Asthma-Neoplasms Relationships: New Insights Using Machine Inference, Epidemiological Reasoning, And Big Data

Authors: Abbas Shojaee [1,2], Jose L. Gomez [1], Xiaochen Wang [3], Naftali Kaminski [1], Jonathan M. Siner [1], Seyedtaghi Takyar [1], Hongyu Zhao [3], Geoffrey Chupp [1*]

Affiliations:

1. Section of Pulmonary, Critical Care and Sleep Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut, USA
2. Center for Precision Pulmonary Medicine (P2MED), Yale University School of Medicine, New Haven, Connecticut, USA
3. Department of Biostatistics, School of Public Health, Yale University, New Haven, Connecticut, USA

*** Author for Correspondence:**

Dr. Geoffrey Chupp,
Section of Pulmonary, Critical Care and Sleep Medicine, Department of Internal Medicine
Yale University School of Medicine, New Haven, Connecticut 06510
Email: geoffrey.chupp@yale.edu
Office: 203 737 5405
Fax: 203-785-6094

Authors Contribution:
Abbas Shojaee designed the study, developed the methodology, performed analysis and wrote the manuscript, Geoffrey Chupp supervised the research and together with Naftali Kaminski and Jose Gomez verified the results and critically contributed to writing the manuscript, Seyedtaghi Takyar conceived the idea of the asthma-neoplasms relationship and contributed to the writing. Hongyu Zhao and Xiaochen Wang contributed to verifying methods and writing.

Disclosures:
The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

At a Glance Commentary:

Over the past three decades, consensus over the potential relationship between asthma-specific inflammation and abnormality in cellular growth and neoplasm has not been reached. The debate rises from controversial, population-level evidence that has emerged from different methods, data sources and cohorts and in the lack of sufficient molecular-level understanding of the association between allergic airway inflammation and malignant transformation.

The present study includes the most extensive cohort of patients with asthma or COPD. We used machine learning methods of causal inference combined with conventional epidemiological reasoning tools to identify potential associations between asthma and neoplasms. Our study reaffirms the well-known connections of COPD and lung cancer and reveals previously unknown relations between allergic asthma and benign neoplasms of various glands. Further, we identified the association between subtypes of COPD and hematological and lymphatic malignancies and confirmed this relation with allergic asthma. Our findings help to clarify the contradictory results from other large studies on asthma and its connection to lung cancer. Moreover, we show that the set of neoplasms in allergic asthma and COPD have significant differences, suggesting the involvement of various pathways.

Abstract

Rationale: The Relationship between asthma and abnormalities in cellular growth and neoplastic transformations has remained controversial due to contradictory population studies and the absence of a mechanistic understanding of such associations.

Objectives: To identify the relationship between asthma and neoplasms.

Methods: We used a large-scale observational administrative dataset from the state of Florida and "Causal Inference Using Composition of Transitions" (CICT) to produce multiple hypotheses on potential relationships between asthma-and neoplasms. Next, we used a similar dataset from the state of California to validate the initial findings. Nine case-control cohorts were created for patients exposed to subtypes of asthma and COPD and corresponding control groups matched on gender, age, race and history of tobacco abuse. In each cohort odds ratio analysis was conducted to measure the association of asthma and COPD with 26 different neoplasms. Bonferroni-Holm correction was applied to adjust for family-wise error.

Results
A total of 234 individual studies were conducted between Asthma (N= 998,585, male= 33%, age= 50) and COPD (N=715,971, male = 50%, age=69) patients and corresponding matched control groups (N=8,400,004, male= 42%, age= 47) of unexposed patients. Subtypes of COPD were strongly associated with bronchial and lung malignancies, with the strongest association for emphysema (OR: 9.94). Allergic asthma was associated with benign neoplasm of the meninges, salivary, pituitary, parathyroid, and thyroid glands ORs between 1.52 to 2.52, and malignant neoplasms of intrahepatic bile ducts, breast, hematopoietic, and lymphatic system with ORs between 1.45 to 2.05. Emphysema was strongly associated with malignancies of the lung, GI system, bladder and secondary malignancies. Allergic asthma diverged from COPD's in the set of neoplasms associations. Associations of obstructive asthma was a blend of COPD and allergic asthma

Conclusions: Our study identified novel associations between glandular neoplasms and allergic asthma in two large-scale US-based populations. Confirmation of previously known associations between COPD and malignancies in the same database supports the use of our approach for the identification of causal associations. Additional studies on the role of allergic inflammation neoplastic transformation of common structures of secretory organs including epithelial cells are required.

Word count:351

Keywords:
Asthma, Endocrine Gland Neoplasms, Etiology, Chronic Obstructive Pulmonary Disease

Introduction

Asthma is a common chronic disease that affects nearly 10% of the US population and is associated with annual costs of approximately 80 billion dollars [1-3]. Asthma has been linked to premature death, lower physical activity levels and comorbidities including obesity, gastroesophageal reflux, psychological problems, chronic infections, and hormonal disturbances [4-7]. It has been suggested that chronic systemic inflammation contributes to the initiation and progression of asthma-induced comorbidities[8-10]. Chronic inflammation is also known for its effect on tumorigenesis and increased risk of cancer[11-13]. Accordingly, it is plausible that pathways activated in asthma contribute to cellular growth and proliferation. Therefore, an improved understanding of the potential association between asthma and neoplasms could inform both the pathobiology of chronic airway inflammation as well as the development of malignancy.

Conclusive clinical and population-level evidence of the relationship between asthma and neoplasms does not exist. A meta-analysis suggested a 1.8 fold increase in lung cancer in patients with asthma based on 18 studies conducted between 1966 to 2002[14]. A subsequent study found a protective effect of asthma on cancer mortality in a long-term, large-scale US-based cohort(1982-2000) [15]. A nationwide study in Sweden identified an increased incidence ratio of malignancies in 15 organ systems including leukemia, gastrointestinal, lung, prostate, nervous system, and thyroid gland with asthma[16]. Additional smaller scale studies have also identified a positive association between asthma and prostate cancer[17], hematopoietic malignancies[18] and a protective effect on adenocarcinoma of the pancreas[19]. However, these studies did not discriminate between the mixed effect of COPD and chronic obstructive asthma, identified by International Classification of Disease version 9 clinical modifications (ICD9CM) code: 493.2, from allergic asthma (ICD: 493.0). Moreover, previous studies have been limited by a lack of statistical power, small sample sizes, potential biases in participant selection and exposure measurement [15,20-24], and the limited type of neoplasms that were studied. Therefore, additional studies are required to address these limitations and to evaluate associations with specific neoplasms [15,21].

To understand the potential association between asthma and the risk of various neoplasms, we studied two large-scale administrative, US-based cohort of over 20 million patients. We focused on the association pattern between neoplasms, three subtypes of asthma and five subtypes of COPD as coded in ICD9CM. We used methods of causal inference in high-dimensional data to generate new insights and employed standard epidemiological reasoning tools to evaluate new findings. Our studies suggest that asthma is associated with a range of benign glandular neoplasms and malignant neoplasms in other organs, especially secretory tissues. Our findings provide new information on the debated association between asthma and neoplasms.

Materials and Methods

Data Source and Preprocessing

Deidentified observational claim data from Healthcare Utilization Project(HCUP) Florida **S**tate **I**npatient **D**atabase (SID) and **E**mergency Department **D**atabase (ED)[25] 2004 to 2014 was used for primary analysis and hypothesis generation. Validations were conducted using HCUP California **S**tate **I**npatient **D**atabase (SID) and **E**mergency Department **D**atabase (ED)[25] from 2005 to 2011. HCUP, or Healthcare Cost and Utilization Project, is a federal-state-industry partnership sponsored by the Agency for Healthcare Research and Quality [25]. The Florida and California SID contains all inpatient discharge records in community hospitals. The ED contains all visits to the affiliated emergency department that do not result in a hospitalization. HCUP includes records for all-payers, including the uninsured, and generally, include all non-Federal acute care hospitals in a state[26]. The data is a census of all discharges[27], and together, the SID

encompass about 95 percent of all U.S. community hospital discharges [28,29 30]. Each HCUP record represents a patient encounter[2] and includes demographic data, clinical diagnoses, comorbidities, procedures, total costs of hospitalization and other information from claim records. Also, HCUP has a pseudo-patient-identifier that allows connecting all encounters of each patient across the SID and ED datasets over time. The pseudo identifier, called 'VisitLink', is assigned through a verification process to ensure the correct assignment of a unique identifier to all admissions of a patient [31].

International Classification of Diseases, Ninth Revision, Clinical Modification(ICD9CM) codes were used for the primary diagnosis and up to 24 comorbidities to identify exposures or events. Race was stratified into six categories: Black, White, Hispanic, Asian or Pacific Islander, Native American, and Other in the original HCUP data. History of smoking was identified by any primary or secondary diagnosis with the ICD9CM code of 305.1 (tobacco use disorder), 649.0 (tobacco use disorder complicating pregnancy, childbirth, or the puerperium), v15.82 (personal history of tobacco use), or 989.84 (toxic effect of tobacco). The study population included patients 18 years or older who had at least two inpatient or emergency department observations in our dataset. As our automated causal inference method uses transitions of patients between clinical conditions to build the transition network, we included patients with more than one observation. We conducted odds ratio analyses on the same population. We excluded patients with missing VisitLink information and excluded patients who died during asthma or COPD hospitalizations. Population characteristics are shown in the second column of Table 1, and the consort diagram in Figure 1 summarizes the data preparation process.

Hypothesis Generation

Evaluation of all potential relationships between subtypes of asthma and various malignancies is not feasible due to theoretical and practical reasons. Accordingly, to limit the number of evaluations and identify more likely hypotheses, we used the Causal Inference using Composition of Transitions (CICT) method[32]. CICT is a method for generating probabilistic component causal hypotheses in a complex network of potential outcomes[33]. CICT is based on the idea that observing a population, the set and frequency of events before and after a causal event are different than for a random event. For example, observing a population and recording frequency of patients' transitions between the two consecutive diagnoses, the frequency of various diagnoses after diabetes should have a different pattern than the frequency of various diagnoses after viral respiratory infection (random event). CICT defines specific features to capture these differences in large-scale population-level data and uses supervised machine learning methods to learn the patterns specific to causal, random and effect phenomena. Next, CICT uses the learned patterns to generate hypotheses about potential causal relationships among various clinical conditions. For this study, we first created a network of patients' transitions between pairs of clinical conditions in our study population. The clinical conditions were identified using the ICD9CM code of primary diagnosis for admission. Next, CICT was applied to identify the relationships between asthma and various neoplasms. Figure 2 and Table 2, show the CICT hypotheses related to potential asthma induced neoplasms.

