



Causal Inference Analysis for Poorly Soluble Low Toxicity Particles, Lung Function, and Malignancy

Philip Harber*

Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

OPEN ACCESS

Edited by:

Vittorio Perduca,
Université de Paris, France

Reviewed by:

Conor MacDonald,
INSERM U1018 Centre de Recherche
en Épidémiologie et Santé des
Populations (CESP), France
Vivian Viallon,
International Agency For Research On
Cancer (IARC), France

*Correspondence:

Philip Harber
pharber@arizona.edu

Specialty section:

This article was submitted to
Environmental Health and Exposome,
a section of the journal
Frontiers in Public Health

Received: 27 January 2022

Accepted: 05 May 2022

Published: 05 July 2022

Citation:

Harber P (2022) Causal Inference
Analysis for Poorly Soluble Low
Toxicity Particles, Lung Function, and
Malignancy.
Front. Public Health 10:863402.
doi: 10.3389/fpubh.2022.863402

Poorly soluble low toxicity particles such as carbon black and titanium dioxide have raised concern about possible nonmalignant and malignant pulmonary effects. This paper illustrates application of causal inference analysis to assessing these effects. A framework for analysis is created using directed acyclic graphs to define pathways from exposure to potential lung cancer or chronic airflow obstruction outcomes. Directed acyclic graphs define influences of confounders, backdoor pathways, and analytic models. Potential mechanistic pathways such as intermediate pulmonary inflammation are illustrated. An overview of available data for each of the inter-node links is presented. Individual empirical epidemiologic studies have limited ability to confirm mechanisms of potential causal relationships due to the complexity of causal pathways and the extended time course over which disease may develop. Therefore, an explicit conceptual and graphical framework to facilitate synthesizing data from several studies to consider pulmonary inflammation as a common pathway for both chronic airflow obstruction and lung cancer is suggested. These methods are useful to clarify potential bona fide and artifactual observed relationships. They also delineate variables which should be included in analytic models for single study data and biologically relevant variables unlikely to be available from a single study.

Keywords: causal inference analysis, directed acyclic graph, carbon black, chronic obstructive pulmonary disease (COPD), lung cancer, particulate toxicity, causation analysis, pulmonary inflammation

INTRODUCTION

Poorly soluble low toxicity particles (PSLTs) have received increased interest over the past few years. Concerns about possible malignant or nonmalignant pulmonary disease effects are being addressed. However, available epidemiologic data leave considerable uncertainty about the presence and magnitude of significant effects in humans. In general, studies of chronic pulmonary disorders (e.g., chronic airflow obstruction, CAO) and malignancies such as lung cancer (CA) are particularly challenging because of long latencies and major changes in typical workplace exposures over time. Although more mechanistic data are suggested by animal and *in vitro* studies, the applicability to human health is constrained by species differences and differences in toxokinetics.

Causal Analysis

Causal inference analysis (CIA) provides an explicit framework for identifying and representing interacting factors of different types in complex multi-step causal pathways (1). It is particularly useful for delineating variable sets to be considered as covariates, defining exposure variables more

clearly, and identifying potential biases leading to misleading conclusions. These methods also allow explicit consideration of mediators between exposure and disease outcomes.

Causal considerations have long been implicit in science; research increasingly seeks to provide explanations rather than simple descriptions. A causal relationship between A and B ($A \rightarrow B$) may be conceptualized as either deterministic or probabilistic. In a deterministic relationship, A is sufficient to cause B. It is probabilistic if an intervention setting A to a fixed value changes the probabilistic distribution of B. Current assessments of the likelihood that PSLTs produce CA and/or CAO follow the latter inferential probabilistic paradigm.

Significant conceptual advances concerning causation occurred in the early 20th and early 21st. In the early 1900's, Popper conceptualized that an explanation may be considered scientific if it is possible to refute it by empirical observations. Similarly, Fisher, Pearson, and others helped create the basis for modern statistical hypothesis testing and statistical inference techniques.

Observational studies such as those assessing exposures and long-term health in humans must consider multiple factors. Such factors may be empirically measurable for inclusion as variables in statistical calculation models or be unmeasured but of potential interest. Such factors may increase or decrease the apparent relationship between the predictors and outcomes of interest; they may also prevent identification of bona fide causal relationships or lead to the appearance of artifactual relationships. Confounders in regression models are a such empirically measurable factors worthy of consideration. Even unmeasurable factors should be considered for interpreting the significance of calculated models. Occasionally, a measurable instrumental variable may be used to infer an important unmeasured factor.

Focus upon causation evaluation methodologies reignited in the early 21st century to deal with this potential complexity. Causal inference analysis helps identify the set of variables, whether measurable or unmeasured, that should be considered. Directed acyclic graphs (DAGs) are increasingly used to both facilitate ascertaining estimands and to force logical thinking and explicit representation of how they might relate. This paper illustrates how such DAGs may be applied to understanding several potential PSLTs effects.

Using pulmonary inflammatory response as an example, the paper then illustrates how DAGs may be applied to consideration of underlying biologic mechanisms in a single study. It then extends these principles to suggest how approaches analogous to DAGs may facilitate explicitly expressing potential underlying mechanisms even if information comes from separate studies. While mechanistic reasoning is already extensively used, the graphic approach facilitates explicit and logical representation.

Directed Acyclic Graphs (DAGs)

DAGs are based upon underlying quantitative probabilistic models as described by Greenland et al. in 1999 (2). These methods may be applied for both facilitating specific calculations and for forcing explicit consideration of potential relationships (1).

A directed acyclic graph is composed of a series of nodes and links. The graph is built by including the set of all variables that affect the relationship between the predictor and outcome. Thus, for a complex topic such as PSLT effects, the graph may become complex. In this paper, the graphs are built sequentially beginning with very simple relationships and then adding more considerations. By including a full set of variables, the formal graph approach helps determine those variables that would not affect the primary relationship of interest. The DAG may include a series of alternate pathways.

