Population-based Large-scale longitudinal analysis of the relationship between asthma and neoplasia

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.

At a Glance Commentary:

Over the past three decades, consensus on a relationship between asthma and risk of neoplasm has not been reached. The debate rises from controversial, population-level evidence that has emerged from different methods, data sources and cohorts as well as the lack of the mechanism by which allergic airway inflammation contributes to malignant transformation.

The present study includes the largest longitudinal dataset cohort of patients, to date, with asthma or COPD. We used machine learning methods for predicting causal relationships combined with standard 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 novel relationships between allergic asthma and benign neoplasms of glands. Further, we identified the association between COPD and hematological and lymphatic malignancies and confirmed the relationship with allergic asthma. These findings rectify contradictory results from other large studies on asthma and its connection to lung cancer. Moreover, we show that the sets of neoplasms in allergic asthma and COPD have significant differences, suggesting the involvement of various pathways.

Abstract

Background: A relationship between asthma and the development of cancer has been reported, but remains controversial because of contradictory results from population studies.

Methods: Two large-scale observational, all payers claim datasets from the US-based Healthcare Cost and Utilization Project (HCUP) were used for discovery and validation. Associations between asthma and neoplasms were discovered using Causal Inference Using Composition of Transitions (CICT) in data from the State of Florida. Validation was completed in eight case-control cohorts for patients exposed to subtypes of asthma and COPD using data from the State of California. Control groups were matched on gender, age, race and history of tobacco abuse. Odds ratio analysis with Bonferroni-Holm correction measured the association of asthma and COPD with 26 different neoplasms. ICD9CM codes were used to identify exposures and outcomes.

Findings
CICT identified 17 association between asthma and neoplasms in the discovery datasets. In the validation dataset, 208 case-control, analyses were conducted between subtypes of Asthma (N= 999,370, male= 33%, age= 50) and COPD (N=715,971, male = 50%, age=69) with the corresponding matched control groups (N=8,400,004, male= 42%, age= 47). Allergic asthma was associated with benign neoplasms of the meninges, salivary, pituitary, parathyroid, and thyroid glands (OR:1.52 to 2.52), and primary malignant neoplasms of breast, hematopoietic, and lymphatic system (OR: 1.45 to 2.05). COPD was associated with neoplasms in the lung, bladder and hematopoietic system.

Interpretation: The combination of machine learning for the discovery of associations with epidemiological methods for validation shows that allergic asthma is associated with neoplastic transformation of common structures of secretory organs.

Funding

None.

Word count:257

Keywords:
Asthma, Neoplasm, Endocrine, Glandular Neoplasia, Etiology, Chronic Obstructive Pulmonary Disease, Malignancy,

Introduction

Asthma is a common chronic disease affecting nearly 10% of the US population that is associated with annual costs of approximately 80 billion dollars (1-3). It has been suggested that the chronic systemic inflammation that is associated with asthma contributes to the development and severity of comorbidities(4-6) including obesity, gastroesophageal reflux, psychological problems, chronic infections, and hormonal disturbances (7-10). Chronic inflammation is also known for its effect on tumorigenesis and increased risk of cancer(11-13), through pathways that some are also activated in asthma. Understanding potential associations between asthma and neoplasia could inform both the pathobiology of chronic airway inflammation as well as the mechanisms that underlie the development of malignancy.

Definitive clinical and population-level evidence on the relationship between asthma and the development of neoplasms does not exist. A meta-analysis suggested a 1.8 fold increase in lung cancer risk 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 allergic asthma, identified by International Classification of Disease version 9 clinical modifications (ICD9CM) code 493.0, and chronic obstructive asthma (ICD: 493.2), which is a blend of obstructive and allergic manifestations. Moreover, previous studies have been limited by a lack of statistical power, small sample sizes, potential biases in participant selection, exposure measurement (15, 20-24), and the limited type of neoplasms that were studied. Considering the substantial cost associated with asthma and neoplasms, and the contradictory results of previous reports, additional studies are required to address these limitations and to evaluate associations with specific neoplasms (15, 21).

To overcome the deficiencies of the previous studies, we studied two large-scale longitudinal administrative, all payers US-based datasets of over one hundred million admission records, covering 95% of community hospital discharges in Florida and California. We focused on the association patterns between neoplasms, three subtypes of asthma and four subtypes of Chronic Obstructive Pulmonary Disease (COPD) as coded in ICD9CM. We developed a novel machine learning method to identify causal relationships in high-dimensional data and validated the new findings with standard statistical methodology. We used a rigorous method and the largest dataset of asthma and COPD patients to be analyzed to date for discovery and validation. Our findings suggest that asthma is associated with a range of benign glandular neoplasms and malignant neoplasms in other organs, especially secretory tissues. The results also provide new information on the debated association between asthma and neoplasms.

Materials and Methods

Data Sources

Deidentified observational claim data from Healthcare Cost and Utilization Project (HCUP) was used for discovery and validation. HCUP is a federal-state-industry partnership sponsored by the Agency for Healthcare Research and Quality(25). The **S**tate **I**npatient **D**atabase (SID) and **E**mergency **D**epartment **D**atabase (EDD)(25) 2004 to 2014 from the state of Florida was used for exploratory analyses. Validation was conducted using HCUP California SID and ED(25) from 2005 to 2011. HCUP data is a census of all discharges(26) and SID contains the records for all-payers, including the uninsured, comprising all non-federal acute care hospitals in participating states(27), and together, about 95 percent of all U.S. community hospital discharges (28, 29) (30). The ED contains all visits to the affiliated emergency department that did not result in a hospitalization. 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 can connect all encounters across the SID and ED to create a longitudinal record for each patient. The pseudo identifier, 'VisitLink,' is assigned through a verification process to ensure the correct assignment of a unique identifier to all admissions of a patient (31).

Exposures and outcome events were identified using ICD9CM codes of primary diagnosis and up to 24 comorbidities. 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 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. The causal prediction method employed in the discovery phase of these studies uses consequent diagnoses that occur over time between visits to build a network of patients transitions between clinical conditions. Accordingly, we included only patients with more than one visit. Patients were excluded if the VisitLink information was missing or if the patient died during asthma or COPD hospitalizations. Population characteristics are shown in Table 1, and the CONSORT (32) diagram in Figure 1 summarizes the data preparation process.