Odds Ratio Analysis

To evaluate CICT generated hypotheses related to asthma and neoplasms, we used odds ratio analysis which is a measure of association. Also, to discriminate the effect of asthma and Chronic Obstructive Pulmonary Disease (COPD) case mix, we further added OR analysis for subtypes of COPD and identified neoplasms. For each exposure to one of the subtypes of asthma or COPD, we identified exposed groups as cases and matched them with a control group based on similar race, gender, age, and smoking status. Four cohorts were designed for asthma and its three subtypes

with ICD-9-CM codes: 493.* for combined asthma (N= 998937), 493.0 for extrinsic asthma (N= 28149), 493.2 chronic obstructive asthma for (N= 235446), and 493.9 for asthma unspecified (N= 853556). We excluded intrinsic asthma groups due to a low number of exposed patients (N= 1167, 0.12% of asthma patients) and that few or no neoplasms of interest were identified in both case and control groups. Six cohorts were designed for subtypes of COPD with ICD9CM codes: 491 for chronic bronchitis (N= 390862), 492 for emphysema (N= 97956), 494 for Bronchiectasis (N= 29539), and 496 for chronic airway obstruction, not elsewhere classified (N= 715957). Next, for each of the 17 neoplasm events suggested by CICT (Table 2 and Figure 2) and nine other relevant neoplasms (appendix table 2), we conducted a separate study. A total of 234 individual studies were designed between ten exposures and 26 neoplasm events.

Constructing matched control groups

In each study, the control group was drawn from subjects with no recorded code for any asthma or COPD subtypes as the primary diagnosis for hospital admission or comorbidities during the study period (2005-2011). To minimize the effects of confounding and to obtain an unbiased estimate of odds ratios, we matched the control group with the exposed patients in each study using coarsened exact matching (CEM)[34,35] on baseline variables including age, gender, race, and smoking. The remaining imbalance and heterogeneity between case and control groups were assessed using the multivariate L1 metric [36]. Previous research used different user-specified criteria for L1 statistics [37-42]. In this study, a threshold of L1 statistics less than 0.2 after CEM matching was considered as the acceptable.

Statistical analysis

Descriptive statistics of individual characteristics at baseline were calculated for the whole population, asthma subtypes cohorts, and COPD subtypes cohorts (Table 1). We computed frequencies for categorical variables and means with standard deviations for continuous variables. The association between asthma and cancers was estimated using both the Cochran-Mantel-Haenszel (CMH) [43,44] common odds ratio and pooled odds ratio over the whole cohort. We report CMH common odds ratios (ORs) and 95% confidence intervals (CIs) for 26 neoplasms (column 1 of table 3) in the ten matched case-control populations for subtypes of asthma and COPD (row 2 of table 3). A Bonferroni-holm [45] correction was applied to each asthma and COPD subtype cohort to control family-wise error rate. The significance threshold was established at $P \leq 0.05$. All statistical analyses were performed using F# and R.

Results

Our large-scale empirical study reveals that benign neoplasms of glands are more frequent in asthma. Also, it shows that COPD is associated with multiple malignancies including the well-known connections between COPD and lung cancer. Moreover, we identified the association between subtypes of COPD and hematological and lymphatic malignancies and confirmed such relation with allergic asthma.

Among 524 neoplasms encoded in ICD9CM codes, our hypothesis generation method(CICT) suggests that twenty could be induced by subtypes of asthma (Figure 2, Table 2). We used epidemilogical reasoning methods to evaluate the 17 of CICT hypotheses along with nine other

competing hypotheses suggested by pulmonology board. Accordingly, to measure the association of these twenty six neoplasms with an exposure to subtypes of asthma and COPD, four cohorts were designed for asthma and its three subtypes with ICD-9-CM codes: 493.* for combined asthma (N= 998937), 493.0 for extrinsic asthma (N= 28149, male= 31%, age= 47), 493.2 chronic obstructive asthma for (N= 235446, male= 38%, age= 65), and 493.9 for asthma unspecified (N= 853556, male= 32%, age= 48). Six cohorts were designed for subtypes of COPD with ICD9CM codes: 491 for chronic bronchitis (N= 390862, male= 48%, age=68), 492 for emphysema (N= 97956 , male= 53% , age= 68), 494 for Bronchiectasis (N= 29539, male= 39%, age= 71), and 496 for chronic airway obstruction, not elsewhere classified (N= 715957, male= 50%, age= 69) Approximately 8.4 million matched controls were included in each cohort: ICD9CM 493.0 (N= 8,398,723), 493.2 (N= 8,400,004), 493.9 (N= 8,400,004), 491 (N= 8,400,004), 492 (N= 8,397,079), 494 (N= 8,399,392), 495 (N=8,348,680), 496 (N= 8,400,004). The matching reduced L1 statistics from the range of 0.21-0.34 to less than 0.17 after matching in various cohorts of asthma and COPD, indicating the case and matched cohorts in 234 different studies are similar.[34]

A total of 139,038 incident cases of neoplasms were identified among patients with exposure to asthma, and 338,418 in those exposed to COPD, and 518,138 in the control cohort that had no record of exposure to either respiratory disorders. The distributions of age, sex, race, and smoking in overall population and subtypes of asthma and COPD are shown in Table 1. A history of asthma in 10.63% of the total included population, and a history of COPD of 6.96% was observed, consistent with national averages [3 1 46]. On average, patients with chronic obstructive asthma and subtypes of COPD tended to be older than those with neither condition, including those with extrinsic asthma. Patients with extrinsic asthma, asthma unspecified and bronchiectasis are less likely to be former smokers than patients with other subtypes of asthma and COPD. Bronchiectasis and subtypes of asthma are more likely to be women than in patients with chronic bronchitis, emphysema, and chronic airway obstruction - not elsewhere classified. Compared with healthy people with no history of asthma or COPD, patients with asthma were more likely to be older adults, females, white or black races, and smokers. Unspecified asthma was the most frequent diagnostic code and observed in 85.41%% of the asthma patients. Obstructive asthma had the highest risk of various pulmonary cancers (OR: 3.40 to 4.78), followed by unspecified asthma (OR: 1.66 to 2.18).

The CMH common OR estimates and 95% CIs are shown in Table 3 and appendix figures 2,3 for asthma, COPD, and their subtypes compared to the control cohort for individual cancers (column 1 of Table 3). The pooled odds ratio is shown in appendix POOLED ODDS RATIO. The pooled odds ratios are usually higher than corresponding common odds ratios. CMH common ORs is a weighted combination of ORs in balanced strata of the population. The balancing, conducted by CEM, ensures for the similarity of age, gender, race, and smoking between case and controls in each stratum. Extrinsic asthma was associated with benign neoplasms of cerebral meninges, major salivary glands, parathyroid gland, and pituitary gland and craniopharyngeal duct (OR:1.57-2.52); and with malignant neoplasms of intrahepatic bile ducts and upper-outer quadrant of female breast (OR 1.45-1.81). Extrinsic asthma shows a weak association with lung cancers compared to other exposures. Emphysema was associated with almost all neoplasms (OR 1.24 – 9.94) except benign neoplasm of the pituitary gland and craniopharyngeal duct, benign neoplasm of thyroid glands, malignant neoplasm of extrahepatic bile ducts and malignant neoplasm of intrahepatic bile ducts. Also, emphysema showed the strongest association with lung cancers (OR 7.93-9.94). Chronic bronchitis and chronic airway obstruction show similar patterns of emphysema COPD regarding associated neoplasms (table 3 and appendix figure 3). The actual number of cases in the COPD and asthma cohort is reported in Table 4, and the pooled odds ratio are presented in appendix table 1.

Subtypes of asthma and COPD showed a weak association with secondary, or metastatic, neoplasms (table 3 and appendix figure 2) with often slightly increased odds ratios. However, emphysema is associated with increased odds for secondary malignant neoplasm of other specified sites (ICD9CM: 198, OR: 2.27, CI: 2.20-2.33), secondary malignant neoplasm of respiratory and digestive systems (ICD9CM: 197, OR: 2.07, CI: 2.01-2.13) and secondary malignancy of intra-abdominal lymph nodes (ICD9CM: 196.2, OR: 1.24, CI: 1.15-1.34). Extrinsic asthma is associated with considerably increased odds of lymphatic and hematopoietic cancers (ICD9CM: 238.7, OR: 2.05, CI: 1.80-2.33). Subtypes of COPD are also associated with hematopoietic and lymphatic cancers (OR: 2.03-3.23).

Discussion

Evaluation of the relationship between asthma and neoplasms is a broad question that would entail the evaluation of over five thousands possible hypotheses between 514 neoplasms coded in the HCUPS administrative data and subtypes of asthma and relevant conditions.

Allergic asthma and COPD have contrasting associations with neoplasms

We found that allergic asthma is associated with a specific profile of neoplasms that is different from those found in obstructive asthma and COPD. Allergic asthma has a small or insignificant effect (CI contains one) on bronchial or lung cancer. This pattern contrasts that of COPD which shows a strong association with lung cancer, consistent with existing knowledge[47-49]. Chronic obstructive asthma, a diagnosis label for the spectrum of coexisting allergic and obstructive manifestations, followed a pattern similar to COPD. This finding may help explain some of the contradictory results described in previous studies where some showed a protective effect of asthma on multiple cancers and some showed a positive and inducing effect. Such contradictions could have raised because of the different case mix in those studies.

Allergic asthma is associated with benign glandular neoplasms

Our analysis indicates that allergic asthma is primarily associated with benign neoplasms in glandular tissues including the pituitary, thyroid, parathyroid, and salivary glands. Increased odds of neoplasia in various glands suggests that neoplasia happens in common glandular structures such as epithelium. This novel finding proposes that hormonal and metabolic disturbances such as obesity, diabetes, osteoporosis and thyroid dysfunction in asthmatic patients might be associated with specific pathway dysregulation in allergic airway inflammation and neoplastic transformation. Traditionally, diabetes, obesity, and osteoporosis have been attributed to high dose inhaled corticosteroid (ICS) and oral corticosteroid (OCS) therapy [50,51]. Also, corticosteroids have been associated with a down-regulation of thyroid function[52-54]. However, a connection between thyroid or parathyroid neoplasms with corticosteroids has not been reported. Our findings are in line with recent studies which failed to establish an association between regular or high dose inhaled corticosteroid (ICS) and osteoporosis[55] or diabetes[56] and suggested that the chronic respiratory disease process itself could be responsible. Interestingly, in our study allergic asthma shows association with malignancies in ductal structures such as intrahepatic biliary system and breast. Chronic obstructive asthma (ICD: 493.2) is a mixed coding that covers the frequent case of patients with mixed asthma and obstructive symptoms.