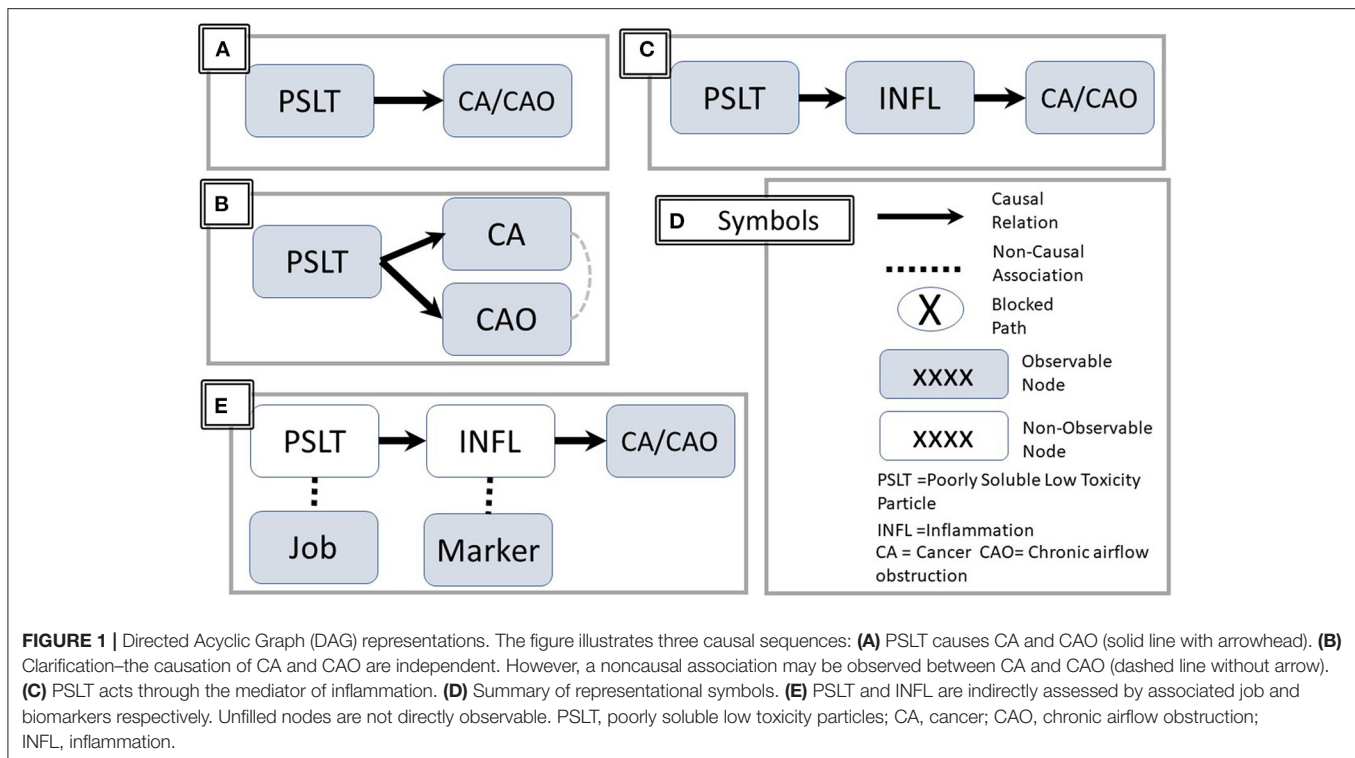
In a DAG, causal relationships must be both directed and unidirectional. The direction of causation is indicated by a single arrow. That is, $A \rightarrow B$ and $B \rightarrow A$ cannot both be included. Causation generally is probabilistic rather than deterministic. In addition to causal links (represented by arrows), associations not implying causation may also be shown (e.g., by dashed lines without arrows).

The DAG is acyclic, meaning that even with a complex series of nodes and links, a variable may not have a directed path leading back to itself. Thus, $A \rightarrow B$ and $A \rightarrow C \rightarrow D \rightarrow B$ may be present but $A \rightarrow C \rightarrow B \rightarrow D \rightarrow A$ is not acceptable.

Paper Overview

This paper applies these methods to examine the specific questions: Do PSLTs such as carbon black (CB) cause chronic airflow obstruction (CAO) or lung cancer (CA), and if so, is this mediated by inflammation? The paper is illustrative and does not presuppose that such relationships exist. The approach is described in three sections: Frameworks (Section Framework for DAG Analysis) describes use of directed acyclic graphs (DAGs) to frame hypotheses and identify variables appropriate for analysis. DAGs consist of a series of nodes and links representing segments of potential causal pathways between the exposure and the possible outcomes. They also represent links to variables, both measured and unmeasured, that should be considered to avoid misleading results. Data for some variables identified as relevant may be challenging to acquire in a single study. Available data (Section Available Data) provides examples of data available for the many of the individual causal pathway link segments, although they may not be directly combined in a specific quantitative model. Implications (Section Implications of the Causal Mechanistic Analysis) suggests approaches for understanding disease development mechanisms by considering information from multiple studies. Potential future studies for CAO and CA are suggested. The approach emphasizes how existing data may be applied rather than providing a comprehensive summary of all relevant research data.

These considerations may contribute to understanding of PSLT effects in several ways: 1) Graphical methods such as DAGs make assumptions about causal pathways explicit. 2) It encourages specifically stating the set of variables deemed necessary and sufficient for understanding causal associations. 3) Limitations of data likely to be available from individual studies are described. 4) Examples illustrate how appropriate or inappropriate variable selection or adjustments may impact accuracy of measures of association. 5) Specific suggestions for qualitative inferences from distinct studies are shown.



FRAMEWORK FOR DAG ANALYSIS

This section describes applications of DAGs for explicitly representing a framework for causal relationships and lays the foundation for assessing the role of empirical data in Section Available Data in specifying the relevant DAG. This section also illustrates methods to reduce the likelihood of misleading results. It also suggests potential intermediate steps along the path from exposure to disease.

Fundamental Hypothesis

Potential pathways from PSLT exposure to final disease outcomes (CAO and CA) are represented as nodes and links. For simplicity, the nodes (PSLT, CAO, and CA) are shown as binary elements (yes, no). However, these methods may be extended to graded effects (e.g., CAO might be represented with a quantified physiologic airflow obstruction), or the strength of a causal link may be quantified. The figures include a series of DAGs showing progressively richer causal networks. Relationships between nodes are represented by a line with an arrow if the relationship is causal, by dashed lines without an arrow if the association is not causal, and by lighter dashed lines if it may be artifactual.

The traditional epidemiologic mortality study hypothesis asking whether PSLT predicts development of CA and CAO is shown in **Figure 1A**. This illustrates a direct causal or “front door path”. This is further clarified in **Figure 1B**, suggesting that PSLT exposure predicts CA and CAO, but they occur independently. In this circumstance, an association of CAO and CA would be observed even if there is no actual causal pathway between these outcomes. In addition, the apparent strength of association

between PSLT and CA (e.g., odds ratio) may be attenuated by adjustment for the presence of CAO.

