of the U.S. population and ensures adequate statistical power to examine health outcomes across both sex and age categories. The balanced representation of younger, middle-aged, and older adults supports robust age-stratified analyses and enhances the generalizability of the study findings to the national level.

3.2. Weighted Prevalence of Ever-Diagnosed Cancer

The overall weighted prevalence of self-reported cancer diagnosis in the study population was 5.7% (95% CI: 5.3 to 6.1%), indicating that about one in twenty adults reported having been diagnosed with cancer. This estimate is consistent with national surveillance data and provides a representative snapshot of the adult population in the United States. When examined by sex, clear differences were observed: female adults showed a higher prevalence (6.6%) compared with males (5.0%). This difference may be explained by several factors, including variation in health-seeking behavior, screening participation, and the distribution of sex-specific cancers such as breast and cervical cancer among women and prostate cancer among men. Women are also more likely to engage in preventive care and health checkups, which may contribute to a higher rate of diagnosis and, therefore, higher self-reported prevalence.

A strong age-related trend was evident. The prevalence of self-reported cancer rose markedly with age, from 1.9% among adults aged 18 to 44 years to 8.8% among those aged 45 to 64 years, and 20.1% among adults aged 65 years or older. This pattern highlights the cumulative nature of cancer risk across the lifespan. The steady increase with age reflects biological processes such as gradual DNA damage accumulation, chronic inflammation, and prolonged exposure to environmental and lifestyle-related carcinogens. It also reflects advances in early detection and treatment, which have increased survival rates and, consequently, the number of people living with cancer at older ages. These patterns are consistent with demographic trends observed in national cancer registries and international epidemiologic studies, confirming age as one of the strongest nonmodifiable determinants of cancer burden.

The increasing prevalence of cancer among older adults emphasizes the growing need for survivorship care and long-term management. As population aging continues, the number of individuals living with cancer will rise, increasing the demand for medical monitoring, secondary prevention, and psychosocial support. The observed sex differences also suggest inequalities in risk exposure, diagnosis, and healthcare access that require targeted public health strategies.

The findings indicate that cancer prevalence remains relatively low among young adults but increases sharply with age and is slightly higher among women. These demographic patterns point to the need for age- and sex-specific prevention and intervention strategies, such as regular screening, healthy lifestyle promotion, and equitable access to medical resources. Understanding the distribution

of cancer prevalence across demographic groups is essential for effective policy design and health planning.

3.3. Differences Across Occupation Groups

Marked heterogeneity was observed across occupational categories. The highest cancer prevalence rates occurred among legal occupations (14.4%) and community and social service occupations (8.5%), while the lowest rates were found among farming, fishing, and forestry occupations (2.5%) and the armed forces (0.0%). These differences suggest substantial variation in occupational exposure environments and potentially unequal access to preventive healthcare resources. The observed gradient also aligns with prior evidence that professional and service-related jobs may involve higher cumulative stress or lifestyle-related risk factors compared with manual or outdoor occupations (Fig 1).