COPD is associated with various malignant neoplasms

Our results on COPD subtypes and combined cohorts are consistent with existing knowledge of the increased risk of lung cancers, specifically the seven to nine-fold increase in odds of lung and bronchial cancers in the presence of emphysema. Both chronic asthma and COPD induces alveolar hypoxia, hypoxemia and consequent tissue hypoxia which also results in systemic inflammation. Recent studies have established a direct link between hypoxia and the composition and organization of the extracellular matrix (ECM). Furthermore, emerging data indicate that ECM has a crucial role in metastasis [57]. Specifically, increased expression of genes encoding proteins that mediate ECM remodeling has been associated with increased mortality in patients with breast, lung and gastric cancers [57-59]. These molecular findings are in line with our findings that show an increased risk of breast, gastrointestinal and lung cancer. However, the higher incidence of lung cancer and lower, or statistically insignificant increased, the rate of breast and gastrointestinal cancer in COPD in comparison with asthma suggests that the proinflammatory effect of COPD on ECM is local, whereas asthma induces a more systemic effect.

The increased odds of lymphatic and hematopoietic cancers in the presence of allergic asthma is confirmed in our study. These findings provide new insight into the inconclusive evidence regarding the role of asthma in lymphatic and hematopoietic cancer [60,61] with different studies reporting highly elevated[62] to reduced risks[23]. Moreover, for the first time, we showed that the subtypes of COPD are strongly associated with hematopoietic and lymphatic cancers.

Automatic hypothesis generation accelerates research

Another significant result of this research is that we employed new approaches to machine inference to generate hypotheses from data and the standard tools of epidemiological reasoning, to validate those hypotheses. The hypotheses were generated using CICT[32] which employes supervised machine learning approach for top-down causal inference. CICT learns the features of causal relationships from labeled known causal relationships. Then it uses the learned features to predict new causal associations with high accuracy. Using CICT on Florida state data we reduced thousands of possible associations of 524 candidate neoplasms to subtypes of asthma, to only 17 promising relationships. Next, we examined those findings using California data, through creating balanced cohorts and conducting odds ratio analysis. The two-step process resulted in a targeted evaluation plus validation of just the promising hypotheses. It is important to clarify that although CICT is a method for identifying causal associations, here our epidemiological reasoning methods are meant only to validate the association claim and not the causality one. Nevertheless, the use of non-reductionist methods such as CICT helps to produce valuable insights by embracing the heterogeneity of the population, finding patterns that are otherwise unidentifiable, and increasing the external validity of those findings. We suggest that adding machine learning to our toolbox of causal inference methods can help in producing evidence in an expedient and non-expensive manner.

Large-scale observational data is an indispensable source of discovery

The vast amount of administrative data can produce valuable insights. The large volume of data when used with proper methods, can compensate for the inherent noise and help find signals that are invisible in small-scale data, limited data acquisition practices, or obsessively filtered data. In this study, we used ten years of unfiltered data from Florida State inpatient and emergency departments, to conduct causal inference and then employed six years of similar data from California State to validate our findings through 234 individual studies in ten balanced cohorts. The large-scale discovery and validation datasets helped us to bring new insights into controversies about asthma, COPD, and neoplasms.

Conclusion

The present study is the most extensive cohort reported of patients exposed to asthma or COPD. We revealed the previously unknown relation between allergic asthma and benign neoplasms of multiple glands. Also, we identified the association between subtypes of COPD and hematological and lymphatic malignancies and confirmed such relation with allergic asthma. Moreover, we discriminated the mixed signal of obstructive asthma which caused contradictory results in previous studies, probably due to selection bias, and showed that the set of neoplasms in allergic asthma and COPD have significant differences that suggest various pathway involvements.

Limitations:

The use of administrative data is constrained by the lack of clinical depth and potential coding errors. For example, an accurate specification of the onset or subtype of asthma might not be possible. Also, attributing the higher odds of benign neoplasms of glands to corticosteroids or

asthma needs further investigation through clinical data. As HCUP data does not contain outpatient medical office visits, our study was limited to inpatient and emergency department records, which usually contain records for patients at a later stage or more severe condition of asthma or COPD. Nevertheless, employing proper methods on large volumes of data provides the opportunity to improve the signal to noise ratio. In this study, the large cohort population allows us to capture the relationship between allergic asthma and benign neoplasms and uncover the differences between obstructive and allergic asthma in the promotion of lung cancers.

Future studies

The molecular mechanism(s) that link asthma with its comorbidities, especially the potential connection between asthma-related dysregulations and cellular growth and neoplasms, has yet to be studied. Further epidemiological etiology research for identifying and profiling upstream and downstream causal associations of asthma can provide clues and guide systems biology level studies[5] and may help in understanding the underlying mechanisms of asthma itself.

Figure1: Consort diagram

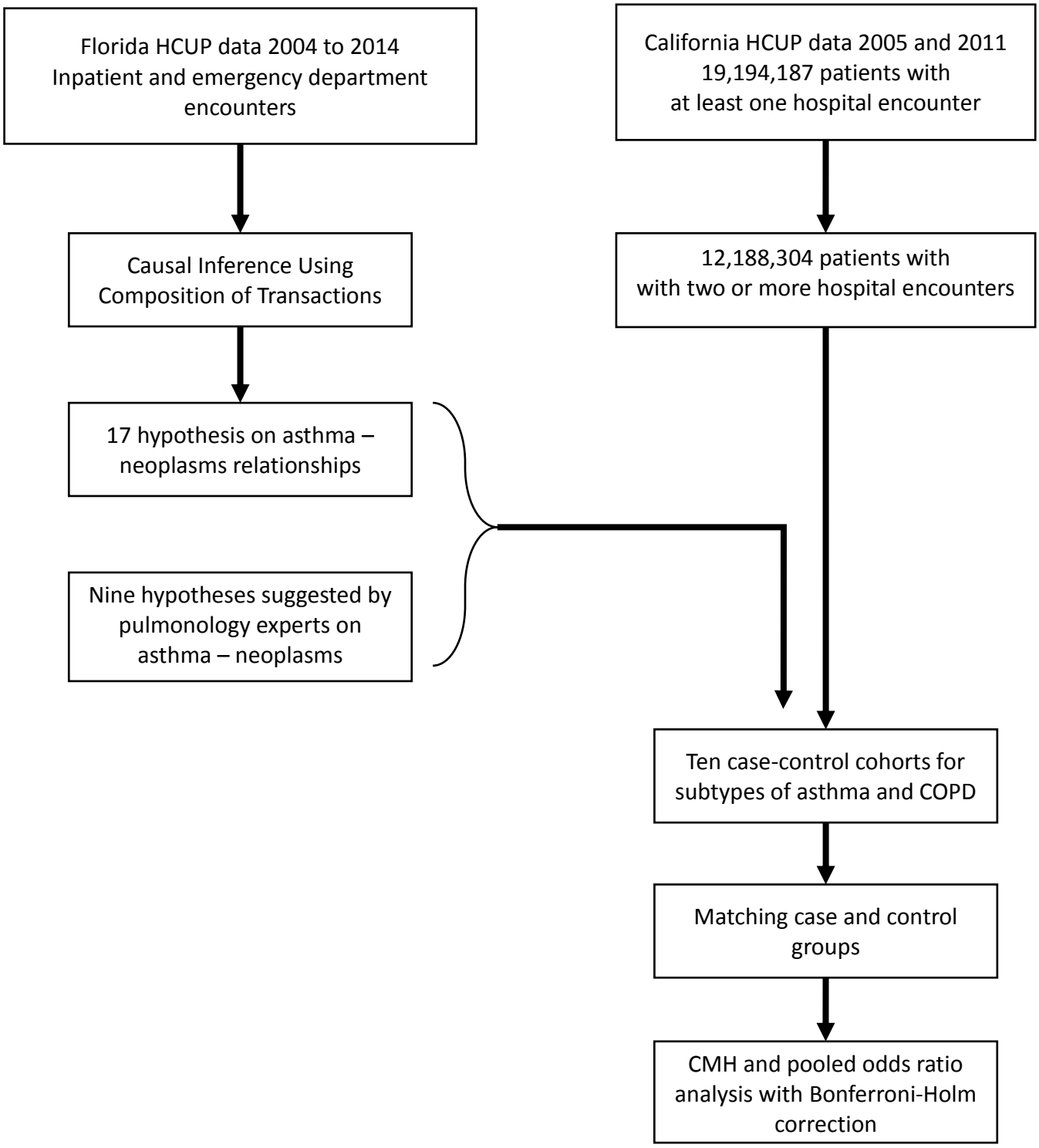


Table 1: Basic characteristics of patients with asthma and COPD by subtype - California HCUP SID, EDD 2005-2011.