The mechanistic hypothesis that a causal relationship between PSLT and CAO or CA is mediated via pulmonary inflammation is shown in **Figure 1C**. For simplicity, the two health outcomes are combined in a single node. **Figure 1D** summarizes the common symbols.

However, there are several ambiguities in the apparently simple representation of the hypothesis.

- Even if inflammation is a critical intermediary step, inflammation *per se* may not be the final effector for producing airway damage or malignant cellular transformation (i.e., there may be additional intermediary steps).
- The number of potentially observable measures of inflammation, particularly in humans, is limited and may not truly reflect the *in vivo* inflammatory process.
- Inflammation is not a specific entity; rather, there are numerous inflammatory pathways, many of which are interactive and/or cross-regulatory. These are often assessed indirectly by measuring associated biomarkers (**Figure 1E**).
- “PSLT” is not a specific entity, and its definition is ambiguous. While many would agree that carbon black and titanium dioxide fall within this group, inclusion of other materials such as coal dust is less certain.
- Even for a specific agent, potential effects and pathways may depend upon the characteristics of size, particle size, dose, dose rate, charge, and surface properties. These were recently reviewed by Borm and Driscoll (3).
- The exposure term (e.g., PSLT) may either refer to the specific agent or to a surrogate of exposure such as job title. Laboratory

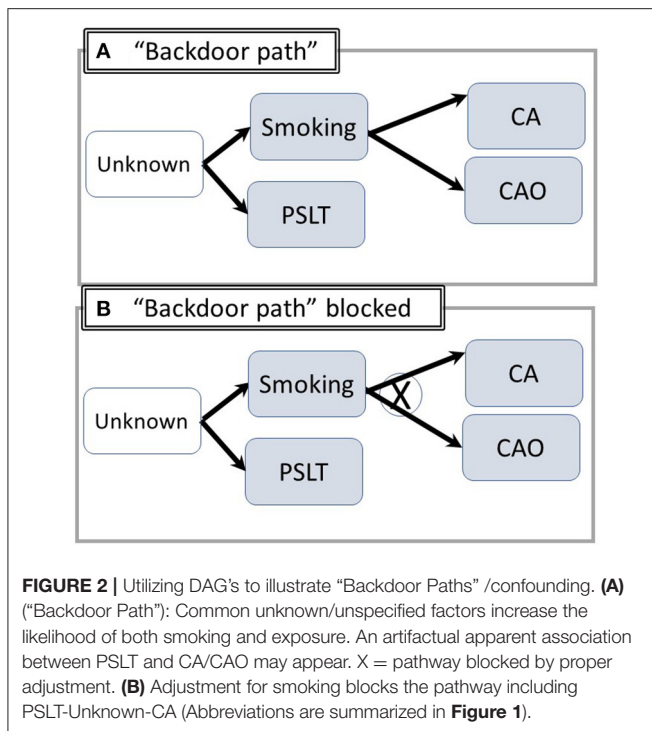


FIGURE 2 | Utilizing DAG's to illustrate "Backdoor Paths" /confounding. **(A)** ("Backdoor Path"): Common unknown/unspecified factors increase the likelihood of both smoking and exposure. An artificial apparent association between PSLT and CA/CAO may appear. X = pathway blocked by proper adjustment. **(B)** Adjustment for smoking blocks the pathway including PSLT-Unknown-CA (Abbreviations are summarized in **Figure 1**).

studies may use a chemically defined agent with specified dose or even assessed initial or retained dose. Neither exposure nor dose are typically precisely measured or controlled in human studies. Rather, these are assessed with proxy measures such as job title, job duration, etc. as approximate surrogate measures of actual dose (**Figure 1E**). The correlation between the surrogate and the actual dose is likely to be heterogeneous over time and among study sites.

- Time course: Inflammatory events early in the time course may initiate events ultimately leading to CAO/CA, but the mechanistic or exposure events later in the pathway may be different. Implications are discussed below.

Explicit consideration of pathways involving both observable and significant but unknown (non-recorded) factors is essential for analysis and interpretation. Backdoor pathways, representing artificial associations such as the well-known confounder of tobacco smoking are illustrated in **Figure 2A**. For example, smoking may be associated with exposure if common but unobserved socioeconomic or cultural factors increase the likelihood of working in a heavily exposed job and of smoking. Members of socially disadvantaged groups may be disproportionately likely to have jobs with heavy exposure and come from homes in polluted areas or have grown up with parental smoking. Adjustment for smoking, which produces both CA and CAO (**Figure 2B**) prevents the appearance of an incorrect causal association between PSLT and the outcomes.

Conversely, adjustment for an intermediary step such as inflammation may reduce the likelihood of identifying a true relationship between PSLT and the outcomes. This potential overadjustment is illustrated in **Figure 3A**. Similar considerations apply if there are both a direct and an

inflammation mediated pathway (**Figure 3B**) (The figure is simplified by assuming each of the nodes is directly observable). In contrast, adjustment for inflammation may create an artificial association between PSLT and CA if both PSLT and smoking cause inflammation, but inflammation is not actually the mediator between PSLT and CA (see **Figure 3C**). In this "collider" situation, a false association between PSLT and CA would be observed if the PSLT-CA measure of association (e.g., odds ratio of CA on PSLT) is adjusted for inflammation (Adjustment for smoking would eliminate this effect, but it is preferable to simply not adjust for inflammation). Thus, assessment of which factors to use and when to adjust is facilitated by DAGs.

