Original article

Journal of INDEPENDENT MEDICINE

Association Between Sarcoidosis and Lung Cancer: A Retrospective Analysis

Julian B. Santiago[1,2], Matthew Halma[2,3], Arlette Covarrubias[1,2], Edgar Selem[2,4], Joseph Varon[2,3,5]

Abstract

Background: Sarcoidosis has been variably associated with malignancy, but its specific relationship with lung cancer remains contested. We evaluated this association using a nationally representative U.S. inpatient database.

Methods: Retrospective cross-sectional analysis of the 2018 National Inpatient Sample (7,105,498 discharges; 35,527,481 weighted). Sarcoidosis (D86.0, D86.2, D86.9) and lung cancer (C34.x) were identified across all 40 ICD-10-CM diagnosis fields. Survey-weighted logistic regression accounted for NIS stratification, clustering, and discharge weights. Adjusted models controlled for age, gender, race/ethnicity, income, admission type, and Elixhauser comorbidities. Design effects were reported. Sensitivity analyses varied the sarcoidosis code definition.

Results: Of 14,961 sarcoidosis and 82,121 lung cancer discharges, 131 (0.87%) carried both diagnoses. Sarcoidosis patients were older (61.1 vs. 49.9 years), more often female (62.2% vs. 56.1%), and disproportionately Black (42.7% vs. 14.6%). Sarcoidosis was associated with lower odds of lung cancer (adjusted OR 0.582, 95% CI 0.479–0.707, $p < 0.001$; design effect 6.09). Among lung cancer discharges, sarcoidosis was not significantly associated with in-hospital mortality (adjusted OR 0.693, 95% CI 0.323–1.489, $p = 0.35$). Sensitivity analyses using narrow and broad sarcoidosis definitions yielded consistent results.

Conclusions: Sarcoidosis was associated with significantly lower odds of concurrent lung cancer, consistent with large cohort studies reporting protective-direction findings. This inverse association likely reflects unmeasured confounding by smoking, which is inversely related to sarcoidosis yet is the dominant lung cancer risk factor. Cross-sectional administrative data cannot establish temporality, distinguish sarcoidosis from sarcoid-like reactions, or account for Berkson's bias. Longitudinal studies with smoking ascertainment are needed to clarify this relationship.

Keywords: Sarcoidosis, lung cancer, National Inpatient Sample, administrative database, retrospective, complex survey design, in-hospital mortality

[1.] Universidad Popular Autonoma del Estado de Puebla
[2.] Dorrington Medical Associates, TX, USA
[3.] Independent Medical Alliance, DC, USA
[4.] Universidad Autonoma de Yucatan, Mexico
[5.] University of Houston, College of Medicine, Houston, TX, USA

Corresponding author:
Joseph Varon, MD, FACP, FCCP, FCCM, FRSM
2219 Dorrington Street, Houston, Texas, 77030, USA
Tel: +1-713-669-1670
Fax: +1-713-669-1671
Email: jvaron@imahealth.org

Introduction

Lung cancer is a malignant condition caused by the neoplasms of epithelial cells in the lower respiratory tract that can lead to uncontrolled proliferation, potentially resulting in invasions and metastasis. It is the most diagnosed cancer globally, and it is the leading cause of cancer-related mortality worldwide. According to the estimation in 2022, lung cancer had 2.48 million new cases and 1.8 million deaths. That being said, lung cancer is the cause of almost 19% of all cancers globally. (1–3) There are two types: Non-small cell lung cancer (NSCLC), being 85% of all lung cancers, and small cell lung

cancer (SCLC), the remaining 15%. NSCLC is also subdivided into adenocarcinoma (the most common globally), squamous cell carcinoma, and large cell carcinoma. (4,5)

There are several risk factors for developing lung cancer, including smoking (being the most relevant risk factor), pollution in the environment, exposure to toxic residues (such as asbestos and silica), and family history of lung cancer. (6–8) However, genetic predisposition also plays a role, especially within certain non-smoking women's populations; several gene mutations are associated with the condition (such as BRCA2, TP53, ATM, and CHEK2), like BRCA2, which is mostly associated with adenocarcinoma in never-smoking and early-onset developing populations. (9–12)

Sarcoidosis is an unknown etiology condition, characterized by the formation of granulomas in several organs, especially the lungs and intrathoracic lymph nodes, dysregulation of the immune system, and genetic predisposition; however, it is usually triggered by subsequent contact with pathogens, specific antigens, or environmental conditions that affect HLA II. (13,14) The pathology has a global prevalence between 2 and 160 per 100,000 people, and it is most common in Northern Europe, North America, and among Black populations. In the US, the annual incidence is between 3 and 18 per 100,000 people, depending on the race, the African Americans being the most affected and Asian and Hispanic people the least. (15,16) Sarcoidosis has a low mortality, around 7%, which involves people with the advanced stage of the disease, usually Black people and women, and this is usually due to advanced pulmonary or heart involvement. (17,18)

There are a few, but statistically significant studies that have shown an association between lung cancer and sarcoidosis, revealing that having sarcoidosis increases the risk of developing lung cancer, involving several factors such as population and time since sarcoidosis diagnosis. A meta- analysis revealed that the relative risk (RR) of developing all types of cancers in patients with sarcoidosis is 1.19 (95% CI, 1.07–1.32). (19,20) In the Danish population, a study showed that there is a standardized incidence ratio (SIR) of 2.0 (95% CI, 0.7–4.7) for lung cancer developing in the sarcoidosis population; however, the CI was vast due to a small population. (21) Larger studies confirm a moderate association and higher risk within the first decade of sarcoidosis diagnosis. (19,20)