Characteristics, N (%)	Population Data	Asthma Subtypes			Asthma (493.*)	COPD Subtypes				COPD (490, 491, 492, 494, 495, 496)
		Extrinsic asthma (493.0)	Chronic obstructive asthma (493.2)	Asthma, unspecified (493.9)		Chronic bronchitis (491)	Emphysema (492)	Bronchiectasis (494)	Chronic airway obstruction, not elsewhere classified. (496)	
Multivate Imbalance L1		0.098	0.141	0.044	0.050	0.157	0.169	0.158	0.157	0.152
Total count	12,188,304	28149 (0.33%)	235446 (2.73%)	853556 (9.22%)	999370 (10.63%)	390862 (4.45%)	97956 (1.15%)	29539 (0.35%)	715957 (7.85%)	715971 (6.95%)
Age (SD)	42.65 (23.86)	47 (18)	65 (15)	48 (19)	50 (20)	68 (15)	68 (14)	71 (15)	69 (14)	69 (14)
Sex										
Men	5,314,100 (43.6%)	8864 (31.49%)	88982 (37.79%)	269308 (31.55%)	329413 (32.96%)	186542 (47.73%)	52571 (53.67%)	11610 (39.30%)	361018 (50.42%)	361027 (50.42%)
Women	6,795,494 (56.4%)	19285 (68.51%)	146464 (62.21%)	584248 (68.45%)	669957 (67.04%)	204320 (52.27%)	45385 (46.33%)	17929 (60.70%)	354939 (49.58%)	354944 (49.58%)
Race										
White	6,194,004 (51.4%)	16392 (58.23%)	152944 (64.96%)	474326 (55.57%)	574255 (57.46%)	274497 (70.23%)	74496 (76.05%)	18864 (63.86%)	520076 (72.64%)	519785 (72.60%)
Black	1,160,655 (9.6%)	3766 (13.38%)	28163 (11.96%)	120800 (14.15%)	134259 (13.43%)	37209 (9.52%)	8345 (8.52%)	1495 (5.06%)	60958 (8.51%)	60616 (8.47%)
Hispanic	3,468,709 (28.8%)	5194 (18.45%)	32482 (13.80%)	183721 (21.52%)	202252 (20.24%)	48715 (12.46%)	9003 (9.19%)	4217 (14.28%)	84978 (11.87%)	85020 (11.87%)
Asian or Pacific Islander	807,814 (6.7%)	2003 (7.12%)	16926 (7.19%)	51158 (5.99%)	61992 (6.20%)	22847 (5.85%)	4468 (4.56%)	4320 (14.62%)	36803 (5.14%)	37095 (5.18%)
Native American	27,493 (0.2%)	75 (0.27%)	585 (0.25%)	2467 (0.29%)	2771 (0.28%)	764 (0.20%)	167 (0.17%)	36 (0.12%)	1322 (0.18%)	1535 (0.21%)
Others	389,883 (3.2%)	719 (2.55%)	4346 (1.85%)	21084 (2.47%)	23841 (2.39%)	6830 (1.75%)	1477 (1.51%)	607 (2.05%)	11820 (1.65%)	11920 (1.66%)
Smoking	1,664,159 (13.8%)	5762 (20.47%)	88443 (37.56%)	212813 (24.93%)	262847 (26.30%)	173756 (44.45%)	44051 (44.97%)	4361 (14.76%)	283411 (39.58%)	283423 (39.59%)

Figure 2: CICT generated hypotheses on asthma neoplasms relationships using Florida 2003-2015 data. Malignant neoplasms are abbreviated to MN and benign neoplasms to BN. Edge values are CICT coefficients on the strength of hypothesis(0 to 1). Nodes are color-coded and organized based on their clinical relevance: asthma subtypes (center, black color), MN of gastrointestinal (left-top , blue), MN of bile ducts(left, yellow), MN or lung (left-bottom, red), BN of cerebral meninges and glands (bottom, green), MN urinary track (right, brown), MN of lymphatic and hematopoietic-top (top, orange), MN of upper outer breast(top, pink). Node size is non-linearly proportional to the patients' frequency.

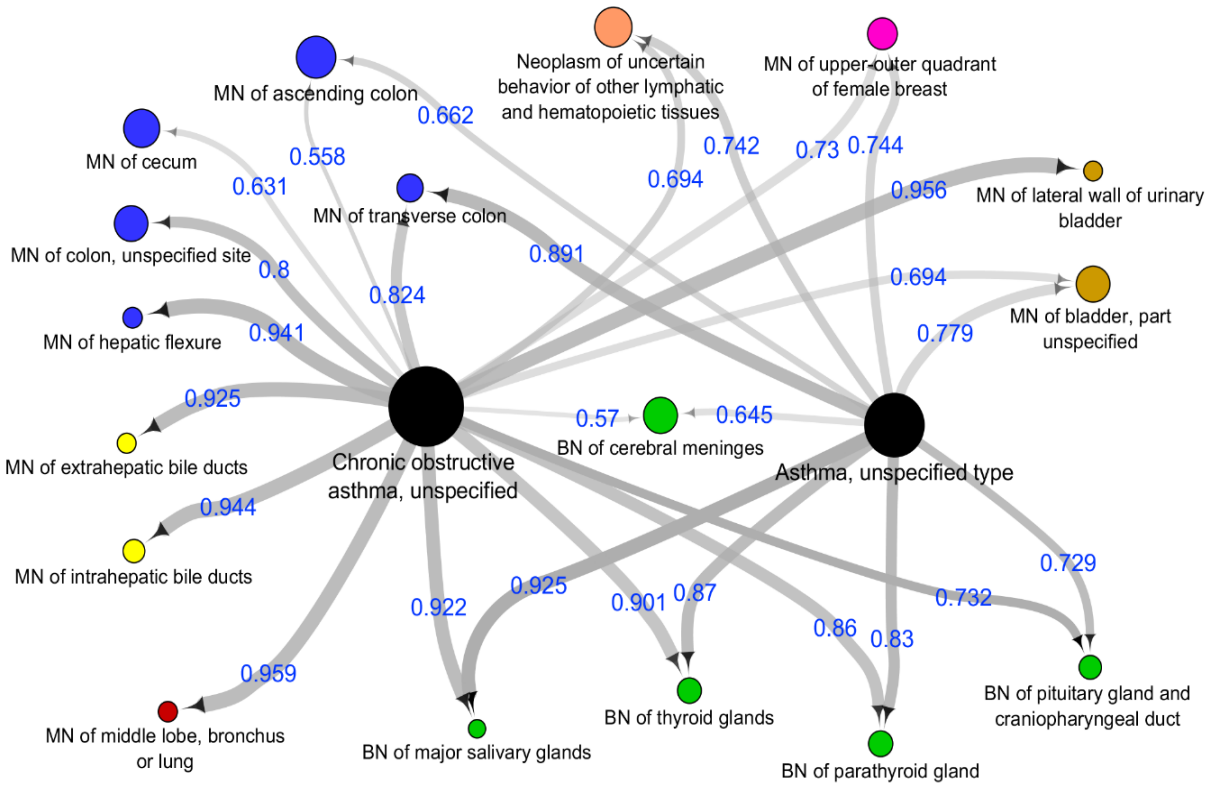


Table 2: CICT generated hypotheses on the effect of subtypes of asthma (columns) on inducing neoplasms (rows) using Florida 2003-2015 SID and ED data. The number represents the predicted strength of causal hypotheses (0-1). The confidence interval of each prediction is given in the parenthesis.

ICD9CM code	**Exposure** / **Event**	Asthma, unspecified type, with (acute) exacerbation (ICD: 493.9)	Chronic obstructive asthma, unspecified (ICD:493.2)
210.2	BN of major salivary glands	0.925 (0.81 - 1.00)	0.922 (0.81 - 1.00)
227.3	BN of pituitary gland and craniopharyngeal duct	0.729 (0.57 - 0.89)	0.732 (0.63 - 0.83)
227.1	BN of parathyroid gland	0.830 (0.72 - 0.94)	0.860 (0.75 - 0.97)
225.2	BN of cerebral meninges	0.645 (0.28 - 1.00)	0.570 (0.33 - 0.81)
226	BN of thyroid glands	0.870 (0.76 - 0.98)	0.901 (0.79 - 1.00)
153.4	MN of cecum	N/A	0.631 (0.43 - 0.83)
153.6	MN of ascending colon	0.662 (0.55 - 0.77)	0.558 (0.44 - 0.67)
153.9	MN of colon, unspecified site	N/A	0.800 (0.70 - 0.90)
153.1	MN of transverse colon	0.891 (0.78 - 1.00)	0.824 (0.72 - 0.93)
153.0	MN of hepatic flexure	N/A	0.941 (0.83 - 1.00)
153.1	MN of transverse colon	0.891 (0.78 - 1.00)	0.824 (0.72 - 0.93)
155.1	MN of intrahepatic bile ducts	N/A	0.944 (0.83 - 1.00)
156.1	MN of extrahepatic bile ducts	N/A	0.925 (0.82 - 1.00)
174.4	MN of upper-outer quadrant of female breast	0.744 (0.64 - 0.84)	0.730 (0.63 - 0.83)
162.4	MN of middle lobe, bronchus or lung	N/A	0.959 (0.85 - 1.00)
188.9	MN of bladder, part unspecified	0.779 (0.56 - 1.00)	0.694 (0.59 - 0.80)
188.2	MN of lateral wall of the urinary bladder		0.956 (0.85 - 1.00)
238.7	Other lymphatic and hematopoietic tissues	0.742 (0.63 - 0.86)	0.694 (0.59 - 0.80)

Table 3: Cochran- Mantel-Haenszel (CMH) common odds ratios of neoplasms in subtypes of asthma and COPD. Values in parenthesis show the lower and upper bounds of the confidence interval. **Bold type indicates significant odds ratios** when the adjusted p-values are less than 0.05. In exposure description, BN and MN correspondingly stand for 'benign neoplasm' and 'malignant neoplasm.'