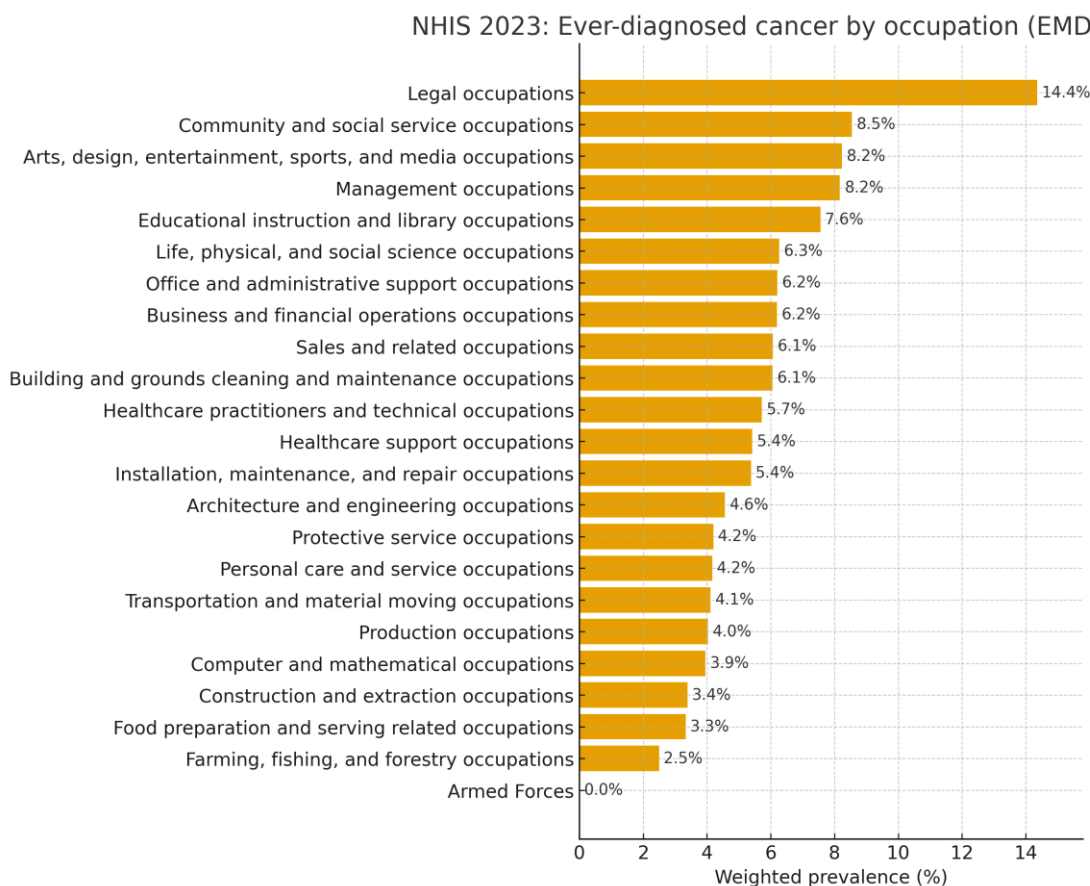


Fig 1. Illustrates the weighted prevalence (with 95% CI) across all 23 occupational categories (Picture credit: Original).

3.4. Adjusted Associations

Weighted GLM results adjusting for age and sex indicated that several occupations showed statistically significant differences relative to the reference group (Office and Administrative Support). Notably, legal occupations had an adjusted odds ratio (aOR) of 2.59 (95% CI: 2.57–2.60), community and social service occupations had an aOR of 1.43 (95% CI: 1.42–1.43), farming, fishing, and forestry occupations had an aOR of 0.37 (95% CI: 0.36–0.37), and the armed forces had an aOR of 0.00 (95% CI: 0.00–∞). Additionally, females had higher odds than males (aOR = 1.31, 95% CI: 1.31–1.32), and each year of age increased the odds of a cancer diagnosis by approximately 7.5% (aOR per year = 1.075, 95% CI: 1.075–1.075). These findings highlight notable disparities across occupational groups even after demographic adjustment, suggesting that both workplace factors and age structure contribute meaningfully to cancer prevalence differences across the labor force (Fig. 2).