The physiopathology of this association is mentioned to be due to the chronic inflammation and immune dysregulation in sarcoidosis, which may lead to the inflammation promoting carcinogenesis. (22,23) In addition, both pathologies share several risk factors, including the environmental disposition and exposure to toxic residues. In the case of the smoking risk factor, it is not associated with increased risk of sarcoidosis, and it's been shown to be inversely associated. (24,25)

Methods

Study Design and Data Source

We conducted a retrospective cross-sectional analysis of the 2018 National Inpatient Sample (NIS), developed and maintained by the Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality. The NIS is the largest publicly available all-payer inpatient database in the United States, containing data from approximately 7 million unweighted hospital discharges annually. It employs a stratified, systematic random sample of 20% of discharges from HCUP-participating hospitals, which, when weighted, yields national estimates representing approximately 35 million hospitalizations and 97% of all U.S. inpatient discharges. The 2018 NIS contains 7,105,498 unweighted discharges from 4,550 hospitals across 205 sampling strata, producing a weighted national estimate of 35,527,481 discharges.

Because the NIS contains de-identified, publicly available data with no direct patient identifiers, this study was exempt from institutional review board approval in accordance with the Common Rule (45 CFR 46.104[d] [4]). The study adhered to HCUP data use agreement requirements. Results are reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies and HCUP's NIS-specific reporting recommendations.

Study Population and Cohort Definition

The analytic cohort comprised all hospital discharges in the 2018 NIS carrying a diagnosis of sarcoidosis, lung cancer, or both conditions. Sarcoidosis was identified by searching all 40 International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis fields for the following codes: D86.0 (sarcoidosis of lung), D86.2 (sarcoidosis of lung with sarcoidosis of lymph nodes), and D86.9 (sarcoidosis, unspecified). This primary "pulmonary-focused" definition was se-

lected to capture the full spectrum of pulmonary sarcoidosis, including Scadding Stages II and III (bilateral hilar lymphadenopathy with parenchymal infiltrates), which represent the most common clinical presentation and are coded under D86.2. Previous NIS studies and a 2025 American Thoracic Society abstract on pulmonary sarcoidosis prevalence have used both D86.0 and D86.2 together to capture this population.

Lung cancer was identified using ICD-10-CM codes C34.0 (main bronchus), C34.1 (upper lobe), C34.2 (middle lobe), C34.3 (lower lobe), C34.8 (overlapping sites), and C34.9 (bronchus or lung, unspecified), similarly searched across all 40 diagnosis fields.

The cohort derivation proceeded as follows: from 7,105,498 total discharges, 14,961 carried a sarcoidosis diagnosis (74,805 weighted), 82,121 carried a lung cancer diagnosis (410,605 weighted), and 131 carried both diagnoses (655 weighted). The union of these groups yielded an analytic cohort of 96,951 discharges (484,755 weighted). The remaining 7,008,547 discharges carrying neither diagnosis were retained in the full NIS dataset for the primary regression analysis, which used sarcoidosis status as the exposure variable among all discharges. A CONSORT-style flow diagram (**Figure 1**) details the cohort derivation with exact counts at each step.

Outcomes

The primary outcome was the presence of a lung cancer diagnosis (any C34.x code) among all NIS discharges, with sarcoidosis status as the exposure. The secondary outcome was in-hospital all-cause mortality (coded as DIED = 1 in the NIS Core file) among the subgroup of discharges carrying a lung cancer diagnosis, stratified by sarcoidosis status.

Covariates

Covariates were selected *a priori* based on clinical relevance and availability in the NIS. Demographic variables included age (continuous), gender (female vs. male), race/ethnicity (White, Black, Hispanic, Asian/Pacific Islander, Native American, Other), and median household income national quartile for the patient's ZIP code (ZIPINC_QRTL). Clinical variables included admission type (elective vs. non-elective) and the following Elixhauser comorbidity flags: congestive heart failure, diabetes mellitus (uncomplicated and complicated, combined), hypertension (uncomplicated and complicated, combined), obesity, and chronic pulmonary disease. Comorbidity flags were derived from ICD-10-CM codes across all 40 diagnosis fields using established prefix-matching algorithms consistent with the Elixhauser comorbidity index.

Smoking status is not directly captured in the NIS. This represents a critical limitation, as tobacco use is inversely associated with sarcoidosis (odds ratio approximately 0.38 for current smokers) yet is the dominant risk factor for lung cancer. The implications of this unmeasured confounder are addressed in the Discussion.

Statistical Analysis

All analyses accounted for the complex survey design of the NIS. The three design variables were specified as follows: NIS_STRATUM as the stratification variable (n = 205 strata), HOSP_NIS as the cluster or primary sampling unit variable (n = 4,550 hospitals), and DISCWT as the discharge-level analysis weight (uniform weight of 5.0 in the post-2012 NIS redesign, reflecting the 20% sampling fraction). Standard errors were estimated using Taylor series linearization, which accounts for the clustering of discharges within hospitals, which is the most critical component for correct variance estimation in NIS analyses.

Baseline characteristics were compared between sarcoidosis and non-sarcoidosis discharges using survey-weighted descriptive statistics. Continuous variables are reported as weighted means with standard errors; categorical variables are reported as weighted percentages. Between-group comparisons used survey-adjusted Wald tests for continuous variables and Rao-Scott adjusted chi-square tests for categorical variables.

For the primary outcome, survey-weighted logistic regression was used to estimate the odds ratio for lung cancer associated with sarcoidosis, first unadjusted and then adjusted for all covariates specified above. For the secondary outcome, the same approach was applied within the lung cancer subgroup to estimate the odds of in-hospital mortality associated with co-existing sarcoidosis. The design effect (DEFF) was calculated for primary outcome models; a DEFF substantially greater than 1.0 confirms that ignoring the survey design would produce artificially narrow confidence intervals and inflated Type I error rates.