ICD9CM Code	Event \ Exposure	Asthma (493.*)	Asthma, unspecified (493.9)	Extrinsic asthma (493.0)	Chronic obstructive asthma (493.2)	Bronchitis, not specified as acute or chronic (490)	Chronic bronchitis (491)	Bronchiectasis (494)	Chronic airway obstruction, not elsewhere classified (496)	Emphysema (492)
210.2	BN of major salivary glands	**1.17 (1.05-1.31)**	1.17 (1.04-1.33)	**2.52 (1.64-3.88)**	1.18 (0.98-1.42)	0.72 (0.57-0.91)	1.00 (0.86-1.17)	1.33 (0.82-2.14)	1.15 (1.02-1.29)	1.33 (1.04-1.7)
227.3	BN of pituitary gland and craniopharyngeal duct	**1.43 (1.35-1.52)**	**1.5 (1.41-1.6)**	**2.43 (1.9-3.11)**	**1.34 (1.19-1.5)**	**1.25 (1.13-1.39)**	0.96 (0.86-1.07)	1.04 (0.74-1.47)	**1.18 (1.09-1.27)**	1.02 (0.84-1.24)
227.1	BN of parathyroid gland	**1.17 (1.08-1.26)**	**1.22 (1.12-1.32)**	**1.99 (1.45-2.74)**	1.06 (0.93-1.22)	0.97 (0.84-1.12)	**0.82 (0.72-0.93)**	1.05 (0.73-1.52)	1.1 (1.01-1.2)	1.26 (1.03-1.54)
225.2	BN of cerebral meninges	**1.26 (1.21-1.32)**	**1.29 (1.23-1.35)**	**1.57 (1.27-1.94)**	**1.34 (1.26-1.43)**	**1.32 (1.23-1.42)**	**1.17 (1.1-1.24)**	**1.33 (1.13-1.56)**	**1.29 (1.24-1.34)**	**1.24 (1.12-1.38)**
226	BN of thyroid glands	**1.33 (1.24-1.43)**	**1.36 (1.26-1.47)**	1.52 (1.07-2.15)	**1.26 (1.09-1.44)**	0.98 (0.85-1.13)	1.09 (0.95-1.23)	1.52 (1.07-2.15)	**1.17 (1.06-1.29)**	1.25 (1.00-1.57)
153.4	MN of cecum	0.98 (0.92-1.05)	0.97 (0.9-1.05)	1.32 (0.95-1.85)	1.05 (0.95-1.15)	**0.81 (0.72-0.91)**	**0.83 (0.77-0.9)**	1.15 (0.92-1.44)	**1.21 (1.15-1.28)**	**1.37 (1.22-1.55)**
153.6	MN of ascending colon	1.02 (0.96-1.08)	1.02 (0.95-1.09)	1.17 (0.84-1.63)	1.1 (1.01-1.2)	**0.78 (0.69-0.87)**	**0.90 (0.83-0.97)**	1.14 (0.92-1.41)	**1.22 (1.16-1.28)**	**1.28 (1.14-1.44)**
153.9	MN of colon, unspecified site	**1.09 (1.05-1.13)**	**1.09 (1.04-1.14)**	1.3 (1.06-1.59)	**1.27 (1.2-1.34)**	1.03 (0.97-1.1)	**1.12 (1.07-1.17)**	1.12 (0.97-1.3)	**1.42 (1.38-1.47)**	**1.5 (1.40-1.62)**
235.2	Neoplasm of uncertain behavior of stomach, intestines, and rectum	**1.49 (1.38-1.61)**	**1.45 (1.32-1.59)**	1.75 (1.14-2.69)	**1.77 (1.58-1.99)**	**1.29 (1.12-1.48)**	**1.63 (1.49-1.79)**	1.40 (1.01-1.93)	**1.78 (1.66-1.92)**	**1.64 (1.39-1.93)**
153.0	MN of hepatic flexure	1.12 (1-1.26)	1.12 (0.98-1.28)	1.72 (0.98-3.04)	1.22 (1.03-1.45)	0.89 (0.72-1.11)	0.96 (0.83-1.11)	1.29 (0.86-1.93)	**1.41 (1.28-1.55)**	**1.57 (1.27-1.95)**
153.1	MN of transverse colon	1.01 (0.93-1.11)	0.95 (0.85-1.05)	1.41 (0.9-2.22)	**1.25 (1.11-1.42)**	0.92 (0.79-1.08)	1.04 (0.93-1.15)	1.13 (0.83-1.56)	**1.33 (1.23-1.43)**	**1.52 (1.29-1.79)**
155.1	MN of intrahepatic bile ducts	0.87 (0.78-0.98)	0.92 (0.81-1.05)	1.81 (1.12-2.92)	0.85 (0.71-1.02)	**0.69 (0.56-0.85)**	**0.60 (0.51-0.71)**	0.89 (0.58-1.37)	0.93 (0.84-1.03)	1.08 (0.85-1.36)
156.1	MN of extrahepatic bile ducts	**0.8 (0.69-0.92)**	0.79 (0.68-0.94)	0.96 (0.43-2.14)	0.89 (0.73-1.1)	**0.63 (0.49-0.83)**	**0.56 (0.46-0.68)**	1.19 (0.77-1.83)	0.95 (0.84-1.07)	1.11 (0.85-1.46)
174.5	MN of lower-outer quadrant of female breast	**1.2 (1.05-1.37)**	**1.24 (1.08-1.42)**	1.65 (0.91-2.99)	1.16 (0.92-1.48)	0.8 (0.6-1.07)	1.08 (0.86-1.34)	0.35 (0.11-1.10)	1.17 (0.99-1.38)	1.63 (1.15-2.31)
174.4	MN of upper-outer quadrant of female breast	**1.16 (1.09-1.23)**	**1.19 (1.11-1.26)**	1.45 (1.09-1.93)	1.09 (0.97-1.21)	0.92 (0.81-1.03)	0.87 (0.78-0.97)	0.84 (0.6-1.18)	1.07 (0.99-1.16)	**1.38 (1.16-1.63)**
162.2	MN of main bronchus	**3.12 (2.88-3.39)**	**2.18 (1.97-2.41)**	2.13 (1.26-3.61)	**4.78 (4.34-5.25)**	**2.43 (2.12-2.78)**	**6.66 (6.16-7.19)**	**4.34 (3.39-5.54)**	**6.88 (6.43-7.35)**	**9.11 (8.24-10.08)**
162	MN of trachea, bronchus, and lung	**2.27 (2.22-2.31)**	**1.66 (1.62-1.7)**	**1.69 (1.48-1.92)**	**3.4 (3.32-3.49)**	**1.96 (1.9-2.03)**	**4.73 (4.64-4.82)**	**3.65 (3.44-3.87)**	**5.24 (5.16-5.32)**	**7.93 (7.73-8.13)**
162.4	MN of middle lobe, bronchus or lung	**2.82 (2.54-3.13)**	**2.11 (1.86-2.4)**	1.55 (0.74-3.27)	**4.39 (3.88-4.98)**	**2.62 (2.21-3.1)**	**5.39 (4.86-5.98)**	**5.10 (3.87-6.72)**	**6.38 (5.85-6.94)**	**9.94 (8.73-11.33)**

162.8	MN of other parts of bronchus or lung	**2.43 (2.29-2.57)**	**1.79 (1.66-1.92)**	1.27 (0.82-1.97)	**3.74 (3.49-4.01)**	**2.19 (1.99-2.4)**	**5.52 (5.23-5.82)**	**4.34 (3.69-5.11)**	**5.89 (5.63-6.16)**	**8.19 (7.62-8.8)**
188.9	MN of bladder, part unspecified	**1.34 (1.28-1.4)**	**1.19 (1.12-1.26)**	1.26 (0.93-1.71)	**1.69 (1.59-1.8)**	**1.17 (1.08-1.27)**	**1.46 (1.39-1.53)**	**1.53 (1.31-1.78)**	**1.84 (1.78-1.9)**	**2.09 (1.95-2.24)**
188.2	MN of lateral wall of urinary bladder	**1.2 (1.06-1.36)**	1.05 (0.9-1.23)	0.81 (0.3-2.17)	**1.49 (1.27-1.75)**	1.00 (0.8-1.25)	**1.22 (1.07-1.39)**	1.30 (0.85-2.01)	**1.75 (1.6-1.91)**	**1.73 (1.43-2.09)**
196.2	Secondary and unspecified MN of intra-abdominal lymph nodes	**0.87 (0.84-0.91)**	**0.88 (0.84-0.92)**	1.22 (1.02-1.47)	**0.86 (0.81-0.91)**	**0.63 (0.58-0.68)**	**0.65 (0.61-0.68)**	0.93 (0.79-1.08)	**1.07 (1.03-1.1)**	**1.24 (1.15-1.34)**
196	Secondary and unspecified MN of lymph nodes	**1.08 (1.06-1.1)**	**1.04 (1.02-1.07)**	**1.38 (1.25-1.53)**	**1.17 (1.14-1.21)**	**0.86 (0.83-0.9)**	**1.15 (1.12-1.18)**	**1.25 (1.14-1.36)**	**1.53 (1.5-1.56)**	**2.09 (2.02-2.17)**
198	Secondary MN of other specified sites	**1.04 (1.02-1.05)**	0.97 (0.96-0.99)	1.06 (0.96-1.17)	**1.2 (1.17-1.23)**	**0.95 (0.92-0.98)**	**1.34 (1.32-1.37)**	**1.18 (1.10-1.26)**	**1.80 (1.77-1.82)**	**2.27 (2.2-2.33)**
197	Secondary MN of respiratory and digestive systems	**1.03 (1.01-1.04)**	**0.97 (0.95-0.99)**	1.12 (1.02-1.22)	**1.18 (1.15-1.21)**	**0.88 (0.85-0.9)**	**1.3 (1.28-1.33)**	**1.14 (1.07-1.22)**	**1.68 (1.65-1.70)**	**2.07 (2.01-2.13)**
238.7	Other lymphatic and hematopoietic tissues	**1.81 (1.76-1.85)**	**1.75 (1.7-1.8)**	**2.05 (1.8-2.33)**	**2.36 (2.28-2.44)**	**2.03 (1.95-2.1)**	**2.36 (2.3-2.43)**	**3.23 (3.01-3.46)**	**2.33 (2.28-2.38)**	**2.59 (2.47-2.71)**

Table 4: Prevalence of who developed the event (row heading) who were exposed (column heading).