Constraints Upon Observable Data in Human Studies

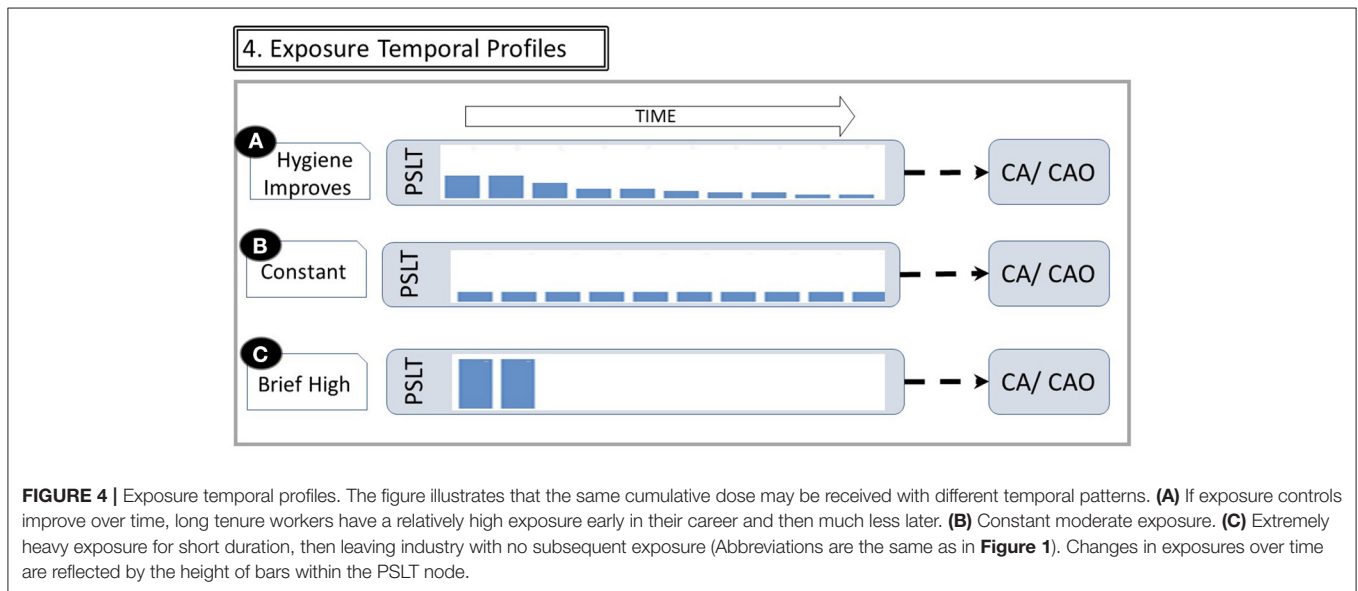
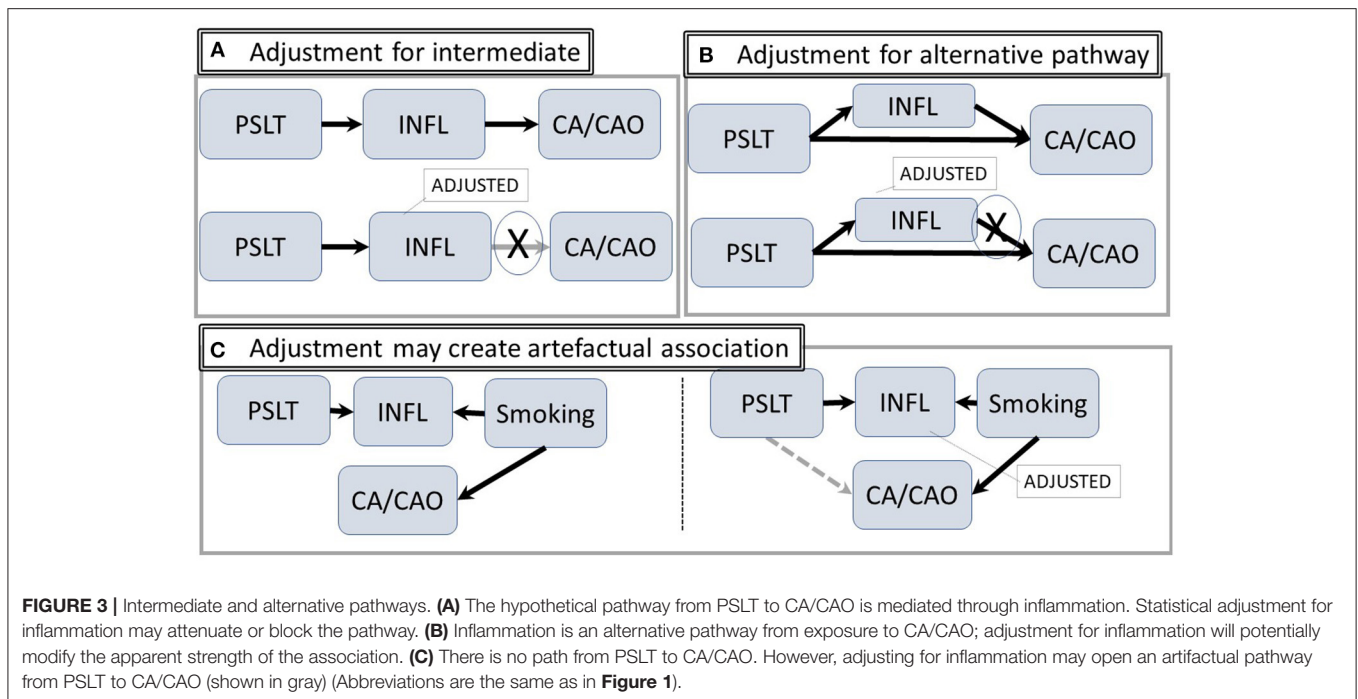
In addition to the consequences of long latency and occasional ambiguity in defining "exposure", observational studies in humans require additional framework considerations. Such ambiguities may contribute to inconsistency among studies. These include distinguishing early vs. late steps in the causal chain and practical constraints upon measures of inflammation that may be collected noninvasively.

The importance of time is summarized in **Figure 4**. A study of whether PSLT predicts subsequent CAO or CA must cover a long time span. Even if the worker remains active in the industry until near the date at which diagnosis is made, it is necessary to reconstruct exposures many years in the past. This is particularly important since workplace air concentrations were often orders of magnitude greater in the past. In many instances, retrospective exposure assessment does not occur until near the date of diagnosis and is therefore subject to recall bias and/or missing data. Other workers may have left the industry long before recognition of CAO or CA, making cohort identification challenging.

Many studies utilize estimated cumulative exposure as the predictor of a possible health effect. However, the same cumulative exposure may be accrued with very different dose rates and latencies as shown in **Figure 4**. For example, some may have moderate exposure over many years, while others may have an early brief but very intense exposure. Misinterpretation of causal relationships may result from not considering the time varying nature of longitudinal exposures (4).

Both biologic and operational consequences of the time course are illustrated in **Figure 5**, in which time is divided into early, mid, and late eras. The latter begins when the fully advanced disease is present.

Biologic considerations about PSLT's warrant consideration of the assumption that cumulative exposure is the optimal predictor. This is particularly true for PSLT's for which clearance and deposition may have thresholds, overloading, or significant nonlinearities. For example, clearance processes may be overwhelmed with very high early exposures but not with more moderate exposures. Limited epidemiologic adjustment methods such as overweighting or underweighting exposure by calendar date or excluding either recent or very remote exposures



does not fully address biologic considerations such as short-term or long-term clearance, tissue repair, threshold doses, or other mechanistic considerations. Underestimation of remote exposures due to applying more recent data to earlier years (5) may lead to overestimation of the regression coefficient of CAO upon exposure.