Discovery analysis

Evaluation of all associations between subtypes of asthma and malignancies could be computationally prohibitive and result in coincidental findings due to a large number of multiple comparison. Accordingly, to identify more likely hypotheses and limit the number of assessments, we used the Causal Inference using Composition of Transitions (CICT) method(33). CICT uses population-level longitudinal data and information about the stochastic process of patients' consequent health encounters to find potential causal relationships about clinical conditions (34). CICT is based on the concept that the type and frequency of events before and after an emerging clinical condition are different than for a random event. For example, observing a population of patients with a particular neoplasm, the type and frequency of diagnoses between the two consecutive time points will be different compared to the diagnoses in a population without that neoplasm. To capture these differences, CICT defines detailed features in the distribution of consequent diagnoses in large-scale population-level data and employes supervised machine learning methods to learn the patterns specific to causal, random and effect phenomena. Then, CICT uses the learned patterns to predict potential causal relationships among various clinical conditions. For this study, we first created a network of patients that transitioned between pairs of consequent clinical conditions in our discovery dataset from the state of Florida. The clinical conditions were identified using the ICD9CM code of primary diagnosis for admission. CICT was applied to identify the relationships between asthma and neoplasms. Figure 2 and Table 2 show the CICT hypotheses related to potential asthma induced neoplasms.

Validation analysis

To validate the hypotheses that we discovered in the previous step, we used odds ratio analysis, which is a measure of association, in the validation dataset from California. Also, to discriminate the effect of asthma and COPD case mix, we added odds ratio analysis for subtypes of COPD with the same neoplasms to further validate the results. 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= 999370), 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. Four cohorts were designed for COPD subtypes 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 208 individual studies were designed between eight 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)(35, 36) 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 (37). Previous research used different user-specified criteria for L1 statistics (38-43). In this study, a threshold of L1 statistics less than 0.2 after CEM matching was considered acceptable.

Statistical analysis

Descriptive statistics of individual characteristics at baseline were calculated for the whole population, asthma and COPD subtype 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) (44, 45) 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 eight matched case-control populations for subtypes of asthma and COPD (row 1 of Table 3). A Bonferroni-Holm (46) correction was applied to each asthma and COPD subtype cohort to control family-wise error rate. The significance threshold was established at $P \leq 0.o5$. All statistical analyses were performed using F# and R.

Results

CICT identified associations between asthma and neoplasms

Among 524 neoplasms encoded in ICD9CM codes, CICT identified seventeen benign and malignant neoplasms associated with asthma(Table 2, Figure 2). The strongest association was predicted between chronic obstructive asthma (ICD9:493.2) and malignant neoplasm of middle lobe, bronchus or lung (ICD9:162.4, estimate: 0.96, 95%CI: 0.85 - 1.00). Other predicted associations with subtypes of asthma were malignant neoplasm of transverse colon(153.1, 0.82, CI:0.72 - 0.93), bile ducts(155.1, 0.94, CI:0.83 - 1.00), bladder (188.9, 0.78, CI: 0.56 - 1.00) and lymphatic and hematopoietic tissues (238.7, 0.74 , CI:0.63 - 0.86). The strongest relationship of asthma with benign neoplasms was for benign neoplasms of salivary glands (210.2, 0.93, CI: 0.81 - 1.00). Further causal associations were predicted for benign glandular neoplasms of parathyroid (227.1, 0.86, CI: 0.75 - 0.97), thyroid (226, 0.90, CI: 0.79 - 1.00) and pituitary glands (227.3, 0.73, CI:0.63 - 0.83). Therefore CICT indicated that there is a relationship between asthma and neoplasia in glands, gastrointestinal, biliary ducts, lymphatic and hematopoietic system, and breast.

Odds ratio analysis of CICT predicted associations

We used epidemiological reasoning methods to evaluate 17 CICT hypotheses along with nine other competing hypotheses suggested by domain experts(GC and JG). We examined the relationship of suggested neoplasms with various subtypes of asthma and COPD. Four cohorts were designed for asthma and its three subtypes: combined asthma (N= 999370), extrinsic asthma (N= 28149, male= 31%, age= 47), chronic obstructive asthma (N= 235446, male= 38%, age= 65), and asthma unspecified (N= 853556, male= 32%, age= 48). Four cohorts were designed for subtypes of COPD: chronic bronchitis (N= 390862, male= 48%, age=68), emphysema (N= 97956 , male= 54% , age= 68), bronchiectasis (N= 29539, male= 39%, age= 71), and chronic airway obstruction, not elsewhere classified (N= 715957, male= 50%, age= 69).

Approximately 8.4 million matched controls were included in each cohort: extrinsic asthma (N= 8,398,723), chronic obstructive asthma (N = 8,400,004), unspecified asthma (N = 8,400,004), chronic bronchitis (N = 8,400,004), emphysema (N = 8,397,079), bronchiectasis (N = 8,399,392) chronic airway obstruction (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 208 different studies are similar.(35)

A total of 139,038 incident cases of neoplasms were identified among patients with exposure to asthma, 338,418 in those exposed to COPD, and 518,138 in the control cohort that had no record of exposure to either respiratory disorder. Age, sex, race, and smoking distribution in the overall validation population and asthma and COPD subtypes are shown in Table 1. A history of asthma in 10.63% of the total included population and a history of COPD of 6.95% was observed, consistent with national averages (3) (1) (47). On average, patients with chronic obstructive asthma and subtypes of COPD tended to be older than those with extrinsic asthma and controls. Patients with extrinsic asthma, asthma unspecified and bronchiectasis were less likely to be former smokers than patients with other subtypes of asthma and COPD. Bronchiectasis and subtypes of asthma were more likely to be women than 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 race, and smokers. Unspecified asthma was the most frequent diagnostic code and observed in 85.41% of the asthma patients.

The CMH common odds ratio for asthma, COPD, and their subtypes compared to the control cohort for individual cancers are presented in Table 3 and Appendix Figures 2 and Figure 3. CMH OR is a weighted aggregation of ORs in balanced strata of the population and is considered a more accurate estimate of association than pooled OR which is a raw, and often higher, estimate of association in the cohort (appendix 4). The balancing, conducted by CEM, ensures for the similarity of potential confounders between case and controls in each stratum. To provide a basis for comparison we presented pooled odds ratios in Appendix Table 1. Extrinsic asthma was associated with benign

neoplasms of salivary, parathyroid, thyroid and pituitary and craniopharyngeal duct (OR:1.52-2.52); and with malignant neoplasms of intrahepatic bile ducts and upper-outer quadrant of female breast (OR 1.45-1.81). Among subtypes of asthma, obstructive asthma had the highest risk of various pulmonary cancers (OR: 3.40 to 4.78), compared with extrinsic asthma which shows a weaker and often insignificant association with lung cancers. Emphysema and Chronic airway obstruction, not elsewhere classified (496) had the strongest associations with various primary malignant neoplasms (OR 1.07 – 9.94). Also, emphysema showed the strongest association with lung cancers (OR 7.93-9.94). (table 3 and appendix figure 3). The actual number of cases in the COPD and asthma cohort are reported in Table 4, and the pooled odds ratio are presented in appendix table 1.

Asthma and COPD subtypes 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: 2.33-3.23).