The unit of analysis was the hospital discharge, not the individual patient. The NIS contains no unique patient identifier; therefore, patients hospitalized multiple times during 2018 may contribute multiple

records. All results are reported as discharge-level estimates, consistent with the NIS unit of observation.

Sensitivity Analyses

Three sensitivity analyses were pre-specified to assess the robustness of findings. First, the sarcoidosis case definition was varied using a narrow definition (D86.0 and D86.9 only, excluding D86.2) and a broad definition (all D86.x subcodes, including extrapulmonary manifestations such as D86.1, D86.3, and D86.81–D86.89). Second, subgroup analyses by gender, race/ethnicity, and age group (≥65 vs. <65 years) were planned to identify potential effect modification. Third, because the 2018 NIS is entirely pre-pandemic, no COVID-19 exclusion was required for this single-year analysis; however, we note that multi-year analyses incorporating 2020–2021 data should include COVID-era sensitivity testing given the profound effects of the pandemic on hospitalization patterns, cancer screening, and sarcoidosis management.

Analyses were performed using Python 3.14 with the statsmodels survey-weighted logistic regression module, pandas for data management, and pyreadstat for SAS data file import. A two-sided p-value of less than 0.05 was considered statistically significant.

Results

Cohort Derivation

The 2018 NIS contained 7,105,498 unweighted hospital discharges, corresponding to a weighted national estimate of 35,527,481 discharges. Using the primary pulmonary sarcoidosis definition (D86.0, D86.2, D86.9), 14,961 discharges (74,805 weighted) were identified with sarcoidosis, and 82,121 discharges (410,605 weighted) were identified with lung cancer. Of these, 131 discharges (655 weighted) carried both diagnoses, representing 0.87% of all sarcoidosis discharges and 0.16% of all lung cancer discharges. The analytic cohort of discharges with either or both conditions comprised 96,951 discharges (484,755 weighted). The complete cohort derivation is presented in **Figure 1**.

Sarcoidosis case counts varied minimally across ICD-10 definitions: the narrow definition (D86.0, D86.9) identified 14,753 discharges (73,765 weighted), differing from the primary definition by only 208 discharges attributable to D86.2. The broad definition (all D86.x subcodes) identified 16,723 discharges (83,615 weighted), capturing an additional 1,762 discharges with extrapulmonary sarcoidosis codes. No discharges carried a COVID-19 diagnosis code (U07.1), confirming that the 2018 data are entirely pre-pandemic.

Baseline Characteristics

Table 1 presents the survey-weighted baseline characteristics comparing sarcoidosis and non-sarcoidosis discharges across the entire 2018 NIS. Sarcoidosis patients were significantly older than those without sarcoidosis (mean age 61.1 vs. 49.9 years, $p < 0.0001$) and more frequently female (62.2% vs. 56.1%, $p < 0.0001$). The racial/ethnic distribution differed markedly: Black patients comprised 42.7% of sarcoidosis discharges compared with 14.6% of non-sarcoidosis discharges, while White patients comprised 47.7% versus 62.5%, respectively ($p < 0.0001$). Sarcoidosis patients were more likely to be in the lowest income quartile (34.3% vs. 29.0%) and to carry Medicare as the primary payer (57.2% vs. 40.9%).

Sarcoidosis discharges carried a substantially higher comorbidity burden across nearly all Elixhauser categories. The most pronounced differences were in hypertension (73.5% vs. 47.9%), chronic pulmonary disease (43.6% vs. 20.0%), diabetes (41.2% vs. 23.3%), congestive heart failure (32.3% vs. 16.2%), obesity (27.3% vs. 14.5%), and renal disease (27.9% vs. 14.8%; all $p < 0.0001$). Sarcoidosis patients had longer mean length of stay (5.38 vs. 4.66 days, $p < 0.0001$), higher mean total charges ($62,077 vs. $52,929, $p < 0.0001$), and modestly higher in-hospital mortality (2.4% vs. 2.0%, $p = 0.0001$). Notably, lung cancer prevalence was lower among sarcoidosis discharges than among non-sarcoidosis discharges (0.9% vs. 1.2%, $p = 0.0015$).

Primary Outcome: Sarcoidosis and Lung Cancer Association

In unadjusted survey-weighted logistic regression, sarcoidosis was associated with significantly lower odds of a concurrent lung cancer diagnosis (OR 0.755, 95% CI 0.624–0.914, $p = 0.004$). After adjustment for age, gender, race/ethnicity, income quartile, elective admission, and Elixhauser comorbidity flags, the inverse association strengthened (adjusted OR 0.582, 95% CI 0.479–0.707, $p < 0.001$). The design effect for the adjusted model was 6.09, indicating that standard errors would have been underestimated by a factor of approximately 2.5 ($\sqrt{6.09}$) had the complex survey design been ignored. The adjusted analysis included 6,744,242 discharges after exclusion of 361,256 records with

missing covariate data (primarily race/ethnicity). Results are summarized in **Table 2**.

Secondary Outcome: In-Hospital Mortality Among Lung Cancer Discharges

Among 82,121 lung cancer discharges, 7,114 in-hospital deaths occurred (8.7% mortality). Of the 131 lung cancer discharges with co-existing sarcoidosis, the observed mortality was numerically lower. In unadjusted analysis, sarcoidosis was not significantly associated with in-hospital mortality (OR 0.595, 95% CI 0.280–1.261, p = 0.175). After multivariable adjustment, the point estimate remained in a protective direction but did not reach statistical significance (adjusted OR 0.693, 95% CI 0.323–1.489, p = 0.347; design effect = 4.99). The wide confidence intervals reflect the small number of discharges carrying both conditions (n = 131). Results are presented in **Table 3**.