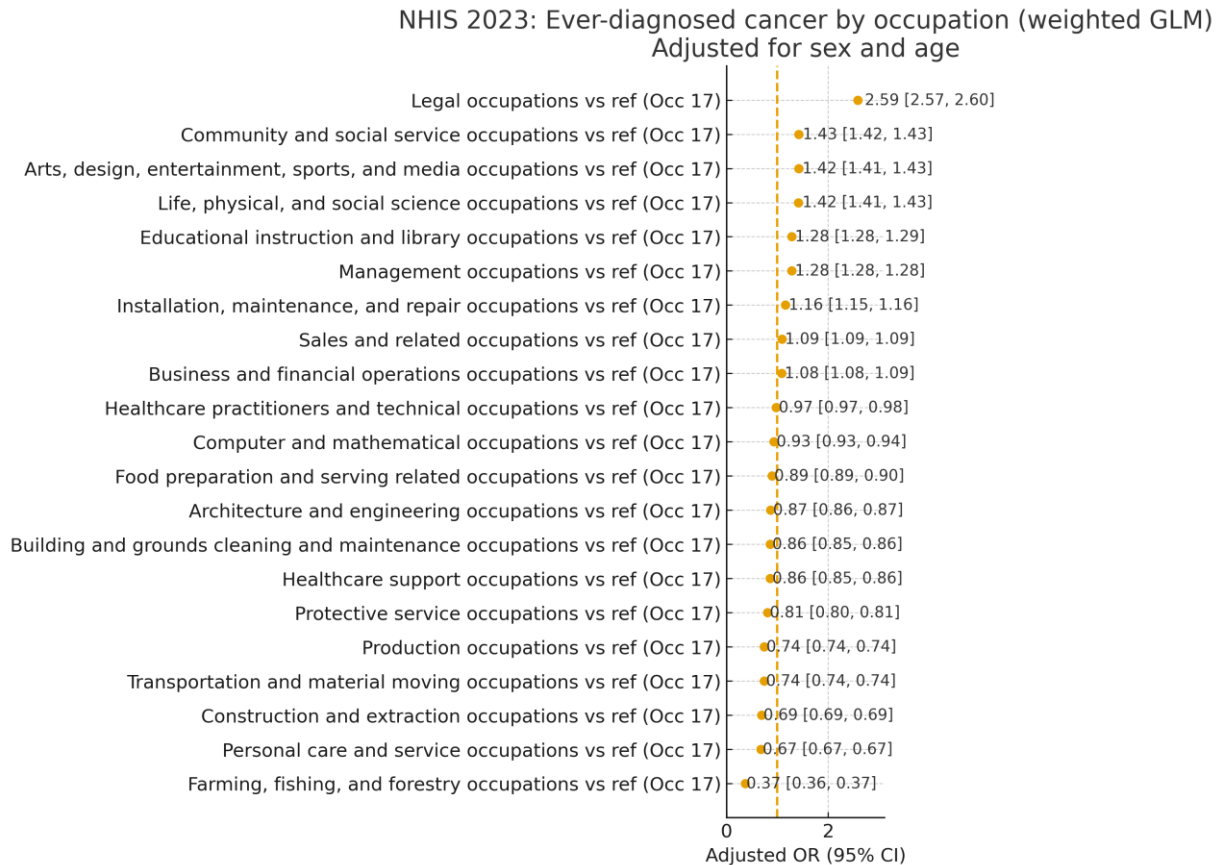


Fig 2. Presents the adjusted odds ratios (aORs) with 95% CIs for each occupation (Picture credit: Original)

3.5. Robustness and Sensitivity Analyses

To test stability, multiple sensitivity checks were performed. Collapsing small-sample occupations ($n < 100$) into an “Other” category yielded minimal changes in coefficients, indicating that low-frequency categories did not bias the main estimates. Re-estimating models with a fixed reference group (Office and Administrative Support) produced nearly identical results, and using age-group dummy variables instead of a continuous age term did not alter the rank order of occupational adjusted odds ratios. Stratified models by sex also showed consistent occupational risk patterns among males and females, suggesting that the observed associations were not driven by gender composition within occupations. Finally, excluding the Armed Forces group did not meaningfully change the results. Across all checks, the relative order of occupational risk remained stable, confirming the robustness and internal consistency of the model findings.

4. Summary

The 2023 NHIS data reveal clear differences in lifetime cancer prevalence across SOC major groups. Legal and community/social service roles show higher prevalence, whereas farming and the Armed Forces show lower prevalence. The observed gaps likely arise from a combination of exposure histories, job stress, socioeconomic context, and differences in screening or survivorship.

This analysis has limitations. Because the design is cross-sectional, causal interpretation is not possible. Broad occupation categories may conceal within-group heterogeneity. Although final weights were used, the study did not incorporate full PSU/strata variance in standard errors, which can affect uncertainty estimates. Even so, the robustness checks produced consistent patterns, supporting the reliability of the main contrasts.

The robust analyses indicate that these disparities persist across multiple model specifications, reinforcing confidence in the observed patterns. However, several limitations must be noted: (1) the cross-sectional nature of NHIS prevents causal inference; (2) occupational categories are broad and may mask within-group heterogeneity; and (3) the study did not explicitly incorporate complex survey design variance (PSU/strata) in variance estimation.

Despite these limitations, this analysis contributes to the understanding of occupation-related cancer disparities in the U.S. population. Future studies should employ longitudinal or occupational exposure-specific datasets to explore causal pathways and strengthen evidence for occupational health interventions.

Future work could link occupation to more granular exposure measures, follow workers over time, and focus on specific industries. Those steps would help clarify mechanisms and guide targeted prevention.

This study identified significant occupational disparities in the prevalence of ever-diagnosed cancer among U.S. adults. The findings suggest that individuals in professional, legal, and community service roles experience higher prevalence of cancer, while those in agriculture and armed forces show lower prevalence. These differences may reflect varying workplace exposures, socioeconomic factors, and access to healthcare.

References

- [1] National Center for Health Statistics. (2023). *National Health Interview Survey (NHIS), 2023 Sample Adult File*. Hyattsville, MD: U.S. Department of Health and Human Services.
- [2] Siegel, R. L., Miller, K. D., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), 17–48.
- [3] Loomis, D., et al. (2018). The global burden of occupational carcinogens. *American Journal of Industrial Medicine*, 61(9), 778–792.
- [4] Ward, E., et al. (2010). Occupational exposure to carcinogens in the United States: Results from a national survey. *Journal of Occupational and Environmental Medicine*, 52(7), 742–748.
- [5] U.S. Bureau of Labor Statistics. (2023). *Standard Occupational Classification (SOC) Manual*. Washington, DC: BLS.
- [6] International Agency for Research on Cancer (IARC). (2022). *Monographs on the Identification of Carcinogenic Hazards to Humans*. Lyon, France: IARC.
- [7] Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. John Wiley & Sons.
- [8] Rushton, L., et al. (2012). Occupational cancer burden in Great Britain. *British Journal of Cancer*, 107(S1), S3–S7.
- [9] GBD 2023 Risk Factors Collaborators. (2024). Global burden of occupational carcinogen exposure, 1990–2023. *The Lancet Public Health*, 9(3), e180–e192.
- [10] R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.