ICD9CM Code	Event / Exposure	Asthma (493.*)	Asthma, unspecified (493.9)	Extrinsic asthma (493.0)	Chronic obstructive asthma (493.2)	Bronchitis, not specified as acute or chronic (490)	Chronic bronchitis (491)	Bronchiectasis (494)	Chronic airway obstruction, not elsewhere classified (496)	Emphysema (492)
210.2	BN of major salivary glands	382	300	21	131	76	206	17	418	71
227.3	BN of pituitary gland and craniopharyngeal duct	1350	1190	64	334	393	395	33	881	105
227.1	BN of parathyroid gland	796	670	38	229	192	271	29	669	104
225.2	BN of cerebral meninges	2741	2166	86	1045	872	1480	151	2993	391
226	BN of thyroid glands	937	817	32	223	204	281	32	546	80
153.4	MN of cecum	1093	797	35	475	283	679	78	1870	291
153.6	MN of ascending colon	1300	952	35	570	310	851	89	2168	316
153.9	MN of colon, unspecified site	3276	2474	95	1442	993	2281	179	5357	791
235.2	Neoplasm of uncertain behavior of stomach, intestines, and rectum	749	543	21	349	207	589	38	1188	157
153.0	MN of hepatic flexure	335	244	12	151	85	222	24	600	94
153.1	MN of transverse colon	569	392	19	281	161	424	39	1028	158
155.1	MN of intrahepatic bile ducts	343	272	17	131	87	162	21	461	76
156.1	MN of extrahepatic bile ducts	217	159	6	99	56	116	21	351	57
174.5	MN of lower-outer quadrant of female breast	276	235	11	78	50	100	3	200	35
174.4	MN of upper-outer quadrant of female breast	1301	1101	47	357	284	414	34	895	151
162.2	MN of main bronchus	995	486	14	692	243	1862	68	3285	705
162	MN of trachea, bronchus, and lung	14672	7644	238	9502	3957	24688	1258	47971	10694
162.4	MN of middle lobe, bronchus or lung	567	306	7	379	160	853	54	1771	423
162.8	MN of other parts of bronchus or lung	1749	928	20	1150	494	3139	153	5809	1260
188.9	MN of bladder, part unspecified	2159	1298	42	1256	643	2369	168	5700	955
188.2	MN of lateral wall of urinary bladder	291	170	4	170	82	310	21	824	124
196.2	Secondary and unspecified MN of intra-abdominal lymph nodes	3163	2446	114	1123	717	1498	168	4556	747
196	Secondary and unspecified MN of lymph nodes	10878	8252	369	4032	2668	6799	549	16439	3147
198	Secondary MN of other specified sites	15895	11348	405	6859	4529	13865	866	33550	6014

197	Secondary MN of respiratory and digestive systems	17325	12577	478	7270	4601	14113	945	32834	5637
238.7	Other lymphatic and hematopoietic tissues	8762	6424	235	4278	3219	7720	844	14126	2210

Appendix Table 1: Pooled odds ratios for cancers in asthma and COPD by subtype. In exposure description, BN and MN correspondingly stand for 'benign neoplasm' and 'malignant neoplasm'.

ICD9CM Code	Event \ Exposure	Asthma (493.*)	Asthma, unspecified (493.9)	Extrinsic asthma (493.0)	Chronic obstructive asthma (493.2)	Bronchitis, not specified as acute or chronic (490)	Chronic bronchitis (491)	Bronchiectasis (494)	Chronic airway obstruction, not elsewhere classified (496)	Emphysema (492)
210.2	BN of major salivary glands	1.40	1.29	2.73	2.04	0.83	1.93	2.11	2.14	2.66
227.3	BN of pituitary gland and craniopharyngeal duct	1.49	1.53	2.50	1.56	1.29	1.11	1.23	1.36	1.18
227.1	BN of parathyroid gland	1.44	1.41	2.43	1.75	1.03	1.25	1.77	1.69	1.92
225.2	BN of cerebral meninges	1.56	1.44	1.73	2.53	1.48	2.16	2.91	2.38	2.27
226	BN of thyroid glands	1.54	1.57	1.87	1.56	1.00	1.18	1.78	1.26	1.34
153.4	MN of cecum	1.17	0.99	1.33	2.15	0.90	1.86	2.82	2.79	3.18
153.6	MN of ascending colon	1.20	1.03	1.15	2.24	0.86	2.02	2.79	2.81	2.99
153.9	MN of colon, unspecified site	1.25	1.11	1.29	2.35	1.14	2.24	2.33	2.88	3.11
235.2	Neoplasm of uncertain behavior of stomach, intestines, and rectum	1.75	1.48	1.74	3.47	1.45	3.53	3.01	3.89	3.75
153.0	MN of hepatic flexure	1.31	1.12	1.67	2.51	1.00	2.23	3.19	3.29	3.77
153.1	MN of transverse colon	1.18	0.95	1.40	2.48	1.00	2.25	2.74	2.99	3.35
155.1	MN of intrahepatic bile ducts	1.00	0.92	1.75	1.62	0.76	1.21	2.07	1.87	2.26
156.1	MN of extrahepatic bile ducts	0.92	0.79	0.90	1.78	0.71	1.26	3.01	2.08	2.46
174.5	MN of lower-outer quadrant of female breast	1.58	1.57	2.23	1.89	0.85	1.46	0.58	1.60	2.04
174.4	MN of upper-outer quadrant of female breast	1.55	1.53	1.98	1.80	1.01	1.26	1.37	1.49	1.84
162.2	MN of main bronchus	4.64	2.65	2.32	13.73	3.39	22.31	10.75	21.50	33.78
162	MN of trachea, bronchus, and lung	3.08	1.87	1.76	8.69	2.47	13.95	9.20	14.87	25.35
162.4	MN of middle lobe, bronchus or lung	3.88	2.45	1.70	11.02	3.27	14.96	12.52	16.97	29.65
162.8	MN of other parts of bronchus or lung	3.36	2.09	1.36	9.42	2.84	15.55	9.99	15.72	25.01
188.9	MN of bladder, part unspecified	1.47	1.04	1.02	3.65	1.31	4.16	3.90	5.47	6.71
188.2	MN of lateral wall of urinary bladder	1.36	0.93	0.66	3.37	1.14	3.70	3.32	5.38	5.91
196.2	Secondary and unspecified MN of intra-abdominal lymph nodes	0.99	0.89	1.26	1.49	0.67	1.20	1.78	1.99	2.39
196	Secondary and unspecified MN of lymph nodes	1.28	1.13	1.54	2.02	0.93	2.06	2.20	2.73	3.85
198	Secondary MN of other specified sites	1.25	1.05	1.13	2.33	1.07	2.86	2.35	3.82	5.08
197	Secondary MN of respiratory and digestive systems	1.23	1.04	1.20	2.22	0.97	2.61	2.30	3.35	4.25
238.7	Other lymphatic and hematopoietic tissues	2.16	1.85	2.05	4.51	2.37	4.92	7.18	4.92	5.63

Appendix table 2: ICD9CM codes for cancers, asthma and COPD.

ICD-9-CM	Description
493.0	Extrinsic asthma
493.2	Chronic obstructive asthma
493.9	Asthma, unspecified
490	Bronchitis, not specified as acute or chronic
491	Chronic bronchitis
492	Emphysema
494	Bronchiectasis
496	Chronic airway obstruction, not elsewhere classified
210.2	Benign neoplasm of major salivary glands
227.3	Benign neoplasm of pituitary gland and craniopharyngeal duct
227.1	Benign neoplasm of parathyroid gland
225.2	Benign neoplasm of cerebral meninges
226	Benign neoplasm of thyroid glands
153.4	Malignant neoplasm of cecum
153.6	Malignant neoplasm of ascending colon
153.9	Malignant neoplasm of colon, unspecified site
235.2	Neoplasm of uncertain behavior of stomach, intestines, and rectum
153.0	Malignant neoplasm of hepatic flexure
153.1	Malignant neoplasm of transverse colon
174.5	Malignant neoplasm of lower-outer quadrant of female breast
174.4	Malignant neoplasm of upper-outer quadrant of female breast
155.1	Malignant neoplasm of intrahepatic bile ducts
156.1	Malignant neoplasm of extrahepatic bile ducts
162.2	Malignant neoplasm of main bronchus
162	Malignant neoplasm of trachea, bronchus, and lung
162.4	Malignant neoplasm of middle lobe, bronchus or lung
162.8	Malignant neoplasm of other parts of bronchus or lung
238.7	Neoplasm of uncertain behavior of other lymphatic and hematopoietic tissues
188.9	Malignant neoplasm of bladder, part unspecified
188.2	Malignant neoplasm of lateral wall of urinary bladder
180.8	Malignant neoplasm of other specified sites of cervix
196.2	Secondary and unspecified malignant neoplasm of intra-abdominal lymph nodes
196	Secondary and unspecified malignant neoplasm of lymph nodes
198	Secondary malignant neoplasm of other specified sites
197	Secondary malignant neoplasm of respiratory and digestive systems

Appendix 3. Coarsened Exact Matching (CEM).

CEM applies "exact matching" algorithm on coarsened variables and produces strata of matched case-control groups. We manually coarsened age into five categories: younger than 30, 30-49, 50-69, 70-89, and 90 and older. Also, the race was arranged in white, black, Hispanic and other (including Asian or Pacific Islander, Native American, and Other in the original HCUP data). CEM generated weights were used in downstream analysis. Compared to other approximate matching methods, CEM shows superior performance in large datasets and comparable or better results with a controllable imbalance rate [34,35,63,64] and includes the maximum possible number of controls in each stratum. CEM created 96 strata for combinations of accounted confounders in each cohort. For all cohorts, all patients in case groups had at least one matched person among controls. L1 statistics is a nonparametric measure that quantifies imbalance by comparing frequencies of the two groups across each of the strata. Values of L1 vary between 0 and 1, where values close to zero indicate perfect matching.

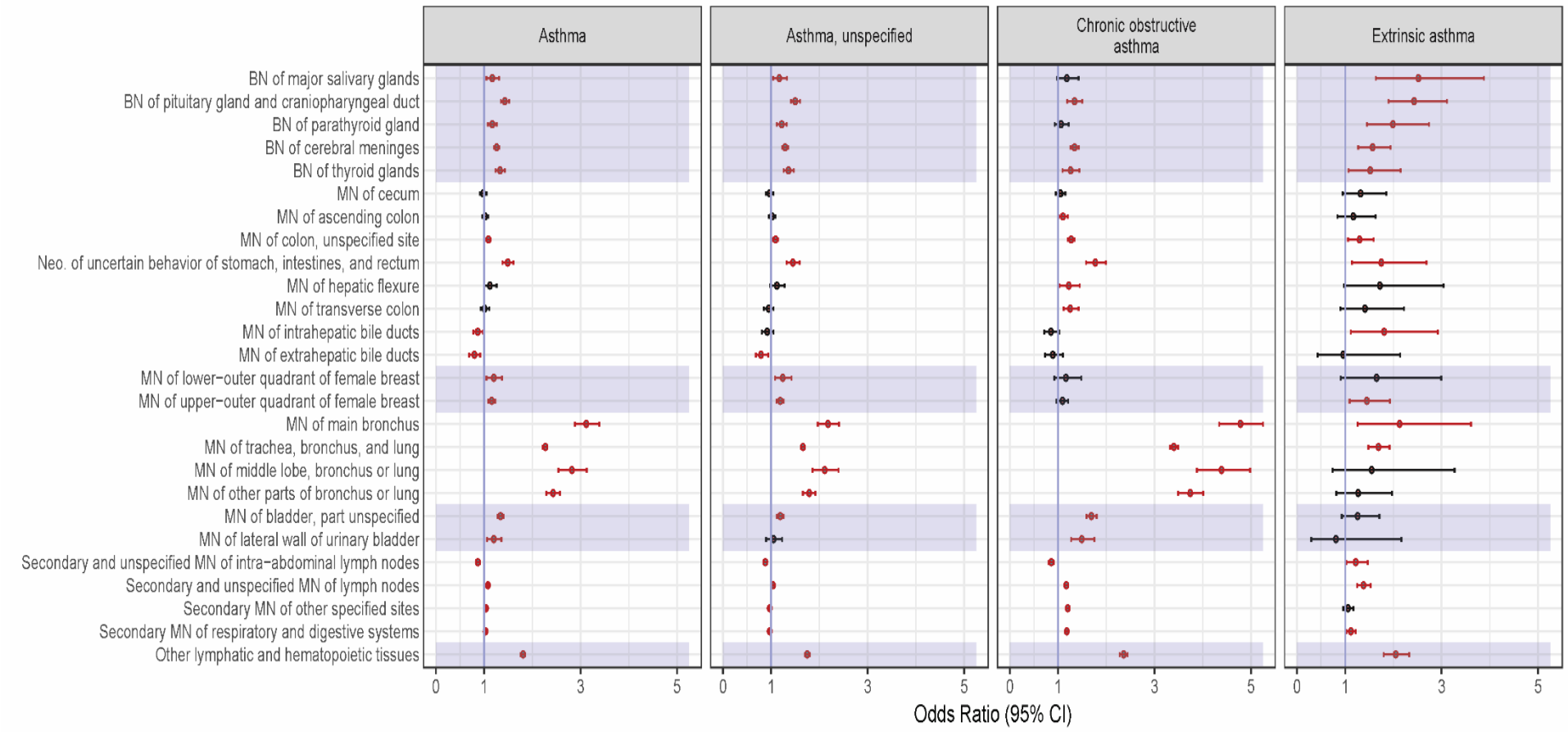
Appendix 4. Advantages of CMH common odds ratio