The simple DAG of **Figure 1** is augmented for analytic clarity in **Figure 5**. This framework expresses several possible considerations. Examples of several human inflammatory markers are shown in **Table 1**. **Figure 5** represents that inflammation is better represented as a cascade of steps rather than as a single node, and there are multiple inflammatory

pathways of which only some are likely to be relevant. Unlike animal and *in vitro* studies, the range of inflammatory biomarkers appropriate for human studies must be limited to noninvasive or minimally invasive procedures. Hence, the observable biomarker for each inflammatory step may be correlated with but not synonymous with the actual mediator (dashed line in the figure); the available biomarkers may therefore be indirect measures rather than direct assessment of the inflammatory steps. They are still useful, but they may only be loosely correlated to the actual inflammatory element.

The figure also suggests that markers measured at various times may reflect various mechanistic aspects differing in their

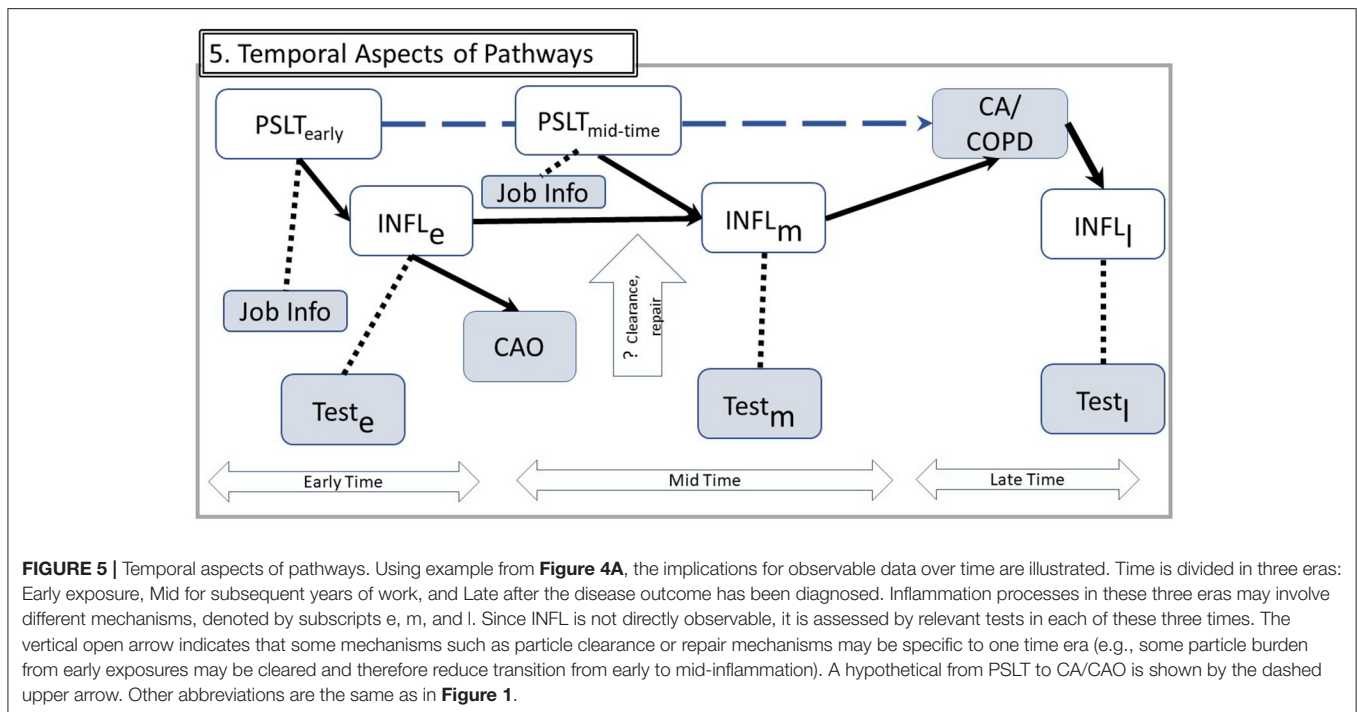


TABLE 1 | Markers of pulmonary inflammation (examples).

Invasive:

Animal sacrifice and histology/biomarker measurement
Bronchial biopsy, bronchial lavage

Less invasive human sampling:

Exhaled breath analysis (frozen or otherwise)

Minimally invasive measures:

Exhaled indicators: FeNO: (Exhaled nitric oxide)
(Subject to upper airway contamination)
(Short-term responsive)
(Does this reflect “important” mediator?)

Non-pulmonary biomarkers:

Nonspecific indicators e.g., CRP- C-reactive protein
CC16 – Club Cell 16: Mediator or a protector?
Micro-RNA

Omics:

Genomics
Epigenomics
Proteomics
Metabolomics
Targeted SNP’s or hypothesis generating GWAS

Markers of Effect

Lung function
Quantitative CT scan

significance. Measurement of an early step may be more relevant for some outcomes in comparison to others. Measures of inflammation are more frequently assessed proximate to the health effect diagnosis (“mid”) than in the early years of exposure. Measurement of inflammatory markers in the several years

immediately preceding diagnosis of CA or CAO may overlook the importance of inflammation in the earlier years in initiating the pathogenic sequence.

Reverse causation is another potential concern. For example, rather than PSLT leading to CAO and CA, it is possible that CAO increases the retained dose of PSLT if airway dysfunction reduces clearance of inhaled PSLT and therefore increases the effective retained dose of PSLT. It also is plausible that CA causes inflammation rather than the reverse causal relationship (in the “late” era). This is illustrated by several of the empirical studies discussed below.

Solution to Complexity

The complexity of a causal inference framework, as illustrated by the DAGs, does not imply it is impossibly challenging to understand whether there are bona fide causal relationships between PSLT and the health outcomes.