Discussion

This study found novel associations between obstructive airway diseases and neoplasms in the largest known population-based analysis of patients with asthma and COPD. Innovative data mining methods were used along with extensive empirical analysis with conventional epidemiological reasoning methods. The novel methods and the large size of the cohort allowed us to identify previously unknown relationships and assess associations with more precision than previous studies. We provide new insights into the long-time debate about the relationship between asthma and neoplasia. Moreover, this is the first study of this scope suggesting that inflammation leads to development of neoplasia. A possible mechanism is that the inflammatory response that contributes to the hormonal and metabolic disturbances that are associated with asthma(e.g., obesity, diabetes, osteoporosis and thyroid dysfunction) (7-10) may also contribute to dysregulation of cellular growth and development of neoplasia in common glandular structures such as secretory epithelium.

An important aspect of this study is the robust two-step design. In the first step, we used CICT, a machine learning method to directly predict potential causal relationships. The use of causality prediction for hypothesis generation is a departure from standard methods that attempt to infer causality using mathematical or statistical models.We used a second large-scale dataset and standard tools of epidemiological reasoning to validate the hypotheses we produced in the first step. Although CICT is a method for identifying potentially causal relationships, the standard epidemiological reasoning methods we used here are designed only to validate the association claim and not the causality one. Nevertheless, the combined use of knowledge discovery methods and confirmation by standard methods helped to draw a broader picture of associations between asthma and neoplasia. We suggest that adding causality prediction to our toolbox can help in identifying the relationships and interactions between diseases in an expedient and cost-effective manner.

The contrast between the allergic and chronic obstructive asthma was critical in the identification of key associations with neoplasms. Chronic obstructive asthma, a mixed coding that covers the frequent case of patients with mixed allergic and obstructive manifestations, showed similarities to both asthma and COPD subtypes, but often followed a neoplasia pattern similar to COPD. This finding may help explain that the contradictory results described in previous studies could be due to differences in the case mix of patients who had various degrees of obstructive symptoms instead of a pure allergic form of asthma. Specifically, allergic asthma showed a small or insignificant effect (CI contains one) on bronchial or lung cancer, contrasting with chronic obstructive asthma which shows a pattern similar to the emphysema COPD, consistent with known association of COPD as a risk factor for lung cancers(48-50). This finding suggests that the diagnosis of obstructive asthma could be a misclassification that is often dominated by obstructive mechanisms rather than asthma.

The different localization pattern of neoplasms related to asthma and COPD in our results, is consistent with our mechanistic understanding of the effects of smoking. Both chronic asthma and COPD can contribute to tissue injury that results in systemic inflammation. In addition, recent studies have established a direct link between hypoxia and the composition and organization of the extracellular matrix (ECM) (51). Emerging data indicate that ECM has a crucial role in metastasis 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 (51-53). These molecular findings are in line with our results that show an increased risk of breast, gastrointestinal and lung cancer in asthma. Compared to chronic bronchitis(491; OR:0.92), bronchiectasis(494, OR:0.87), and chronic airway obstruction(496; OR:1.07), the higher odds of malignant upper-outer quadrant of female breast cancer in allergic asthma(493.0, OR:1.45, CI:1.09-1.93) and lower odds of lung and bronchial cancer suggests that the proinflammatory effect of asthma is a more systemic effect on ECM whereas COPD induces a local effect.

Exposure to high dose inhaled corticosteroid (ICS) and oral corticosteroid (OCS) therapy could also play a role in the association between asthma and gland neoplasms. Additionally corticosteroids have been associated with a down-regulation of thyroid function(54-56). However, a connection between thyroid or parathyroid neoplasms with corticosteroids has not been established. This suggests that the associations we identified are not related to corticosteroid. Traditionally, diabetes, obesity, and osteoporosis have been attributed to ICS and OCS therapy in patients with asthma(57, 58). Nevertheless, our findings are in line with recent studies which failed to establish an association between regular or high dose inhaled corticosteroid (ICS) and osteoporosis(59) or diabetes(60) and suggested that the chronic respiratory disease process itself could be responsible. Interestingly, in our study, allergic asthma showed an association with malignancies in ductal structures including breast and the intrahepatic biliary system. Both elevated(61) and reduced risks(23) of lymphatic and hematopoietic cancer have been reported. Our results, show increased odds of lymphatic and hematopoietic cancers in the presence of allergic asthma. This finding provides new insight into the relationship of asthma and lymphatic and hematopoietic cancer (62, 63).

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. Nevertheless, our study provides the basis for study design by generating ORs that can be used in power calculation by providing accurate effect sizes. 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 with more severe asthma or later stages of COPD. Still, employing proper methods on these large datasets representative of the population at large provides the opportunity to improve the signal to noise ratio. In this study, the large cohort population allowed us to capture the relationship between allergic asthma and benign neoplasms and uncover the differences between obstructive and allergic asthma in the development of lung cancer.

In conclusion these results show, the value of using new analytics along with standard epidemiological methods with large-scale data to reveal signals that are invisible in small-scale data, limited data acquisition practices, or obsessively filtered data. We suggest that, in the era of data abundance, data-driven mass production, validation, and reporting of insights on a specific topic can be a new way of reporting clinical findings. This type of research responds to the increasing demand for accelerated research and better evidence. The uniformity of data preparation, analysis, and replication of 208 individual studies in our research makes the results readily comparable. This type of research could be used in parallel with review articles and meta and systematic analyses which integrate evidence from multiple sources. Such an approach helped us bring fresh insights into the controversies surrounding asthma, COPD, and neoplasms.

The molecular mechanism(s) that link asthma with its comorbidities, especially the potential connection between asthma-related dysregulation and cellular growth and neoplasms, has yet to be studied. Further research to identify and understand causal associations between asthma and neoplasms can provide important clues to biological processes and assist in the identification of asthma-specific comorbidities.