Pooled odds ratio works on the whole cohort and ignores the fact that differences in the distribution of baseline covariates can have an uncontrolled effect on the measure of association. To address this problem, we used Cochran- Mantel-Haenszel (CMH) [43,44] common odds ratio. CMH common odds ratio calculates odds ratio in each matched subgroups of a population and combines the odds ratios using a weight factor based on the size of subgroups. We calculated CMH common odds ratio overall balanced subgroups, or strata, returned by CEM algorithm. Effectively, CMH common odds ratio gives the estimate of odds ratios on a population that has been balanced on relevant confounders, based on the following formula:

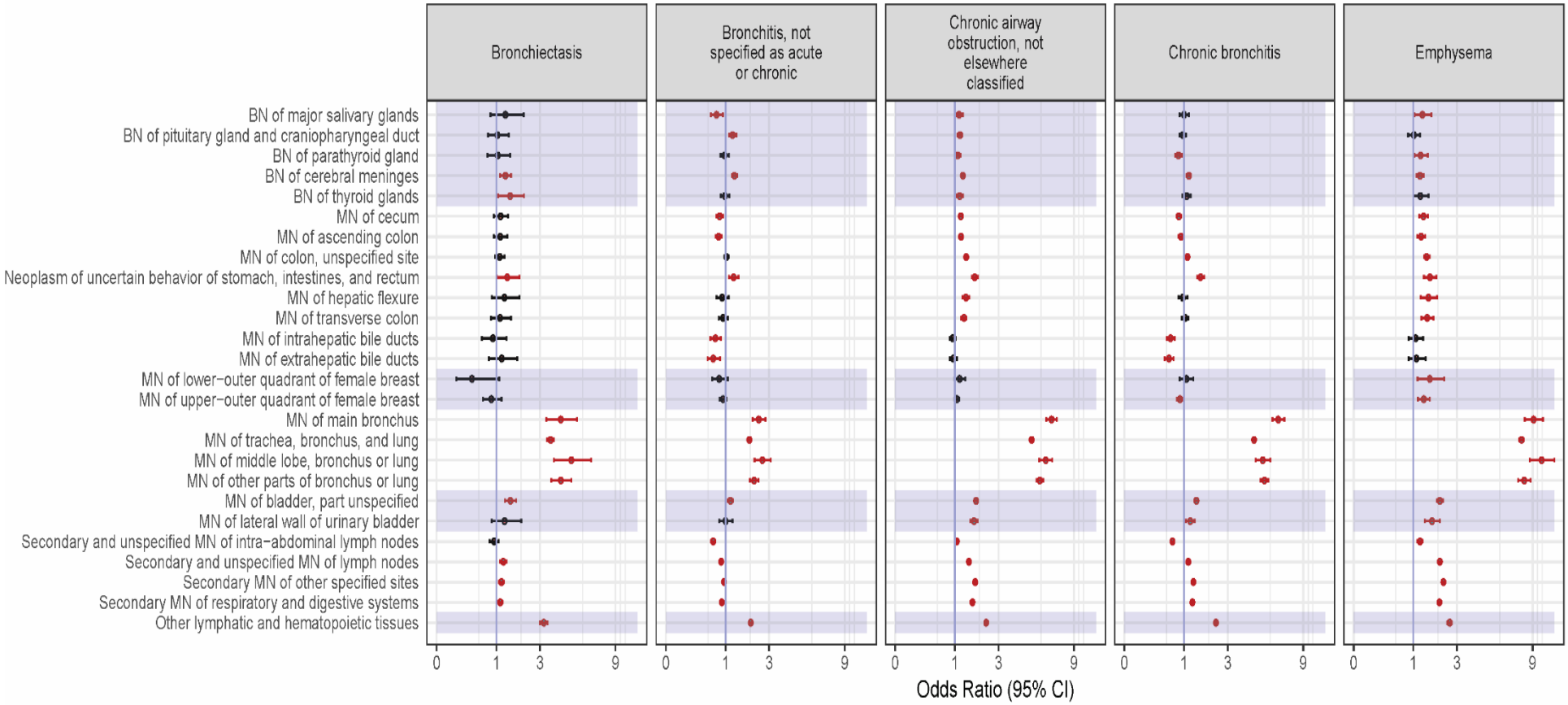
$$\mathrm{OR}_{\{\mathrm{CMH}\}} = \frac{\sum_{\{\mathrm{i}=1\}}^{\mathrm{m}} \frac{a_{11}^{i} a_{22}^{i}}{n_i}}{\sum_{\{\mathrm{i}=1\}}^{\mathrm{m}} \frac{a_{12}^{i} a_{21}^{i}}{n_i}}$$

where m is the number of strata, $a_{\{lj\}}^{i}$ is the (l, j) entry of confusion matrix in i_{th} strata, and n_i is the number of patients in i_{th} stratum.

Appendix Figure 2: Results of individual asthma-neoplasm studies. Each line fragment shows the confidence interval for the odds ratio (dot) of a specific neoplasm(rows) with subtypes of asthma(columns). Clinically significant OR that does not cross one (the blue line) are in the red. In row labels, MN stands for malignant neoplasm, and BN stands for benign neoplasm.



Appendix Figure 3: Results of individual COPD-neoplasm studies. Each line fragment shows the confidence interval for the odds ratio (black dot) of a specific neoplasm(rows) with subtypes of COPD(columns). Significant OR that does not cross one (blue line) are in the red. In row labels, MN stands for malignant neoplasm, and BN stands for benign neoplasm.



Bibliography

1 Control, C. f. D. & Prevention. Asthma facts—CDC's national asthma control program grantees. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention* (2013).

2 Akinbami, O. J., Moorman, J. E. & Liu, X. *Asthma prevalence, health care use, and mortality: United States, 2005-2009*. (US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics Washington, DC, 2011).

3 Akinbami, O. J., Moorman, J. E., Bailey, C., Zahran, H. S., King, M., Johnson, C. A. & Liu, X. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. (2012).

4 Boulet, L.-P. & Boulay, M.-È. Asthma-related comorbidities. *Expert review of respiratory medicine* (2011).

5 Cazzola, M., Segreti, A., Calzetta, L. & Rogliani, P. Comorbidities of asthma: current knowledge and future research needs. *Current opinion in pulmonary medicine* **19**, 36-41 (2013).

6 Luong, K. v. q. & Nguyen, L. T. H. Hyperthyroidism and asthma. *Journal of Asthma* **37**, 125-130 (2000).

7 Zacharisen, M. C. & Fink, J. N. in *Allergy and Asthma Proceedings.* 71-74 (OceanSide Publications, Inc).

8 Wouters, E. F., Reynaert, N. L., Dentener, M. A. & Vernooy, J. H. Systemic and local inflammation in asthma and chronic obstructive pulmonary disease: is there a connection? *Proceedings of the American Thoracic Society* **6**, 638-647 (2009).

9 Togias, A. Systemic effects of local allergic disease. *Journal of allergy and clinical immunology* **113**, S8-S14 (2004).

10 Magnussen, H. & Watz, H. Systemic inflammation in chronic obstructive pulmonary disease and asthma: relation with comorbidities. *Proceedings of the American Thoracic Society* **6**, 648-651 (2009).

11 Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860 (2002).

12 Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, Inflammation, and Cancer. *Cell* **140**, 883-899, doi:http://dx.doi.org/10.1016/j.cell.2010.01.025 (2010).

13 Schottenfeld, D. & Beebe-Dimmer, J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. *CA: a cancer journal for clinicians* **56**, 69-83 (2006).

14 Santillan, A. A., Camargo, C. A. & Colditz, G. A. A meta-analysis of asthma and risk of lung cancer (United States). *Cancer causes and control* **14**, 327-334 (2003).

15 Turner, M. C., Chen, Y., Krewski, D., Ghadirian, P., Thun, M. J. & Calle, E. E. Cancer mortality among US men and women with asthma and hay fever. *American journal of epidemiology* **162**, 212-221 (2005).

16 Liu, X., Hemminki, K., Försti, A., Sundquist, J., Sundquist, K. & Ji, J. Cancer risk and mortality in asthma patients: A Swedish national cohort study. *Acta Oncologica* **54**, 1120-1127, doi:10.3109/0284186X.2014.1001497 (2015).

17 Su, Y.-L., Chou, C.-L., Rau, K.-M. & Lee, C. T.-C. Asthma and risk of prostate cancer: a population-based case-cohort study in Taiwan. *Medicine* **94** (2015).