Much of the complexity is addressable by collecting complete and accurate data on as many variables as feasible in a specific study. This may require long-term studies of a worker group or acquiring data previously obtained (an example is provided in Section Implications of the Causal Mechanistic Analysis). Appropriate adjustments to block backdoor pathways, avoiding adjustments that may introduce artifactual biases, or use of analytical models incorporating do-calculus may help (6). Careful analysis may also identify variables that need not be studied, thereby improving study efficiency (7). In special well-circumscribed circumstances, estimation of effect in one population may be aided by incorporating data from other studies; Pearl applied terms such as transportability and meta-synthesis to these approaches (6, 7). While this paper emphasizes specifying

the DAGs using expert knowledge derived from existing studies, causal structures may sometimes be derived from data themselves with causal distributional notation beyond associational measures (7).

Similarly, information necessary to thoroughly assess the postulated intermediary steps from a single work history—health outcome study. Rather, explicitly delineating the individual node and link segments permits considering available data for each of these segments. For some, analogous human data are available, and for others relevant animal data are applicable. Thus, explicit framework analysis promises to overcome inherent constraints of using a single study and integrating empirical data with conceptual understanding. This approach is illustrated in Part 2 below.

AVAILABLE DATA

The complex causal inference links and nodes shown in **Figure 5** make it unlikely that a single set of human data can traverse all the nodes in the causal pathway. However, data applicable to the each of the links in aggregate may guide consideration of mechanistic hypotheses. The analytic framework described in this paper illustrates how information of several types may be synthesized. It does not provide a specific method for calculating a precise quantitative measure of association (2).

The section includes examples of data types relevant to each of the segments linking particles to inflammation and inflammation to disease. Categories of supportive information include biologic plausibility, disease associations, common exposure associations, common genetic risk factors (e.g., proinflammatory genes), early/pre-lung cancer studies, common mediators, prospective population-based studies, consistency with mediators measured in animal toxicology studies, and internal consistency of postulated mechanisms.

Overall Associations

Several epidemiologic studies have examined the hypotheses that PSLT produces CAO and that PSLT produces CA with studies of carbon black (CB). The CB-CAO relationship was addressed by studies of Harber in North America (8, 9) and Gardiner in Europe and the United Kingdom (5) examining the relationship between estimated cumulative exposure and respiratory outcomes. Both observed that chronic exposure led to a reduction of the forced expiratory volume in one second (FEV1), with the average slope of 0.7 ml and 1.2 m/ per cumulative mg-year/m³ inhalable dust respectively. And effect upon the forced vital capacity was observed in the North American but not the UK/European study. These were cross-sectional studies in which cumulative exposure was retrospectively estimated for the total duration of work for each subject; North American subjects had an average of 14.1 years of exposure (8, 9).

These studies illustrate the practical impact of retrospective exposure assessments. Cumulative exposure estimates were considerably higher in the Harber than the Gardiner studies since the former used all available data to assign exposures in the earlier calendar periods, whereas the latter applied measurements from the early 1990s to earlier years (5, 8). This methodologic

difference is likely to account for differences in regression coefficients for FEV1. Neither considers the temporal pattern or dose rate.

Several epidemiologic mortality studies have addressed the relationship between CB and CA (10–13). In addition, several comprehensive reviews consider this possible association and pointed out the inconclusive nature of the studies (14).

Indicators of Inflammation

As discussed in Section Framework for DAG Analysis, “inflammation” is itself complex since it changes over time and includes many alternate pathways. For human data to be germane to the mechanistic hypothesis under consideration, the measures must be minimally or noninvasive, capable of repeated measurement over time, and reasonably anticipated to correlate well with the actual biologic processes. Potential markers are shown in **Table 1** and their timing is described in **Figure 5**.

Associations Between Diseases

Cross-sectional studies show that several established lung diseases with persistent inflammation are associated with an increased cancer incidence (e.g., tuberculosis, idiopathic pulmonary fibrosis, and CAO) (15–17). These relationships might arise either because inflammation is an underlying cause of both CA and the nonmalignant disease or is a consequence rather than cause of the nonmalignant disease and/or CA. A recent review summarized data supporting a common antecedent for both CAO and CA (18). Many such studies suffer from only looking at “late stage” findings and depend upon fully established clinical diagnoses.

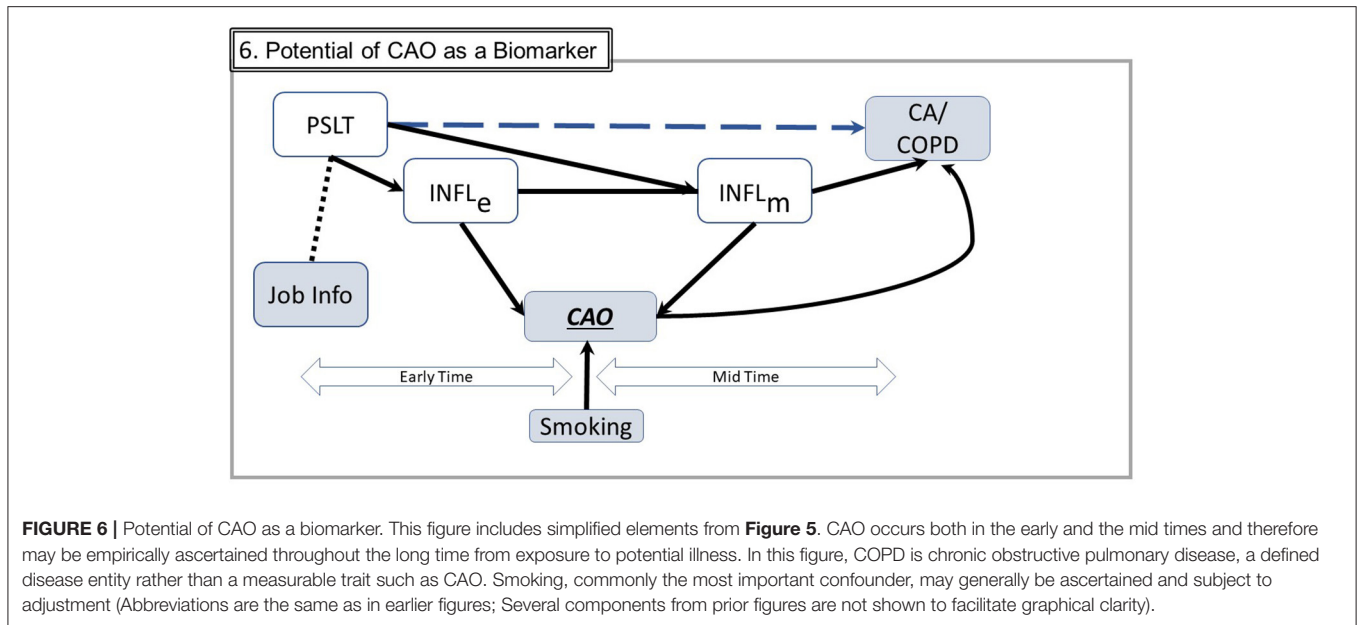
Longitudinal studies overcome this limitation. The US National Health and Nutrition Examination Survey (NHANES) identified 113 incident lung cancers during the 17 year follow-up (19). Moderate or severe CAO at baseline was associated with a significantly elevated risk of subsequent incident CA (OR = 2.6, CI = 1.5–3.8). Analogous data were reported by O’Callaghan (20). Thus, despite limitations, disease association studies provide some support for the proposed mechanism.