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

Figure1: CONSORT diagram

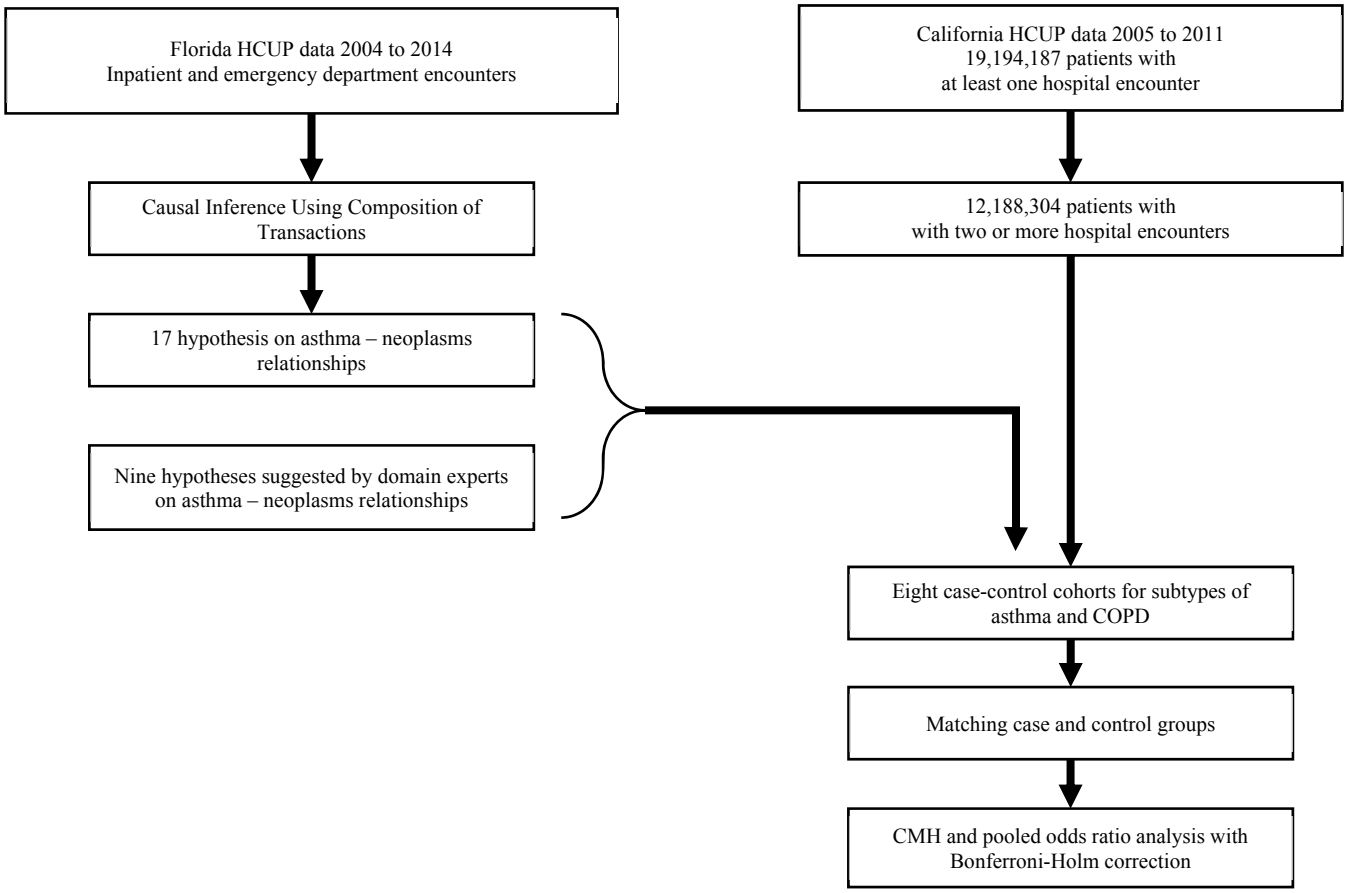


Table 1: Basic characteristics of patients with asthma and COPD by subtype - California HCUP SID, EDD 2005-2011.

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 2004-2014 data. Malignant neoplasms are abbreviated to MN and benign neoplasms to BN. Edge values are CICT predicted strength for the causal hypothesis (0 to 1). Nodes are color-coded and organized based on their clinical relevance: asthma subtypes (center, black), 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.

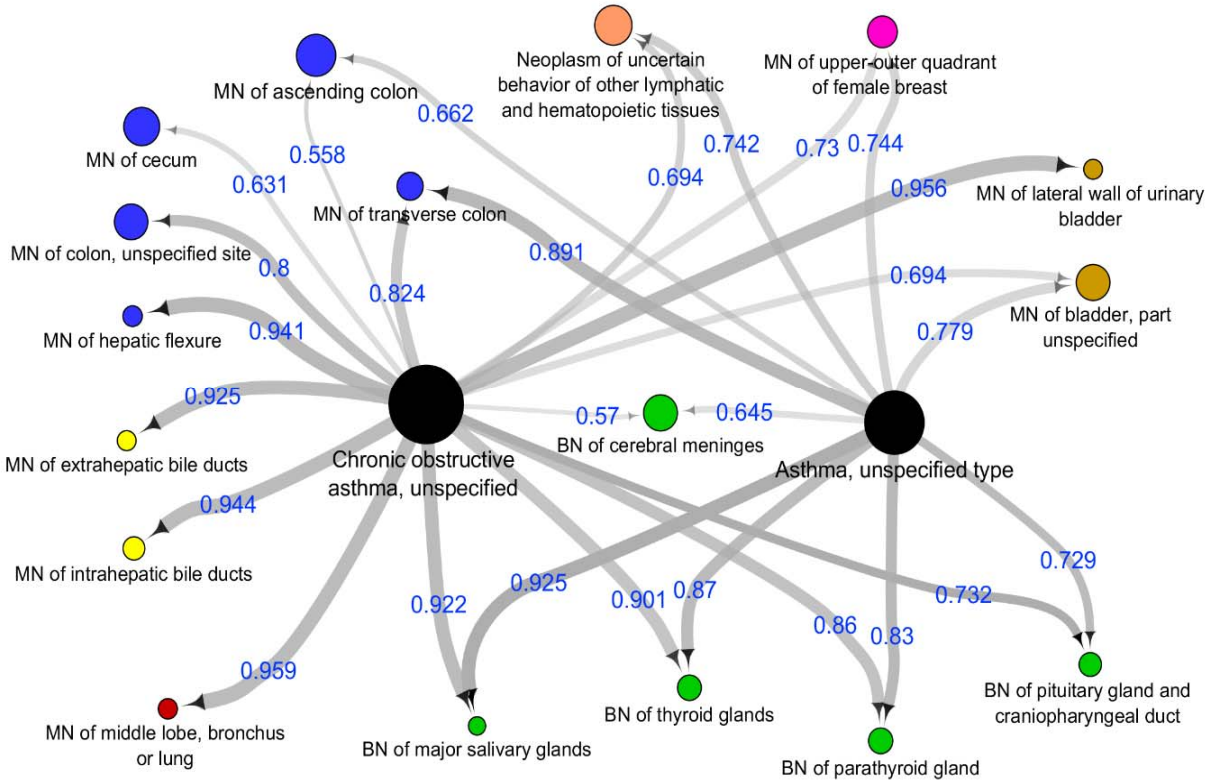


Table 2: CICT generated predicted strength for the causal hypothesis on the effect of subtypes of asthma (columns) on inducing neoplasms (rows) using Florida 2004-2014 SID and ED data. The number represents the predicted strength of causal hypotheses, range 0-1; 1 indicating a strong probability of causality, 0, indicating a random relationship. The 95% confidence interval of each prediction is given in the parenthesis. Malignant neoplasms are abbreviated to MN and benign neoplasms to BN.

