Evaluation of the probability of causation approach for lung cancer: Scoping review protocol

Javier Mancilla Galindo1, Martijn A. H. Oude Voshaar2, Lützen Portengen1, David van Klaveren2, Roel Vermeulen1, Susan Peters1

1. Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands
2. Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands

Table of contents

[Scoping Review Protocol Summary 2](#_Toc205194972)

[Background 3](#_Toc205194973)

[Research Questions 4](#_Toc205194974)

[Inclusion criteria 6](#_Toc205194975)

[Types of sources 6](#_Toc205194976)

[Methods 8](#_Toc205194977)

[References 16](#_Toc205194978)

[Appendix 1: Pilot Search Strategy (Lung Cancer) 18](#_Toc205194979)

[Appendix 2: Pilot Search Strategy (All types of cancer) 27](#_Toc205194980)

[Appendix 3: Screening Procedures 32](#_Toc205194981)

[Appendix 4. Extraction Instructions 34](#_Toc205194982)

# Scoping Review Protocol Summary

#### Type of Review

**Scoping Review**

#### Review Stages

1. Protocol
2. Search
3. Pilot Screening (50 to 100 records)
4. Pilot Extraction (10 items)
5. Preregistration
6. Search update
7. Screening
8. Extraction
9. Synthesis
10. Report

#### Current Review Stage

1. Screening

#### Start Date

* 28-10-2024 (actual start date)

#### End Date

* 31-10-2025 (estimated end date)

#### Disclosure

This protocol has been written and is kept updated by JMG. Coauthors have provided feedback on the review objectives and plan and are listed here to recognize their intellectual contributions. However, this does not mean that all coauthors approved all version updates of the protocol.

# Background

The determination of the likelihood that an event was caused by a single exposure is of scientific, legal, philosophical, and of practical interest. However, health outcomes are often multi-causal and it is practically difficult and often not possible to establish causality due to a single exposure, although examples for which assessment is relatively straight-forward exist (i.e., asbestos exposure and mesothelioma).[1](#ref-Coenen2024) In occupational health, the determination of individual causality due to exposures at the workplace has important consequences for the financial compensation of workers, but this is limited by the ability to establish a certain diagnosis, the strength of the knowledge base for the disease, among other factors reviewed by Moon and Yoo.[2](#ref-MoonYoo2021)

The probability of causation (PoC) concept has been used in the the past to approximate what share of a health outcome is attributable to a single exposure. For a health outcome to be “more likely than not” caused by the exposure under investigation, the probability that it is responsible for the outcome should be greater than 0.5. The way of obtaining the PoC can be by scoring more than half of the criteria for occupational dermatitis,[3](#ref-Mathias1989) through the attributable fraction (AF) when there is a doubling in the relative risk (RR) for the outcome in the exposed compared to the non-exposed (RR = 2),[4](#ref-GreenlandPoC1999) or by a quantitative exposure threshold above which the RR is greater than 0.5 obtained from the exposure-response curve in single epidemiological studies and meta-analysis of published studies.[5](#ref-Siemiatycki2014)

In the United Kingdom, partial or full financial compensation due to radiation exposure was informed by using PoC thresholds to the benefit of the claimant, by lowering the compensation threshold to values lower than 0.5 to account for uncertainties such as error (systematic and random) and unknown generalizability of epidemiological estimates in different populations.[6](#ref-Wakeford1998) In the United States[7](#ref-NCICDCreport2003) and Canada[8](#ref-Armstrong1988), upper confidence intervals have been used to establish a PoC = 0.5 threshold to compensate workers exposed to radiation.

Despite its use in worker compensation schemes, there have been criticisms to the PoC, such as it not accounting for accelerated outcome occurrence (i.e., the disease would have occurred at a later time without the exposure) or the modifications in PoC due to varying background risk.[4](#ref-GreenlandPoC1999),[9](#ref-Greenland2015) Thus, further examination of the use of the PoC in the scientific literature, alternatives, and applications in health policy is relevant. A preliminary search of PubMed, Embase, and OpenAlex was conducted and no current or underway systematic or scoping reviews of the application of the PoC for the financial compensation of lung cancer cases were found, except for one review which does not fully report the methods to allow to judge if a systematic approach was used.[10](#ref-Kim2018)

# Research Questions

### Primary Research Question:

How has the Probability of Causation (PoC) calculation been applied to evaluate the contribution of occupational exposures to lung cancer in worker populations and what alternatives are there to the PoC?