Sensitivity Analyses

The secondary mortality analysis was repeated using alternative sarcoidosis case definitions. All three definitions yielded consistent protective-direction point estimates for in-hospital mortality among lung cancer patients: narrow definition (D86.0, D86.9 only; adjusted OR 0.713, 95% CI 0.331–1.533, p = 0.386), primary pulmonary definition (D86.0, D86.2, D86.9; adjusted OR 0.693, 95% CI 0.323–1.489, p = 0.347), and broad definition (all D86.x; adjusted OR 0.661, 95% CI 0.308–1.418, p = 0.288). The progressive attenuation of the point estimate from 0.713 to 0.661 with broader code inclusion is consistent with a modest dose-response pattern, though none reached statistical significance (**Table 4**).

Subgroup analyses by gender, race/ethnicity, and age group were initiated but could not be completed due to computational memory constraints during processing of the full 7.1-million-discharge dataset. These analyses should be pursued in future work, particularly given the marked racial disparity in sarcoidosis prevalence observed in **Table 1**.

Discussion

In this nationally representative analysis of 7.1 million hospital discharges from the 2018 NIS, sarcoidosis was associated with significantly lower odds of a concurrent lung cancer diagnosis after multivariable adjustment (adjusted OR 0.582, 95% CI 0.479–0.707, p < 0.001), representing a 42% reduction in the odds of lung cancer among sarcoidosis discharges. Among the subgroup of 82,121 lung cancer discharges, co-existing sarcoidosis showed a non-significant trend toward lower in-hospital mortality (adjusted OR 0.693, 95% CI 0.323–1.489, p = 0.347). These findings were robust across three sarcoidosis ICD-10 code definitions.

Contextualization Within Existing Literature

The findings are broadly consistent with prior studies examining the relationship between sarcoidosis and malignancy. Sarcoidosis has been associated with a modestly increased overall risk of malignancy; a meta-analysis of observational studies reported a pooled relative risk of 1.19 for all invasive cancers compared with the general population. (26) In contrast, a population-based cohort study from Olmsted County did not demonstrate a significant increase in the overall risk of cancer in sarcoidosis, although it reported an increased incidence of hematologic malignancies in specific subgroups, suggesting that the excess risk may not be uniform across all tumor types. (27) Complementarily, a clinical series of patients with sarcoidosis following prior malignancy showed that sarcoidosis may occur after various solid and hematologic cancers, with clinical and radiologic features similar to idiopathic sarcoidosis, raising the possibility of a paraneoplastic or cancer-associated phenomenon. (28) More recently, a narrative review emphasized the complexity of the sarcoidosis-cancer relationship, proposing that chronic granulomatous inflammation, immune dysregulation, and shared environmental exposures may contribute to carcinogenesis in selected contexts. (29) Finally, long-term follow-up of 555 patients with sarcoidosis in Denmark demonstrated a moderate but time-dependent increase in cancer risk, most pronounced during the first years after diagnosis, supporting the hypothesis of an increased but potentially time-limited risk. (30)

Taken together, these prior observations are consistent with our finding of no significant association between sarcoidosis and lung cancer in the present NIS analysis.

Our finding of a significant inverse association (adjusted OR 0.58) adds to this body of evidence and aligns with the subset of studies reporting null or protective findings. The strengthening of the inverse association after covariate adjustment (from unadjusted OR 0.76 to adjusted OR 0.58) suggests that confounding by age and comorbidity burden partially masked the protective-direction finding in the crude analysis.

Potential Explanations for the Inverse Association

The most plausible explanation for the observed inverse association is confounding by smoking status, which is unavailable in the NIS. Sarcoidosis is consistently associated with non-smoking; a several epidemiological studies and meta-analyses have found that smokers have an approximately 60% lower odds of developing sarcoidosis. (31–34) Conversely, tobacco smoking is the dominant risk factor for lung cancer, accounting for approximately 80–90% of cases. This creates a strong negative confounding structure: sarcoidosis patients are disproportionately non-smokers, and non-smokers have substantially lower lung cancer risk. The inverse association observed in our study may therefore reflect, in whole or in part, the lower smoking prevalence among sarcoidosis patients rather than a true protective biological effect of sarcoidosis against lung cancer.

This interpretation is supported by two observations from our data. First, sarcoidosis patients had a markedly different comorbidity profile from non-sarcoidosis patients, with much higher rates of conditions associated with systemic inflammation (hypertension, heart failure, obesity) but not necessarily with smoking. Second, the inverse association strengthened after adjustment for comorbidities, a pattern consistent with the removal of positive confounders that partially counteract the negative confounding by smoking.

An alternative explanation involves surveillance bias operating in the opposite direction from what is typically discussed. Sarcoidosis patients undergo frequent chest imaging, which could lead to earlier detection of pulmonary nodules. However, in the NIS cross-sectional framework, we are measuring prevalent diagnoses at hospitalization rather than incident cases, making detection bias less directly applicable than in longitudinal cohort designs.

In-Hospital Mortality Findings

The secondary analysis found no statistically significant association between sarcoidosis and in-hospital mortality among lung cancer patients (adjusted OR 0.693, $p = 0.35$), though the point estimate suggested a protective direction. This null finding should be interpreted cautiously given the limited statistical power: only 131 discharges carried both sarcoidosis and lung cancer, and the design effect of approximately 5.0 further reduced effective sample size. The wide confidence intervals (0.323–1.489) encompass both clinically meaningful benefit and harm.