ICD9CM code	Event \ Exposure	Asthma, unspecified type, with (acute) exacerbation (ICD: 493.9)	Chronic obstructive asthma, unspecified (ICD:493.2)
210.2	BN of major salivary glands	0.93 (0.81 - 1.00)	0.92 (0.81 - 1.00)
227.3	BN of pituitary gland and craniopharyngeal duct	0.73 (0.57 - 0.89)	0.73 (0.63 - 0.83)
227.1	BN of parathyroid gland	0.83 (0.72 - 0.94)	0.86 (0.75 - 0.97)
225.2	BN of cerebral meninges	0.65 (0.28 - 1.00)	0.57 (0.33 - 0.81)
226	BN of thyroid glands	0.87 (0.76 - 0.98)	0.90 (0.79 - 1.00)
153.4	MN of cecum	N/A	0.63 (0.43 - 0.83)
153.6	MN of ascending colon	0.66 (0.55 - 0.77)	0.56 (0.44 - 0.67)
153.9	MN of colon, unspecified site	N/A	0.80 (0.70 - 0.90)
153.0	MN of hepatic flexure	N/A	0.94 (0.83 - 1.00)
153.1	MN of transverse colon	0.89 (0.78 - 1.00)	0.82 (0.72 - 0.93)
155.1	MN of intrahepatic bile ducts	N/A	0.94 (0.83 - 1.00)
156.1	MN of extrahepatic bile ducts	N/A	0.93 (0.82 - 1.00)
174.4	MN of upper-outer quadrant of female breast	0.74 (0.64 - 0.84)	0.73 (0.63 - 0.83)
162.4	MN of middle lobe, bronchus or lung	N/A	0.96 (0.85 - 1.00)
188.9	MN of bladder, part unspecified	0.78 (0.56 - 1.00)	0.69 (0.59 - 0.80)
188.2	MN of lateral wall of the urinary bladder	N/A	0.96 (0.85 - 1.00)
238.7	Other lymphatic and hematopoietic tissues	0.74 (0.63 - 0.86)	0.69 (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 95% confidence interval. **Bold type indicates significant odds ratios** when the Bonferroni-Holm adjusted p-values are less than 0.05. In event 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)	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)	1.00 (0.86-1.17)	1.33 (0.82-2.14)	1.15 (1.02-1.29)	1.33 (1.04-1.70)
227.3	BN of pituitary gland and craniopharyngeal duct	**1.43 (1.35-1.52)**	**1.50 (1.41-1.6)**	**2.43 (1.90-3.11)**	**1.34 (1.19-1.50)**	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)	**0.82 (0.72-0.93)**	1.05 (0.73-1.52)	1.10 (1.01-1.20)	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)**	**1.17 (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)**	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 (0.90-1.05)	1.32 (0.95-1.85)	1.05 (0.95-1.15)	**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 (1.01-1.20)	**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)**	**1.12 (1.07-1.17)**	1.12 (0.97-1.30)	**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)**	**1.63 (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 (1.00-1.26)	1.12 (0.98-1.28)	1.72 (0.98-3.04)	1.22 (1.03-1.45)	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 (0.90-2.22)	**1.25 (1.11-1.42)**	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)	**0.60 (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)	**0.56 (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)	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)	0.87 (0.78-0.97)	0.84 (0.60-1.18)	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)**	**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 (1.62-1.70)**	**1.69 (1.48-1.92)**	**3.40 (3.32-3.49)**	**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 (1.86-2.40)**	1.55 (0.74-3.27)	**4.39 (3.88-4.98)**	**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)**	**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 (1.59-1.80)**	**1.46 (1.39-1.53)**	**1.53 (1.31-1.78)**	**1.84 (1.78-1.90)**	**2.09 (1.95-2.24)**
188.2	MN of lateral wall of urinary bladder	**1.20 (1.06-1.36)**	1.05 (0.90-1.23)	0.81 (0.30-2.17)	**1.49 (1.27-1.75)**	**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)**	**0.65 (0.61-0.68)**	0.93 (0.79-1.08)	**1.07 (1.03-1.10)**	**1.24 (1.15-1.34)**
196	Secondary and unspecified MN of lymph nodes	**1.08 (1.06-1.10)**	**1.04 (1.02-1.07)**	**1.38 (1.25-1.53)**	**1.17 (1.14-1.21)**	**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)**	**1.34 (1.32-1.37)**	**1.18 (1.10-1.26)**	**1.80 (1.77-1.82)**	**2.27 (2.20-2.33)**
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)**	**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 (1.70-1.80)**	**2.05 (1.80-2.33)**	**2.36 (2.28-2.44)**	**2.36 (2.30-2.43)**	**3.23 (3.01-3.46)**	**2.33 (2.28-2.38)**	**2.59 (2.47-2.71)**

Table 4: Prevalence of outcome (row heading) in each exposure group(column heading). 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)	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	206	17	418	71
227.3	BN of pituitary gland and craniopharyngeal duct	1350	1190	64	334	395	33	881	105
227.1	BN of parathyroid gland	796	670	38	229	271	29	669	104
225.2	BN of cerebral meninges	2741	2166	86	1045	1480	151	2993	391
226	BN of thyroid glands	937	817	32	223	281	32	546	80
153.4	MN of cecum	1093	797	35	475	679	78	1870	291
153.6	MN of ascending colon	1300	952	35	570	851	89	2168	316
153.9	MN of colon, unspecified site	3276	2474	95	1442	2281	179	5357	791
235.2	Neoplasm of uncertain behavior of stomach, intestines, and rectum	749	543	21	349	589	38	1188	157
153.0	MN of hepatic flexure	335	244	12	151	222	24	600	94
153.1	MN of transverse colon	569	392	19	281	424	39	1028	158
155.1	MN of intrahepatic bile ducts	343	272	17	131	162	21	461	76
156.1	MN of extrahepatic bile ducts	217	159	6	99	116	21	351	57
174.5	MN of lower-outer quadrant of female breast	276	235	11	78	100	3	200	35
174.4	MN of upper-outer quadrant of female breast	1301	1101	47	357	414	34	895	151
162.2	MN of main bronchus	995	486	14	692	1862	68	3285	705
162	MN of trachea, bronchus, and lung	14672	7644	238	9502	24688	1258	47971	10694
162.4	MN of middle lobe, bronchus or lung	567	306	7	379	853	54	1771	423
162.8	MN of other parts of bronchus or lung	1749	928	20	1150	3139	153	5809	1260
188.9	MN of bladder, part unspecified	2159	1298	42	1256	2369	168	5700	955
188.2	MN of lateral wall of urinary bladder	291	170	4	170	310	21	824	124
196.2	Secondary and unspecified MN of intra-abdominal lymph nodes	3163	2446	114	1123	1498	168	4556	747
196	Secondary and unspecified MN of lymph nodes	10878	8252	369	4032	6799	549	16439	3147
198	Secondary MN of other specified sites	15895	11348	405	6859	13865	866	33550	6014
197	Secondary MN of respiratory and digestive systems	17325	12577	478	7270	14113	945	32834	5637