Inflammation Concurrent With CA or CAO

Studies have also demonstrated that inflammation is frequently present when CAO or CA is diagnosed. Limited weight should be given to studies showing of inflammation at the time of CA diagnosis such as, the adverse prognostic impact of inflammation at the time of diagnosis (21, 22). These may be important for guiding treatment but have limited import for describing the significant pathway (see **Figure 5**).

Associations Between Particle Exposure and Inflammation

Human and animal data support a role of pulmonary inflammation. These studies include both population-based air pollution studies with long-term exposures as well as shorter-term acute exposures (e.g., air pollution episodes, wildland fires), and laboratory studies in chambers. Animal studies have been particularly fruitful in showing that with adequate dose, PSLT’s can lead to inflammation. This is likely to depend upon the dose



time profile and may proceed through several distinct pathways. Human studies are more limited. For example, a study among CB workers and nonexposed controls found increased eosinophils with CB exposure. The differences in peripheral inflammatory markers between CB exposed and unexposed workers were more evident when analyses were stratified by smoking status (23, 24).

Associations Between Inflammation and CA/CAO

Several prospective studies support causal associations between inflammation and CA and CAO. For example, in an 8 year follow-up, elevated C-reactive protein (CRP) was associated with an elevated risk of subsequent lung cancer (hazard ratio = 3.39) (25). Since the increased CRP antedated the CA, reverse causation (CA → inflammation) is unlikely. An analogous study using high-sensitivity CRP was reported by Muller (26). Association of CRP with lung functional morbidity such as airway hyperresponsiveness or hospitalization has also been reported (27, 28). Genetic studies support this thesis (29, 30).

IMPLICATIONS OF THE CAUSAL MECHANISTIC ANALYSIS

The two preceding sections establish an analytic framework and provide suggestive data applicable to each segment of the pathway. This part suggests their practical implications.

First, causal inferential mechanistic analysis is useful for identifying relevant factors to be considered and suggesting considerations for appropriate and inappropriate adjustments in specific statistical models. Additionally, analogous principles extend to synthesizing data across studies for inferential purposes and informing underlying mechanistic considerations.

Second, translating observed data into practical policy may benefit from mechanistic considerations. Dose and dose rate are particularly germane if the mechanism suggests that clearance may either reduce the effective dose or become saturated at high dose rates. Reliance upon estimated total cumulative dose is not robust to such considerations. In addition, it is useful to consider whether intermediary steps are potentially irreversible or may be counterbalanced by repair mechanisms. Irreversible mutational changes may be contrasted with inflammatory sequences subject to internal homeostatic controls. The analysis presented in this paper does not in itself address these questions specifically for the PSLT's, but it establishes a framework through which other data such as animal toxicology studies may be integrated with observable human data.

Third, these approaches facilitate planning future research studies to strengthen the empirical database. As shown in Section Framework for DAG Analysis, it is unlikely to be practical to measure inflammatory markers throughout a working lifetime and thereafter to link early inflammatory responses to subsequent CAO or CA identified decades later. However, the individual segments (node-link-node) may be studied, albeit in different individuals. For example, the relationship between PSLT and inflammation may be assessed in exposed workers with different temporal dose patterns (see **Figure 3**). Such results may contribute to assessing the significance of current lower-level industrial worker exposures and/or effects upon end-users. Similarly, resolution of inflammatory markers after cessation of exposure may be assessed.

Fourth, the different time courses for detecting CAO and CA might be leveraged to gain insight. Unlike CA, which is typically diagnosed at a specific "late" point in time, CAO develops gradually and is detectable early in its course well before advanced disease such as chronic obstructive pulmonary disease is present (see **Figure 6**). If inflammation is central to the development of

both outcomes, monitoring CAO may provide insight into the temporal/concentration exposure characteristics associated with inflammatory responses. CAO is easily measured repetitively and noninvasively at low cost with spirometry or associated inflammation biomarkers such as exhaled nitric oxide. Assessing temporal course of CAO with different temporal exposure patterns would provide insight relevant to the significance of dose-time exposure patterns. Ongoing industry medical surveillance programs may already have data permitting such analyses. This approach, while currently hypothetical, may be practically implementable.

SUMMARY

Causal inference analysis such as use of DAGs is a very useful tool to help clarify many of the significant questions concerning the health significance of poorly soluble low toxicity particles. It