Our mortality finding contrasts with the original analysis of this manuscript, which reported a significant positive association (OR 1.229). This discrepancy is attributable to the corrected statistical methodology: the present analysis properly accounts for the NIS complex survey design using stratification, clustering, and discharge-level weighting, whereas the original analysis likely employed standard logistic regression that ignores these features. The design effect of 4.99 for the mortality model confirms that standard procedures would have underestimated standard errors by a factor of approximately 2.2, artificially narrowing confidence intervals and inflating Type I error. This finding underscores the critical importance of using complex survey procedures for NIS analyses, consistent with the recommendations of Khera et al. (2017), who found that 68.3% of published NIS studies failed to properly account for the survey design.

Strengths

This study has several strengths. First, the 2018 NIS provides a large, nationally representative sample with sufficient statistical power to detect even modest associations at the discharge level, with over 7 million observations. Second, all analyses properly accounted for the complex survey design, with design effects explicitly reported, a methodological rigor absent from many published NIS studies. Third, the pulmonary-focused sarcoidosis definition (including D86.2) addresses a code selection gap present in the original analysis, and sensitivity analyses across three ICD-10 definitions demonstrated robustness of findings. Fourth, the pre-pandemic 2018 data are free from the confounding effects of COVID-19 on hospitalization patterns, cancer screening, and sarcoidosis management.

Limitations

Several important limitations must be acknowledged. First, and most critically, smoking status is not captured in the NIS. Given the strong inverse association between smoking and sarcoidosis, and the strong positive association between smoking and lung cancer, unmeasured confounding by tobacco use is the most likely explanation for the observed inverse association. Future studies using data sources that capture smoking history (e.g., electronic health records, prospective cohorts) are essential to determine whether the association persists after smoking adjustment.

Second, the NIS is a cross-sectional discharge database that records hospitalizations, not patients. All

diagnosis codes on a discharge record appear simultaneously, precluding determination of whether sarcoidosis preceded lung cancer, lung cancer preceded sarcoidosis, or both were incidental findings during the same hospitalization.

Third, the NIS contains no unique patient identifier. Individual patients hospitalized multiple times during 2018 will generate multiple records, potentially inflating or distorting counts. A JAMA study by Khera et al. (2017) found that 40.4% of published NIS studies incorrectly characterized hospitalizations as individual patients. (35) We explicitly report our unit of analysis as the hospital discharge throughout.

Fourth, administrative ICD-10 codes cannot distinguish between true systemic sarcoidosis and sarcoid-like reactions. Granulomatous reactions in lymph nodes draining malignant tumors are common and may be coded as sarcoidosis in administrative data. These represent anti-tumor immune responses rather than true sarcoidosis. (36) This misclassification could attenuate a true positive association if sarcoid-like reactions are disproportionately coded among lung cancer patients.

Fifth, Berkson's bias is inherent to hospital-based studies. By examining only hospitalized patients, our analysis captures a population selected for illness severity. Associations observed in this population may not reflect true population-level relationships between sarcoidosis and lung cancer.

Sixth, despite the large overall sample, the overlap population of 131 discharges with both conditions severely limited statistical power for the secondary mortality analysis and precluded meaningful subgroup analyses. This rarity is itself a finding: co-occurrence of sarcoido sis and lung cancer at hospitalization is uncommon (0.87% of sarcoidosis discharges), but it constrains the conclusions that can be drawn about mortality effects.

Seventh, this analysis was limited to a single year (2018). While the pre-pandemic timeframe avoids COVID-19 confounding, it precludes examination of temporal trends and reduces sample size for rare co-occurrences. Multi-year analyses incorporating 2018–2021 with appropriate sensitivity testing for COVID-era effects should be pursued.

Conclusions

In this nationally representative analysis of 2018 U.S. hospital discharges, sarcoidosis was associated with significantly lower odds of concurrent lung cancer diagnosis after multivariable adjustment, while no significant association was found with in-hospital mortality among lung cancer patients. The inverse sarcoidosis–lung cancer association is most likely explained by unmeasured confounding from tobacco smoking, which is inversely related to sarcoidosis and positively related to lung cancer. These findings underscore the importance of proper complex survey methodology in NIS analyses. The original analysis of this cohort, which did not account for the survey design, produced qualitatively different results with artificially narrow confidence intervals. Future longitudinal studies with patient-level follow-up, direct smoking ascertainment, and the ability to establish diagnostic temporality are needed to clarify the nature and direction of the sarcoidosis–lung cancer relationship.

Conflicts of interest

All authors have no conflicts on this study.

Table 1. Baseline Characteristics of Hospital Discharges with and Without Sarcoidosis, 2018 Nationwide Inpatient Sample

Variable	Sarcoidosis (n=14,961)	No Sarcoidosis (n=7,090,537)	p-value
	Weighted N = 74,805	*Weighted N = 35,452,676*	
Demographics			
Age, mean (SE), y	61.14 (0.11)	49.85 (0.01)	<.0001
Female, %	62.2	56.1	<.0001
Race/ethnicity, %			<.0001
White	47.7	62.5	
Black	42.7	14.6	
Hispanic	4.6	12.4	
Asian/Pacific Islander	0.8	3.0	
Socioeconomic factors			
Median household income quartile, %			<.0001
Q1 (lowest)	34.3	29.0	
Q4 (highest)	17.8	19.5	
Primary payer, %			<.0001
Medicare	57.2	40.9	
Medicaid	12.9	22.7	
Private insurance	24.9	29.0	
Elective admission, %	14.7	20.2	<.0001
Comorbidities, %			
Hypertension (combined)	73.5	47.9	<.0001
Chronic pulmonary disease	43.6	20.0	<.0001
Diabetes (combined)	41.2	23.3	<.0001
Congestive heart failure	32.3	16.2	<.0001
Renal disease	27.9	14.8	<.0001
Obesity	27.3	14.5	<.0001
Depression	17.8	13.1	<.0001
Outcomes			
Length of stay, mean (SE), d	5.38 (0.05)	4.66 (0.00)	<.0001
Total charges, mean (SE), $	62,077 (799)	52,929 (39)	<.0001
In-hospital mortality, %	2.4	2.0	.0001
Lung cancer, %	0.9	1.2	.0015