238.7	Other lymphatic and hematopoietic tissues	8762	6424	235	4278	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	Exposure / Event	Asthma (493.*)	Asthma, unspecified (493.9)	Extrinsic asthma (493.0)	Chronic obstructive asthma (493.2)	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	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	1.11	1.23	1.36	1.18
227.1	BN of parathyroid gland	1.44	1.41	2.43	1.75	1.25	1.77	1.69	1.92
225.2	BN of cerebral meninges	1.56	1.44	1.73	2.53	2.16	2.91	2.38	2.27
226	BN of thyroid glands	1.54	1.57	1.87	1.56	1.18	1.78	1.26	1.34
153.4	MN of cecum	1.17	0.99	1.33	2.15	1.86	2.82	2.79	3.18
153.6	MN of ascending colon	1.20	1.03	1.15	2.24	2.02	2.79	2.81	2.99
153.9	MN of colon, unspecified site	1.25	1.11	1.29	2.35	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	3.53	3.01	3.89	3.75
153.0	MN of hepatic flexure	1.31	1.12	1.67	2.51	2.23	3.19	3.29	3.77
153.1	MN of transverse colon	1.18	0.95	1.40	2.48	2.25	2.74	2.99	3.35
155.1	MN of intrahepatic bile ducts	1.00	0.92	1.75	1.62	1.21	2.07	1.87	2.26
156.1	MN of extrahepatic bile ducts	0.92	0.79	0.90	1.78	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	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	1.26	1.37	1.49	1.84
162.2	MN of main bronchus	4.64	2.65	2.32	13.73	22.31	10.75	21.50	33.78
162	MN of trachea, bronchus, and lung	3.08	1.87	1.76	8.69	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	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	15.55	9.99	15.72	25.01
188.9	MN of bladder, part unspecified	1.47	1.04	1.02	3.65	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	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	1.20	1.78	1.99	2.39
196	Secondary and unspecified MN of lymph nodes	1.28	1.13	1.54	2.02	2.06	2.20	2.73	3.85
198	Secondary MN of other specified sites	1.25	1.05	1.13	2.33	2.86	2.35	3.82	5.08
197	Secondary MN of respiratory and digestive systems	1.23	1.04	1.20	2.22	2.61	2.30	3.35	4.25
238.7	Other lymphatic and hematopoietic tissues	2.16	1.85	2.05	4.51	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, race was categorized as 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 (35, 36, 64, 65) 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.

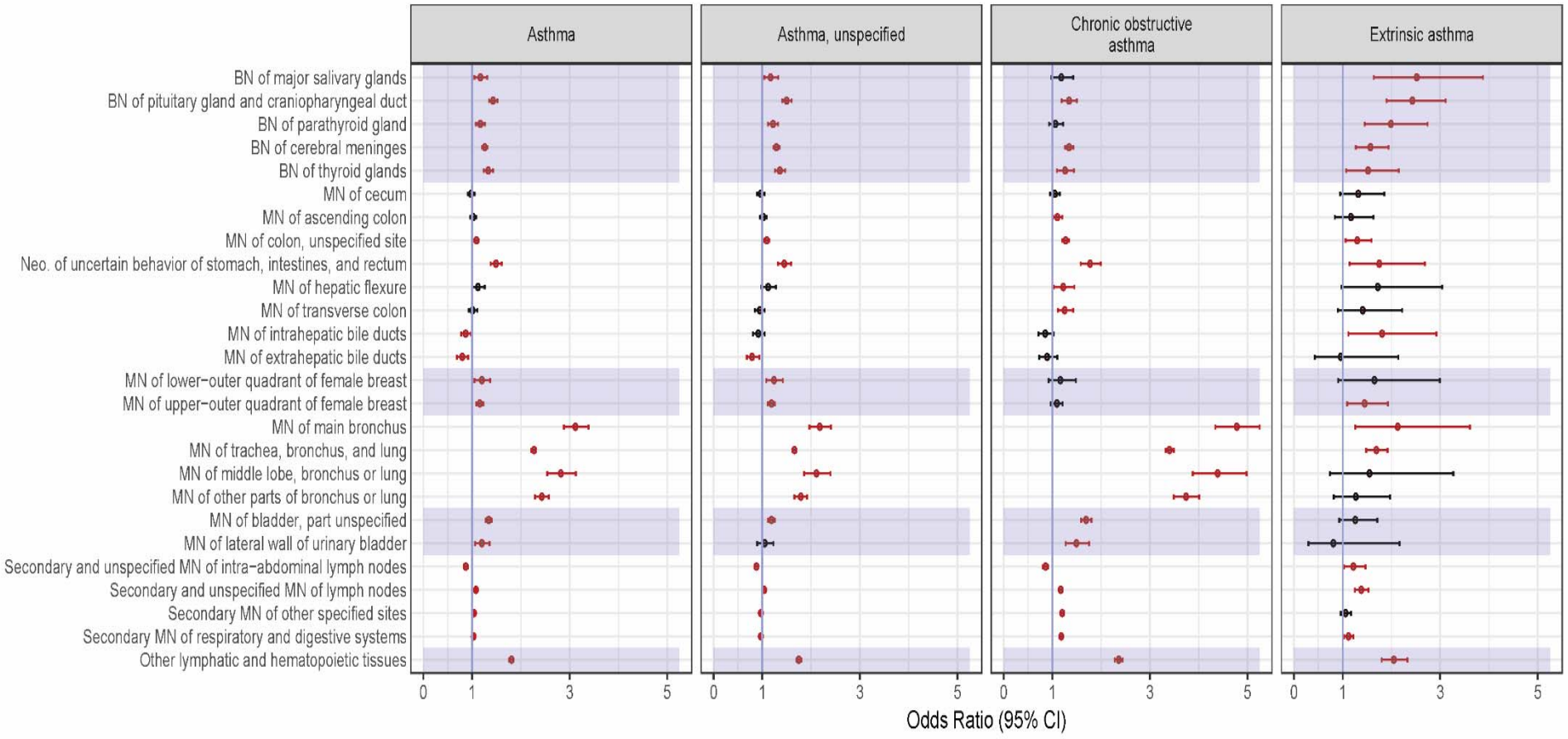
Appendix 4. Advantages of CMH common odds ratio

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) (44, 45) 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:

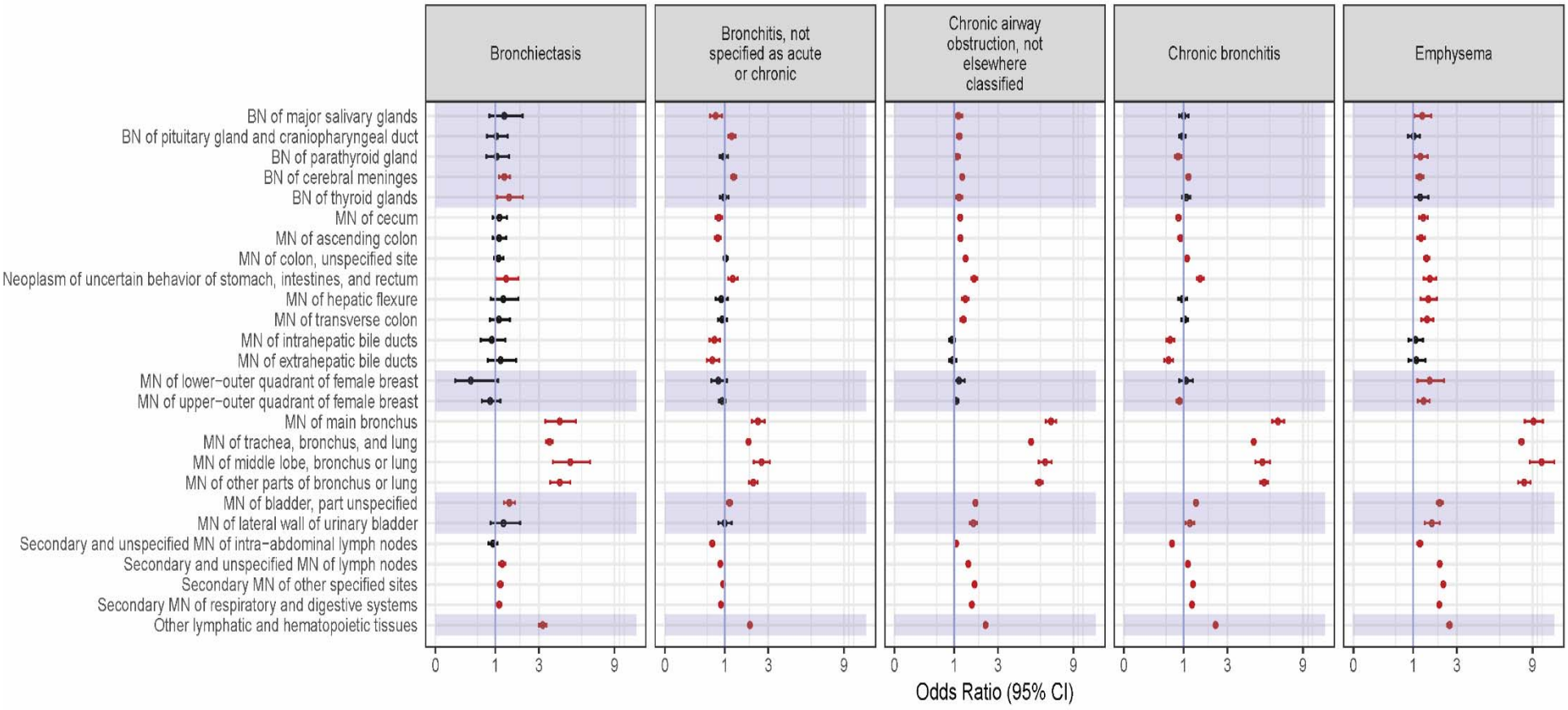
$$\mathrm{OR}_{\{\mathrm{CMH}\}} = \frac{\sum_{\{\mathrm{i}=1\}}^{\mathrm{m}} \frac{a_{11}^{i} a_{22}^{i}}{n_i}}{\sum_{\{\mathrm{i}=1\}}^{\mathrm{m}} \frac{a_{12}^{i} a_{21}^{i}}{n_i}}$$

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 do 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 do 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 CfD, 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 OJ, Moorman JE, 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 OJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. 2012.
4. Wouters EF, Reynaert NL, Dentener MA, Vernooy JH. Systemic and local inflammation in asthma and chronic obstructive pulmonary disease: is there a connection? Proceedings of the American Thoracic Society. 2009;6(8):638-47.
5. Togias A. Systemic effects of local allergic disease. Journal of allergy and clinical immunology. 2004;113(1):S8-S14.
6. Magnussen H, Watz H. Systemic inflammation in chronic obstructive pulmonary disease and asthma: relation with comorbidities. Proceedings of the American Thoracic Society. 2009;6(8):648-51.
7. Boulet L-P, Boulay M-È. Asthma-related comorbidities. Expert review of respiratory medicine. 2011.
8. Cazzola M, Segreti A, Calzetta L, Rogliani P. Comorbidities of asthma: current knowledge and future research needs. Current opinion in pulmonary medicine. 2013;19(1):36-41.
9. Luong Kvq, Nguyen LTH. Hyperthyroidism and asthma. Journal of Asthma. 2000;37(2):125-30.
10. Zacharisen MC, Fink JN, editors. Hyperthyroidism complicating asthma treatment. Allergy and Asthma Proceedings; 2000: OceanSide Publications, Inc.
11. Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002;420(6917):860.
12. Grivennikov SI, Greten FR, Karin M. Immunity, Inflammation, and Cancer. Cell. 2010;140(6):883-99.
13. Schottenfeld D, Beebe-Dimmer J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. CA: a cancer journal for clinicians. 2006;56(2):69-83.
14. Santillan AA, Camargo CA, Colditz GA. A meta-analysis of asthma and risk of lung cancer (United States). Cancer causes and control. 2003;14(4):327-34.
15. Turner MC, Chen Y, Krewski D, Ghadirian P, Thun MJ, Calle EE. Cancer mortality among US men and women with asthma and hay fever. American journal of epidemiology. 2005;162(3):212-21.
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. 2015;54(8):1120-7.
17. Su Y-L, Chou C-L, Rau K-M, Lee CT-C. Asthma and risk of prostate cancer: a population-based case-cohort study in Taiwan. Medicine. 2015;94(36).
18. Linabery AM, Prizment AE, Anderson KE, Cerhan JR, Poynter JN, Ross JA. 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. 2014;23(9):1903-12.
19. Gomez-Rubio P, Zock J-P, Rava M, Marquez M, Sharp L, Hidalgo M, et al. Reduced risk of pancreatic cancer associated with asthma and nasal allergies. Gut. 2017;66(2):314-22.
20. Kallen B, Gunnarskog J, Conradson T. Cancer risk in asthmatic subjects selected from hospital discharge registry. European Respiratory Journal. 1993;6(5):694-7.
21. Talbot-Smith A, Fritschi L, Divitini ML, Mallon DF, Knuiman MW. Allergy, atopy, and cancer: a prospective study of the 1981 Busselton cohort. American journal of epidemiology. 2003;157(7):606-12.
22. VENA JE, BONA JR, BYERS TE, MIDDLETON JR E, SWANSON MK, GRAHAM S. Allergy-related diseases and cancer: an inverse association. American journal of epidemiology. 1985;122(1):66-74.
23. VESTERINEN E, PUKKALA E, TIMONEN T, AROMAA A. Cancer incidence among 78 000 asthmatic patients. International journal of epidemiology. 1993;22(6):976-82.
24. Eriksson N, Holmen A, Högstedt B, Mikoczy Z, Hagmar L. A prospective study of cancer incidence in a cohort examined for allergy. Allergy. 1995;50(9):718-22.
25. Agency for Healthcare Research and Quality. HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP) 2005-2011.
26. Houchens R. Inferences with HCUP State Databases Final Report: U.S. Agency for Healthcare Research and Quality; [updated October 12, 20102018/02/03]. Available from: http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp.