18 Linabery, A. M., Prizment, A. E., Anderson, K. E., Cerhan, J. R., Poynter, J. N. & Ross, J. A. Allergic diseases and risk of hematopoietic malignancies in a cohort of postmenopausal women: A report from the Iowa Women's Health Study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **23**, 1903-1912, doi:10.1158/1055-9965.EPI-14-0423 (2014).

19 Gomez-Rubio, P., Zock, J.-P., Rava, M., Marquez, M., Sharp, L., Hidalgo, M., Carrato, A., Ilzarbe, L., Michalski, C., Molero, X., Farré, A., Perea, J., Greenhalf, W., O'Rorke, M., Tardón, A., Gress, T., Barberà, V., Crnogorac-Jurcevic, T., Domínguez-Muñoz, E., Muñoz-Bellvís, L., Alvarez-Urturi, C., Balcells, J., Barneo, L., Costello, E., Guillén-Ponce, C., Kleeff, J., Kong, B., Lawlor, R., Löhr, M., Mora, J., Murray, L., O'Driscoll, D., Peláez, P., Poves, I., Scarpa, A., Real, F. X. & Malats, N. Reduced risk of pancreatic cancer associated with asthma and nasal allergies. *Gut* **66**, 314-322, doi:10.1136/gutjnl-2015-310442 (2017).

20 Kallen, B., Gunnarskog, J. & Conradson, T. Cancer risk in asthmatic subjects selected from hospital discharge registry. *European Respiratory Journal* **6**, 694-697 (1993).

21 Talbot-Smith, A., Fritschi, L., Divitini, M. L., Mallon, D. F. & Knuiman, M. W. Allergy, atopy, and cancer: a prospective study of the 1981 Busselton cohort. *American journal of epidemiology* **157**, 606-612 (2003).
22 VENA, J. E., BONA, J. R., BYERS, T. E., MIDDLETON JR, E., SWANSON, M. K. & GRAHAM, S. Allergy-related diseases and cancer: an inverse association. *American journal of epidemiology* **122**, 66-74 (1985).
23 VESTERINEN, E., PUKKALA, E., TIMONEN, T. & AROMAA, A. Cancer incidence among 78 000 asthmatic patients. *International journal of epidemiology* **22**, 976-982 (1993).
24 Eriksson, N., Holmen, A., Högstedt, B., Mikoczy, Z. & Hagmar, L. A prospective study of cancer incidence in a cohort examined for allergy. *Allergy* **50**, 718-722 (1995).
25 Agency for Healthcare Research and Quality. HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP) *2005-2011*.
26 Andrews, R. M. Statewide Hospital Discharge Data: Collection, Use, Limitations, and Improvements. *Health Services Research* **50**, 1273-1299, doi:10.1111/1475-6773.12343 (2015).
27 Houchens, R. *Inferences with HCUP State Databases Final Report*, <http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp.> (
28 *HCUP Databases. Healthcare Cost and Utilization Project (HCUP). April 2017. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp; Last modified 4/18/17.*
29 SEDD Database Documentation. Healthcare Cost and Utilization Project (HCUP). December 2017. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/db/state/sedddbdocumentation.jsp; Last modified 9/13/17.
30 Fingar KR, O. P., Barrett ML, Steiner CA. *Using the HCUP Databases to Study Incidence and Prevalence*, <http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp.> (
31 *USER GUIDE: HCUP SUPPLEMENTAL VARIABLES FOR REVISIT ANALYSES*, <https://www.hcup-us.ahrq.gov/toolssoftware/revisit/UserGuide_SuppRevisitFilesCD.pdf> (
32 Shojaee, A., Ranasinghe, I. & Ani, A. Revisiting Causality Inference in Memory-less Transition Networks. *eprint arXiv:1608.02658* (2016).
33 Rothman, K. J. & Greenland, S. Causation and Causal Inference in Epidemiology. *http://dx.doi.org/10.2105/AJPH.2004.059204*, doi:S144.pdf (2011).
34 Iacus, S. M., King, G. & Porro, G. CEM: software for coarsened exact matching. *Journal of statistical Software* **30**, 1-27 (2009).
35 Iacus, S. M., King, G., Porro, G. & Katz, J. N. Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 1-24 (2012).
36 Iacus, S. M., King, G. & Porro, G. CEM: software for coarsened exact matching. *Journal of Statistical Software* **30**, 1-27 (2009).
37 Hamar, G. B., Rula, E. Y., Coberley, C., Pope, J. E. & Larkin, S. Long-term impact of a chronic disease management program on hospital utilization and cost in an Australian population with heart disease or diabetes. *BMC health services research* **15**, 174 (2015).
38 Hicks, C. W., Hashmi, Z. G., Velopulos, C., Efron, D. T., Schneider, E. B., Haut, E. R., Cornwell, E. E. & Haider, A. H. Association between race and age in survival after trauma. *JAMA surgery* **149**, 642-647 (2014).
39 Nazari, S. S. H., Noroozi, M., Soori, H., Noroozi, A., Mehrabi, Y., Hajebi, A., Sharifi, H., Higgs, P. & Mirzazadeh, A. The effect of on-site and outreach-based needle and syringe programs in people who inject drugs in Kermanshah, Iran. *International Journal of Drug Policy* **27**, 127-131 (2016).
40 Mason, R. J., Moazzez, A., Moroney, J. R. & Katkhouda, N. Laparoscopic vs open appendectomy in obese patients: outcomes using the American College of Surgeons National Surgical Quality Improvement Program database. *Journal of the American College of Surgeons* **215**, 88-99 (2012).
41 Haider, A. H., David, J.-S., Zafar, S. N., Gueugniaud, P.-Y., Efron, D. T., Floccard, B., MacKenzie, E. J. & Voiglio, E. Comparative Effectiveness of In-Hospital Trauma Resuscitation at a French Trauma Center and Matched Patients Treated in the United States. *Annals of surgery* **258**, 178 (2013).
42 Hametner, C., Kellert, L. & Ringleb, P. A. Impact of sex in stroke thrombolysis: a coarsened exact matching study. *BMC neurology* **15**, 10 (2015).
43 Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute* **22**, 719-748 (1959).
44 Cochran, W. G. Some methods for strengthening the common χ 2 tests. *Biometrics* **10**, 417-451 (1954).

45 Abdi, H. Holm's sequential Bonferroni procedure. *Encyclopedia of research design* **1**, 1-8 (2010).
46 Zhang, X., Holt, J. B., Lu, H., Wheaton, A. G., Ford, E. S., Greenlund, K. J. & Croft, J. B. Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *American journal of epidemiology* **179**, 1025-1033 (2014).
47 de-Torres, J. P., Wilson, D. O., Sanchez-Salcedo, P., Weissfeld, J. L., Berto, J., Campo, A., Alcaide, A. B., García-Granero, M., Celli, B. R. & Zulueta, J. J. Lung cancer in patients with chronic obstructive pulmonary disease. Development and validation of the COPD Lung Cancer Screening Score. *American journal of respiratory and critical care medicine* **191**, 285-291 (2015).
48 Yao, H. & Rahman, I. Current concepts on the role of inflammation in COPD and lung cancer. *Current opinion in pharmacology* **9**, 375-383 (2009).
49 Young, R. P., Hopkins, R. J., Christmas, T., Black, P. N., Metcalf, P. & Gamble, G. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *European Respiratory Journal* **34**, 380-386 (2009).
50 Barnes, P. J. Therapeutic approaches to asthma–chronic obstructive pulmonary disease overlap syndromes. *Journal of Allergy and Clinical Immunology* **136**, 531-545, doi:http://dx.doi.org/10.1016/j.jaci.2015.05.052 (2015).
51 Pandya, D., Puttanna, A. & Balagopal, V. Systemic Effects of Inhaled Corticosteroids: An Overview. *The Open Respiratory Medicine Journal* **8**, 59-65, doi:10.2174/1874306401408010059 (2014).
52 Hong, J. C. in *Essentials of Pharmacology for Anesthesia, Pain Medicine, and Critical Care* 313-325 (Springer, 2015).
53 Nicoloff, J. T., Fisher, D. A. & Appleman, M. D. The role of glucocorticoids in the regulation of thyroid function in man. *Journal of Clinical Investigation* **49**, 1922-1929 (1970).
54 HILLS JR, S. R., REISS, R. S., FORSHAM, P. H. & GEORGE, W. T. The effect of adrenocorticotropin and cortisone on thyroid function: thyroidadrenocortical interrelationships. *The Journal of Clinical Endocrinology* **10**, 1375-1400 (1950).
55 Van Staa, T., Leufkens, H. & Cooper, C. Use of inhaled corticosteroids and risk of fractures. *Journal of Bone and Mineral Research* **16**, 581-588 (2001).
56 Faul, J. L., Wilson, S. R., Chu, J. W., Canfield, J. & Kuschner, W. G. The effect of an inhaled corticosteroid on glucose control in type 2 diabetes. *Clinical medicine & research* **7**, 14-20 (2009).
57 Gilkes, D. M., Semenza, G. L. & Wirtz, D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nature Reviews Cancer* **14**, 430-439 (2014).
58 Ma, X.-J., Dahiya, S., Richardson, E., Erlander, M. & Sgroi, D. C. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast cancer research* **11**, R7 (2009).
59 Chang, H. Y., Sneddon, J. B., Alizadeh, A. A., Sood, R., West, R. B., Montgomery, K., Chi, J.-T., Van De Rijn, M., Botstein, D. & Brown, P. O. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS biology* **2**, e7 (2004).
60 Merrill, R. M., Isakson, R. T. & Beck, R. E. The association between allergies and cancer: what is currently known? *Annals of Allergy, Asthma & Immunology* **99**, 102-117, doi:https://doi.org/10.1016/S1081-1206(10)60632-1 (2007).
61 Rittmeyer, D. & Lorentz, A. Relationship between Allergy and Cancer: An Overview. *International Archives of Allergy and Immunology* **159**, 216-225 (2012).
62 McWhorter, W. P. Allergy and risk of cancer. A prospective study using NHANESI followup data. *Cancer* **62**, 451-455 (1988).
63 Wells, A. R., Hamar, B., Bradley, C., Gandy, W. M., Harrison, P. L., Sidney, J. A., Coberley, C. R., Rula, E. Y. & Pope, J. E. Exploring Robust Methods for Evaluating Treatment and Comparison Groups in Chronic Care Management Programs. *Population Health Management* **16**, 35-45, doi:10.1089/pop.2011.0104 (2013).
64 Esareyy, J. Causal Inference with Observational Data. (2015).