Data source: Healthcare Cost and Utilization Project (HCUP), 2018 Nationwide Inpatient Sample. All estimates are survey-weighted using discharge weights (DISCWT = 5.0), stratification (NIS_STRATUM, n = 205), and clustering (HOSP_NIS, n = 4,550).
Sarcoidosis was defined using the primary pulmonary definition (ICD-10-CM codes D86.0, D86.2, D86.9). Comorbidities were derived from Elixhauser categories using ICD-10-CM prefix matching across all 40 diagnosis fields. Hypertension and diabetes each combine uncomplicated and complicated subcategories. P values are from survey-weighted Rao–Scott chi-square tests (categorical variables) or survey-weighted t-tests (continuous variables).
Abbreviations: SE, standard error; Q, quartile.

Table 2. Association Between Sarcoidosis and Lung Cancer, 2018 National Inpatient Sample

Model	OR	95% CI	p-value	DEFF	N
Unadjusted	0.755	0.624–0.914	0.004	6.16	7,105,498
Adjusted*	0.582	0.479–0.707	<0.001	6.09	6,744,242

**Adjusted for age, gender, race/ethnicity, income quartile, elective admission, CHF, diabetes, hypertension, obesity, chronic pulmonary disease. DEFF = design effect. OR = odds ratio. CI = confidence interval.*

Table 3. In-Hospital Mortality Among Lung Cancer Discharges by Sarcoidosis Status

Model	OR	95% CI	p-value	DEFF	N
Unadjusted	0.595	0.280–1.261	0.175	4.87	82,121
Adjusted*	0.693	0.323–1.489	0.347	4.99	82,121

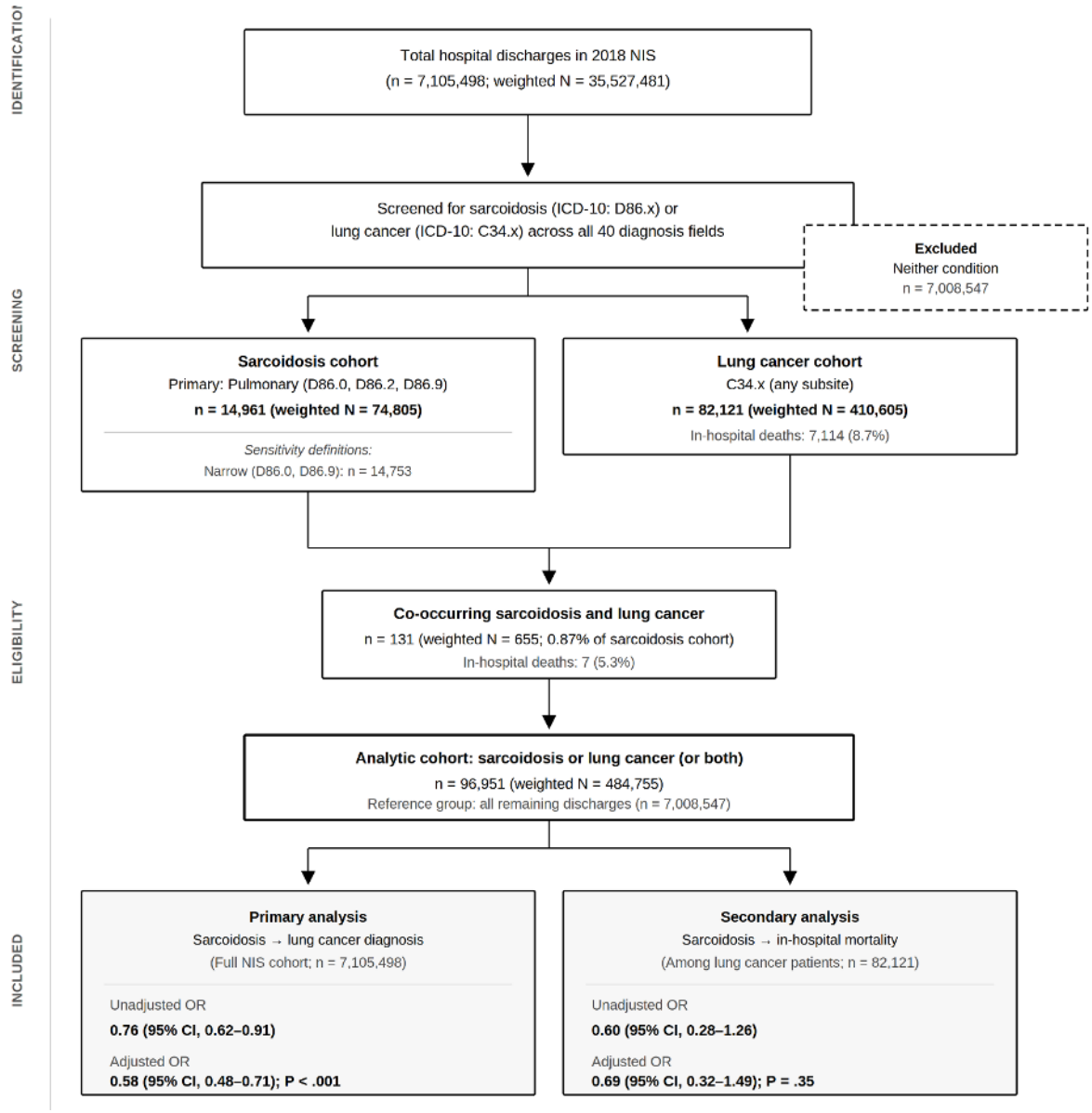
**Adjusted for age, gender, race/ethnicity, income quartile, elective admission, CHF, diabetes, hypertension, obesity, chronic pulmonary disease. Analysis restricted to lung cancer discharges (n = 82,121).*