27. Andrews RM. Statewide Hospital Discharge Data: Collection, Use, Limitations, and Improvements. Health Services Research. 2015;50(Suppl 1):1273-99.
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 OP, Barrett ML, Steiner CA. Using the HCUP Databases to Study Incidence and Prevalence: U.S. Agency for Healthcare Research and
Quality; [updated December 6, 20162018/02/03]. Available from: http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp.
31. USER GUIDE: HCUP SUPPLEMENTAL VARIABLES FOR REVISIT ANALYSES: U.S. Agency for Healthcare Research and
Quality; [updated March 16, 20162018/02/03]. Available from: https://www.hcup-us.ahrq.gov/toolssoftware/revisit/UserGuide_SuppRevisitFilesCD.pdf.
32. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC medicine. 2010;8(1):18.
33. Shojaee A, Ranasinghe I, Ani A. Revisiting Causality Inference in Memory-less Transition Networks. eprint arXiv:160802658. 2016.
34. Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. http://dxdoiorg/102105/AJPH2004059204. 2011.
35. Iacus SM, King G, Porro G. CEM: software for coarsened exact matching. Journal of statistical Software. 2009;30(9):1-27.
36. Iacus SM, King G, Porro G, Katz JN. Causal inference without balance checking: Coarsened exact matching. Political analysis. 2012:1-24.
37. Iacus SM, King G, Porro G. CEM: software for coarsened exact matching. Journal of Statistical Software. 2009;30(13):1-27.
38. Hamar GB, Rula EY, Coberley C, Pope JE, 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. 2015;15(1):174.
39. Hicks CW, Hashmi ZG, Velopulos C, Efron DT, Schneider EB, Haut ER, et al. Association between race and age in survival after trauma. JAMA surgery. 2014;149(7):642-7.
40. Nazari SSH, Noroozi M, Soori H, Noroozi A, Mehrabi Y, Hajebi A, et al. 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. 2016;27:127-31.
41. Mason RJ, Moazzez A, Moroney JR, 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. 2012;215(1):88-99.
42. Haider AH, David J-S, Zafar SN, Gueugniaud P-Y, Efron DT, Floccard B, et al. Comparative Effectiveness of In-Hospital Trauma Resuscitation at a French Trauma Center and Matched Patients Treated in the United States. Annals of surgery. 2013;258(1):178.
43. Hametner C, Kellert L, Ringleb PA. Impact of sex in stroke thrombolysis: a coarsened exact matching study. BMC neurology. 2015;15(1):10.
44. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the national cancer institute. 1959;22(4):719-48.
45. Cochran WG. Some methods for strengthening the common χ 2 tests. Biometrics. 1954;10(4):417-51.
46. Abdi H. Holm's sequential Bonferroni procedure. Encyclopedia of research design. 2010;1(8):1-8.
47. Zhang X, Holt JB, Lu H, Wheaton AG, Ford ES, Greenlund KJ, et al. 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. 2014;179(8):1025-33.
48. de-Torres JP, Wilson DO, Sanchez-Salcedo P, Weissfeld JL, Berto J, Campo A, et al. 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. 2015;191(3):285-91.
49. Yao H, Rahman I. Current concepts on the role of inflammation in COPD and lung cancer. Current opinion in pharmacology. 2009;9(4):375-83.

50. Young RP, Hopkins RJ, Christmas T, Black PN, Metcalf P, Gamble G. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. European Respiratory Journal. 2009;34(2):380-6.
51. Gilkes DM, Semenza GL, Wirtz D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. Nature Reviews Cancer. 2014;14(6):430-9.
52. Ma X-J, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast cancer research. 2009;11(1):R7.
53. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS biology. 2004;2(2):e7.
54. Hong JC. Hormones, Part 1: Thyroid and Corticosteroid Hormones. Essentials of Pharmacology for Anesthesia, Pain Medicine, and Critical Care: Springer; 2015. p. 313-25.
55. Nicoloff JT, Fisher DA, Appleman MD. The role of glucocorticoids in the regulation of thyroid function in man. Journal of Clinical Investigation. 1970;49(10):1922-9.
56. HILLS JR SR, REISS RS, FORSHAM PH, GEORGE WT. The effect of adrenocorticotropin and cortisone on thyroid function: thyroidadrenocortical interrelationships. The Journal of Clinical Endocrinology. 1950;10(11):1375-400.
57. Barnes PJ. Therapeutic approaches to asthma–chronic obstructive pulmonary disease overlap syndromes. Journal of Allergy and Clinical Immunology. 2015;136(3):531-45.
58. Pandya D, Puttanna A, Balagopal V. Systemic Effects of Inhaled Corticosteroids: An Overview. The Open Respiratory Medicine Journal. 2014;8:59-65.
59. Van Staa T, Leufkens H, Cooper C. Use of inhaled corticosteroids and risk of fractures. Journal of Bone and Mineral Research. 2001;16(3):581-8.
60. Faul JL, Wilson SR, Chu JW, Canfield J, Kuschner WG. The effect of an inhaled corticosteroid on glucose control in type 2 diabetes. Clinical medicine & research. 2009;7(1-2):14-20.
61. McWhorter WP. Allergy and risk of cancer. A prospective study using NHANESI followup data. Cancer. 1988;62(2):451-5.
62. Merrill RM, Isakson RT, Beck RE. The association between allergies and cancer: what is currently known? Annals of Allergy, Asthma & Immunology. 2007;99(2):102-17.
63. Rittmeyer D, Lorentz A. Relationship between Allergy and Cancer: An Overview. International Archives of Allergy and Immunology. 2012;159(3):216-25.
64. Wells AR, Hamar B, Bradley C, Gandy WM, Harrison PL, Sidney JA, et al. Exploring Robust Methods for Evaluating Treatment and Comparison Groups in Chronic Care Management Programs. Population Health Management. 2013;16(1):35-45.
65. Esareyy J. Causal Inference with Observational Data. 2015.