is useful both for representing the set of appropriate variables and guiding analytic models for individual studies. Its principles of identifying relevant variables and their potential effects may be applied as a heuristic to graphically represent relationships explicitly to foster qualitative synthesis of information from disparate studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Lederer DJ, Bell SC, Branson RD, Chalmers JD, Marshall R, Maslove DM, et al. Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care. *J Ann Am Thorac Soc.* (2019) 16:22–8. doi: 10.1513/AnnalsATS.162erratum
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* (1999) 10:37–48. doi: 10.1097/00001648-199901000-00008
- Borm PJA, Driscoll KE. The hazards and risks of inhaled poorly soluble particles - where do we stand after 30 years of research? *Part Fibre Toxicol.* (2019) 16:11. doi: 10.1186/s12989-019-0294-4
- Étiévant L, Viallon V. Causal Inference under over-Simplified Longitudinal Causal Models. *Int J Biostat.* (2021). doi: 10.1515/ijb-2020-0081
- Gardiner K, van Tongeren M, Harrington M. Respiratory health effects from exposure to carbon black: results of the phase 2 and 3 cross sectional studies in the European carbon black manufacturing industry. *Occup Environ Med.* (2001) 58:496–503. doi: 10.1136/oem.58.8.496
- Pearl J. The do-calculus revisited. *arXiv.* (2012). doi: 10.48550/arXiv.1210.4852
- Pearl J. Causal Inference. In: Isabelle G, Dominik J, Bernhard S, editors. *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008. Proceedings of Machine Learning Research: PMLR.* Brookline, MA: Microtome Publishing (2010). p. 39–58.
- Harber P, Muranko H, Shvartsblat S, Solis S, Torossian A, Oren T, et al. Triangulation approach to historical exposure assessment for the carbon black industry. *J Occup Environ Med.* (2003) 45:131–43. doi: 10.1097/01.jom.0000052956.59271.bd
- Harber P, Muranko H, Solis S, Torossian A, Merz B. Effect of carbon black exposure on respiratory function and symptoms. *J Occup Environ Med.* (2003) 45:144–55. doi: 10.1097/01.jom.0000052955.59271.66
- Dell LD, Gallagher AE, Crawford L, Jones RM, Mundt KA. Cohort study of carbon black exposure and risk of malignant and nonmalignant respiratory disease mortality in the US carbon black industry. *J Occup Environ Med.* (2015) 57:984–97. doi: 10.1097/JOM.0000000000000511
- Dell LD, Mundt KA, Luippold RS, Nunes AP, Cohen L, Burch MT, et al. A cohort mortality study of employees in the US carbon black industry. *J Occup Environ Med.* (2006) 48:1219–29. doi: 10.1097/01.jom.0000218701.62658.a2
- Sorahan T, Harrington JM. A “Lugged” analysis of lung cancer risks in UK carbon black production workers, 1951–2004. *Am J Ind Med.* (2007) 50:555–64. doi: 10.1002/ajim.20481
- Wellmann J, Weiland SK, Neiteler G, Klein G, Straif K. Cancer mortality in German carbon black workers 1976–98. *Occup Environ Med.* (2006) 63:513–21. doi: 10.1136/oem.2006.026526
- Yong M, Anderle L, Levy L, McCunney RJ. Carbon black and lung cancer mortality—a meta-regression analysis based on three occupational cohort studies. *J Occup Environ Med.* (2019) 61:949–0. doi: 10.1097/JOM.0000000000001713
- Engels EA, Shen M, Chapman RS, Pfeiffer RM, Yu YY, He X, et al. Tuberculosis and subsequent risk of lung cancer in Xuanwei, China. *Int J Cancer.* (2009) 124:1183–7. doi: 10.1002/ijc.24042
- Shiels MS, Albanes D, Virtamo J, Engels EA. Increased risk of lung cancer in men with tuberculosis in the alpha-tocopherol, beta-carotene cancer prevention study. *Cancer Epidemiol Biomarkers Prev.* (2011) 20:672–8. doi: 10.1158/1055-9965.EPI-10-1166
- Yoon JH, Nouraei M, Chen X, Zou RH, Sellares J, Veraldi KL, et al. Characteristics of lung cancer among patients with idiopathic pulmonary fibrosis and interstitial lung disease - analysis of institutional and population data. *Respir Res.* (2018) 19:195. doi: 10.1186/s12931-018-0899-4
- Parris BA, O’Farrell HE, Fong KM, Yang IA. Chronic obstructive pulmonary disease (COPD) and lung cancer: common pathways for pathogenesis. *J Thorac Dis.* (2019) 11:S2155–S72. doi: 10.21037/jtd.2019.10.54
- Mannino DM, Aguayo SM, Petty TL, Redd SC. Low lung function and incident lung cancer in the United States: data from the first national health and nutrition examination survey follow-up. *Arch Intern Med.* (2003) 163:1475–80. doi: 10.1001/archinte.163.12.1475
- O’Callaghan DS, O’Donnell D, O’Connell F, O’Byrne KJ. The role of inflammation in the pathogenesis of non-small cell lung cancer. *J Thorac Oncol.* (2010) 5:2024–36. doi: 10.1097/JTO.0b013e3181f387e4
- Cho WC, Kwan CK, Yau S, So PP, Poon PC, Au JS. The role of inflammation in the pathogenesis of lung cancer. *Expert Opin Ther Targets.* (2011) 15:1127–37. doi: 10.1517/14728222.2011.599801
- Yao H, Rahman I. Current concepts on the role of inflammation in COPD and lung cancer. *Curr Opin Pharmacol.* (2009) 9:375–83. doi: 10.1016/j.coph.2009.06.009
- Dai Y, Niu Y, Duan H, Bassig BA, Ye M, Zhang X, et al. Effects of occupational exposure to carbon black on peripheral white blood cell counts and lymphocyte subsets. *Environ Mol Mutagen.* (2016) 57:615–22. doi: 10.1002/em.22036
- Ramage L, Proudfoot L, Guy K. Expression of C-reactive protein in human lung epithelial cells and upregulation by cytokines and carbon particles. *Inhal Toxicol.* (2004) 16:607–13. doi: 10.1080/08958370490464599
- Van’t Klooster CC, Ridker PM, Hjortnaes J, van der Graaf Y, Asselbergs FW, Westerink J, et al. The relation between systemic inflammation and incident cancer in patients with stable cardiovascular disease: a cohort study. *Eur Heart J.* (2019) 40:3901–9. doi: 10.1093/eurheartj/ehz748.0088
- Muller DC, Larose TL, Hodge A, Guida F, Langhammer A, Grankvist K, et al. Circulating high sensitivity C-reactive protein concentrations and risk of lung

- cancer: nested case-control study within lung cancer cohort consortium. *BMJ*. (2019) 364:k4981. doi: 10.1136/bmj.k4981
27. Kony S, Zureik M, Driss F, Neukirch C, Leynaert B, Neukirch F. Association of bronchial hyperresponsiveness and lung function with C-reactive protein (Crp): a population based study. *Thorax*. (2004) 59:892–6. doi: 10.1136/thx.2003.015768
 28. Dahl M, Tybjaerg-Hansen A, Vestbo J, Lange P, Nordestgaard BG. Elevated plasma fibrinogen associated with reduced pulmonary function and increased risk of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. (2001) 164:1008–11. doi: 10.1164/ajrccm.164.6.2010067
 29. Shiels MS, Engels EA, Shi J, Landi MT, Albanes D, Chatterjee N, et al. Genetic variation in innate immunity and inflammation pathways associated with lung cancer risk. *Cancer*. (2012) 118:5630–6. doi: 10.1002/cncr.27605
 30. Watzka D, Lusk CM, Dyson G, Purrington KS, Wenzlaff AS, Neslund-Dudas C, et al. Copd-dependent effects of genetic variation in key inflammation pathway genes on lung cancer risk. *Int J Cancer*. (2020) 147:747–56. doi: 10.1002/ijc.32780

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Harber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.