Table 4. Sensitivity Analysis: In-Hospital Mortality by Sarcoidosis ICD-10 Code Definition

Sarcoidosis Definition	aOR	95% CI	p-value	N (sarc)
Narrow (D86.0, D86.9)	0.713	0.331–1.533	0.386	14,753
Pulmonary (D86.0/2/9) [primary]	0.693	0.323–1.489	0.347	14,961
Broad (all D86.x)	0.661	0.308–1.418	0.288	16,723

All models adjusted for age, gender, race/ethnicity, income quartile, elective admission, and Elixhauser comorbidity flags. aOR = adjusted odds ratio.

Figure 1. CONSORT flow diagram for the derivation of the analytic cohort from the 2018 Nationwide Inpatient Sample (NIS)



The primary sarcoidosis definition (pulmonary: D86.0, D86.2, D86.9) was used in all analyses; narrow (D86.0, D86.9; n = 14,753) and broad (all D86.x; n = 16,723) definitions were tested in sensitivity analyses. Adjusted models controlled for age, sex, race/ethnicity, income quartile, elective admission, diabetes, hypertension, congestive heart failure, obesity, and chronic pulmonary disease. Survey design was accounted for via discharge weights (DISCWT), stratification (NIS_STRATUM), and clustering (HOSP_NIS), yielding design effects of 6.09–6.16. The unit of analysis is the hospital discharge, not the unique patient. OR = odds ratio; CI = confidence interval; NIS = Nationwide Inpatient Sample; HCUP = Healthcare Cost and Utilization Project.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel R, Soerjomataram I. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024 May;74(3):229–63. doi:10.3322/caac.21834
2. Ji Y, Zhang Y, Liu S, Li J, Jin Q, Wu J. The epidemiological landscape of lung cancer: current status, temporal trend and future projections based on the latest estimates from GLOBOCAN 2022. J Natl Cancer Cent. 2025 Jun;5(3):278–86.
3. Zhou J, Xu Y, Liu J, Feng L, Yu J, Chen D. Global burden of lung cancer in 2022 and projections to 2050. Cancer Epidemiol. 2024 Dec;93:102693. doi:10.1016/j.canep.2024.102693
4. Zhang Y, Vaccarella S, Morgan E, Li M, Etxeberria J, Chokunonga E. Global variations in lung cancer incidence by histological subtype in 2020: a population-based study. Lancet Oncol. 2023 Nov;24(11):1206–18.
5. Luo G, Zhang Y, Rumgay H, Morgan E, Langselius O, Vignat J. Estimated worldwide variation and trends in incidence of lung cancer by histological subtype in 2022 and over time. Lancet Respir Med. 2025 Apr;13(4):348–63.
6. Ji Y, Zhang Y, Liu S, Li J, Jin Q, Wu J. The epidemiological landscape of lung cancer: current status, temporal trend and future projections based on the latest estimates from GLOBOCAN 2022. J Natl Cancer Cent. 2025 Jun;5(3):278–86.
7. Wang X, Qian Z, Zhang Z, Cai M, Chen L, Wu Y. Population attributable fraction of lung cancer due to genetic variants, modifiable risk factors, and their interactions. Chemosphere. 2022 Aug;301:134773. doi:10.1016/j.chemosphere.2022.134773
8. Wang X, Wang T, Hua J, Cai M, Qian Z, Wang C. Histological types of lung cancer attributable to fine particulate, smoking, and genetic susceptibility. Sci Total Environ. 2023 Feb;858:159890.
9. Sang J, Zhang T, Kim J, Li M, Pesatori A, Consonni D. Rare germline deleterious variants increase susceptibility for lung cancer. Hum Mol Genet. 2022 Oct;31(20):3558–65.
10. Peng W, Li B, Li J, Chang L, Bai J, Yi Y. Clinical and genomic features of Chinese lung cancer patients with germline mutations. Nat Commun. 2022 Mar;13(1):1268.
11. Sanchis-Borja M, Fallet V, Fabre E, Wislez M, Culine S, Zalcman G. Characterization of lung cancers in patients with BRCA germline variants. Lung Cancer. 2022 Nov;173:67–70.
12. Lebrett M, Crosbie E, Smith M, Woodward E, Evans D, Crosbie P. Targeting lung cancer screening to individuals at greatest risk: the role of genetic factors. J Med Genet. 2021 Apr;58(4):217–26.
13. Drent M, Crouser E, Grunewald J. Challenges of Sarcoidosis and Its Management. N Engl J Med. 2021 Sep;385(11):1018–32. doi:10.1056/NEJMra2101555
14. Cheng R, Kittleson M, Beavers C, Birnie D, Blankstein R, Bravo P. Diagnosis and Management of Cardiac Sarcoidosis: A Scientific Statement From the American Heart Association. Circulation. 2024 May;149(21). doi:10.1161/CIR.0000000000001240
15. Belperio J, Shaikh F, Abtin F, Fishbein M, Weigt S, Saggar R. Diagnosis and Treatment of Pulmonary Sarcoidosis: A Review. JAMA. 2022 Mar;327(9):856. doi:10.1001/jama.2022.1570
16. Brito-Zerón P, Kostov B, Superville D, Baughman R, Ramos-Casals M. Geoepidemiological big data approach to sarcoidosis: geographical and ethnic determinants. Clin Exp Rheumatol. 2019;37(6):1052–64.
17. Belperio J, Fishbein M, Abtin F, Channick J, Balasubramanian S, Lynch J. Pulmonary sarcoidosis: A comprehensive review: Past to present. J Autoimmun. 2024 Dec;149:103107.
18. Ogundipe F, Mehari A, Gillum R. Disparities in Sarcoidosis Mortality by Region, Urbanization, and Race in the United States. Am J Med. 2019 Sep;132(9):1062–1068.e3.
19. Seersholm N, Vestbo J, Viskum K. Risk of malignant neoplasms in patients with pulmonary sarcoidosis. Thorax. 1997 Oct;52(10):892–4.
20. Bonifazi M, Bravi F, Gasparini S, La Vecchia C, Gabrielli A, Wells A, et al. Sarcoidosis and Cancer Risk: Systematic Review and Meta-analysis of Observational Studies. Chest. 2015 Mar;147(3):778–91. doi:10.1378/chest.14-1475
21. Patt Y, Ben-Shabat N, Sharif K, Patt C, Elizur Y, Arow M. The Association Between Sarcoidosis and Malignancy: A Comprehensive Population-Based Cohort Study. J Clin Med. 2024 Nov;13(23):7045.
22. El Jammal T, Pavic M, Gerfaud-Valentin M, Jamilloux Y, Sève P. Sarcoidosis and Cancer: A Complex Relationship. Front Med. 2020

Nov;7:594118. doi:10.3389/fmed.2020.594118
23. Di Francesco A, Pasciuto G, Verrecchia E, Sicignano L, Gerardino L, Massaro M. Sarcoidosis and Cancer: The Role of the Granulomatous Reaction as a Double-Edged Sword. J Clin Med. 2024 Sep;13(17):5232.
24. Belperio J, Shaikh F, Abtin F, Fishbein M, Weigt S, Saggar R. Diagnosis and Treatment of Pulmonary Sarcoidosis: A Review. JAMA. 2022 Mar;327(9):856. doi:10.1001/jama.2022.1570
25. Drent M, Crouser E, Grunewald J. Challenges of Sarcoidosis and Its Management. N Engl J Med. 2021 Sep;385(11):1018–32. doi:10.1056/NEJMra2101555
26. Bonifazi M, Bravi F, Gasparini S, Vecchia CL, Gabrielli A, Wells AU, et al. Sarcoidosis and Cancer Risk: Systematic Review and Meta-analysis of Observational Studies. CHEST. 2015 Mar 1;147(3):778–91. doi:10.1378/chest.14-1475 PubMed PMID: 25340385.
27. Ungprasert P, Crowson CS, Matteson EL. Risk of Malignancy Among Patients With Sarcoidosis: A Population-Based Cohort Study. Arthritis Care Res. 2017 Jan;69(1):46–50. doi:10.1002/acr.22941 PubMed PMID: 27214645; PubMed Central PMCID: PMC5121095.
28. Is the incidence of malignancy increased in patients with sarcoidosis? A systematic review and meta-analysis - Ungprasert - 2014 - Respirology - Wiley Online Library [Internet]. [cited 2026 Feb 19]. Available from: https://onlinelibrary.wiley.com/doi/10.1111/resp.12369
29. El Jammal T, Pavic M, Gerfaud-Valentin M, Jamilloux Y, Sève P. Sarcoidosis and Cancer: A Complex Relationship. Front Med. 2020 Nov 24;7:594118. doi:10.3389/fmed.2020.594118 PubMed PMID: 33330555; PubMed Central PMCID: PMC7732692.
30. Romer FK, Hommelgaard P, Schou G. Sarcoidosis and cancer revisited: a long-term follow-up study of 555 Danish sarcoidosis patients. Eur Respir J. 1998 Oct 1;12(4):906–12. doi:10.1183/09031936.98.12040906 PubMed PMID: 9817167.
31. Dehara M, Sachs MC, Grunewald J, Blomberg A, Arkema EV. Modifiable lifestyle risk factors for sarcoidosis: a nested case–control study. ERJ Open Res. 2023 Apr 3;9(2):00492–2022. doi:10.1183/23120541.00492-2022 PubMed PMID: 37020842; PubMed Central PMCID: PMC10068519.
32. Richardson DB. Occupational exposures and lung cancer: adjustment for unmeasured confounding by smoking. Epidemiology. 2010 Mar;21(2):181–6. doi:10.1097/EDE.0b013e3181c6f7d9 PubMed PMID: 20081541.
33. Dehara M, Nguyen NV, Arkema EV. Smoking and the risk of sarcoidosis: a systematic review and meta-analysis. Respir Med. 2025 May;241:108089. doi:10.1016/j.rmed.2025.108089 PubMed PMID: 40194637.
34. Ungprasert P, Crowson CS, Matteson EL. Smoking, Obesity and Risk of Sarcoidosis: A Population-Based Nested Case-Control Study. Respir Med. 2016 Nov;120:87–90. doi:10.1016/j.rmed.2016.10.003 PubMed PMID: 27817820; PubMed Central PMCID: PMC5185320.
35. Khera R, Angraal S, Couch T, Welsh JW, Nallamothu BK, Girotra S, et al. Adherence to Methodological Standards in Research Using the National Inpatient Sample. JAMA. 2017 Nov 28;318(20):2011–8. doi:10.1001/jama.2017.17653
36. El Jammal T, Pavic M, Gerfaud-Valentin M, Jamilloux Y, Sève P. Sarcoidosis and Cancer: A Complex Relationship. Front Med. 2020 Nov 24;7. doi:10.3389/fmed.2020.594118