### Secondary Research Questions:

1. **Study Design:**  
   What are the study designs employed in lung cancer research that utilize the PoC?
2. **Outcomes:**  
   What types of outcomes and assessment methods are used for the PoC in lung cancer studies?
3. **Exposures:**  
   Which occupational exposures have been examined in lung cancer studies applying the PoC?
4. **Population:**  
   In which worker populations has the PoC been applied to study lung cancer?
5. **Causal Assumptions:**  
   What causal assumptions underlie the models used to apply the PoC in lung cancer research?
6. **Uncertainties:**  
   What sources of uncertainty are described when applying the PoC?
7. **Alternatives:**  
   What methodological improvements could be used to estimate the PoC or what alternatives can be used to assess the contribution of occupational exposures to lung cancer in worker compensation schemes?

# Inclusion criteria

Written according to the (Participants, Concept, Context) PCC framework.

#### Population

Adults of both sexes at risk of developing lung cancer due to an occupational exposure.

#### Concept

The development, application, or evaluation of the Probability of Causation (PoC) using an exposure-response function to estimate the probability that a health outcome was caused by a chemical, physical, or biological workplace exposure.

#### Context

Lung cancer and histological subtypes.

# Types of sources

This scoping review will include studies that report the probability of causation (PoC) of lung cancer due to at least one occupational exposure, with the condition that values are calculated using a continuous exposure-response curve to determine the probability that the outcome was due to the workplace exposure. Experimental (randomized controlled trials and non-randomized studies) and observational studies including prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies, and ecological studies will be considered. Systematic reviews using meta-regression to determine the exposure-response curve and calculating the PoC will also be included. Case series applying previously derived PoC functions will be included. Other types of studies, such as case reports, narrative reviews, editorials, commentaries, conference abstracts, among others will be excluded from data charting. Both peer-reviewed studies and pre-prints will be considered for inclusion. In the case of the inclusion of a pre-print, a clear mention of the non-peer reviewed status will be made.

For the alternatives of the PoC, studies identified in the first stage of the review will undergo backward and forward citation snowball review to identify potential alternatives. Additional methodological improvements and alternatives to the PoC considered relevant by the research team will be included in the review.

Both the ascendancy approach and descendancy approach will be employed to find other relevant references. The descendancy algorithm will be implemented in OpenAlex through the openalexr package.

During the pilot phase of this scoping review, other types of articles were considered (case reports, narrative reviews, editorials, commentaries), but none were labelled as potentially relevant. Therefore, these will be excluded as they significantly delayed the progress of the review. Lastly, the research question has been narrowed to studies with lung cancer as the outcome because the searches retrieved a manageable number of records, compared to a broader search strategy including all types of cancer. These points are documented in more detailed in **appendixes 1 and 2**.

# Methods

## Search strategy

### Databases

The search strategy aims to find published studies in two academic databases with large coverage of the medical literature (PubMed and Embase), plus one database with a broad coverage of the grey literature (OpenAlex). Some of the databases that will be covered by the review are: MEDLINE (PubMed, Embase, OpenAlex), PubMed Central (PubMed), Embase (Embase), Crossref (OpenAlex), DOAJ (OpenAlex), preprint repositories (OpenAlex and PubMed), Zenodo (OpenAlex), and Internet Archive (OpenAlex).

### Terms and queries

The text words contained in the titles and abstracts of relevant articles, and the index terms used to describe the articles were used to develop a full search strategy during the pilot search and screening phases. The studies recovered were used to refine the search strategy and to identify additional relevant keywords through litsearchr.[11](#ref-Grames2019) The full list of terms according to the PCC framework is as follows (shown here according to PubMed syntax only for simplicity):

### Population

**Mesh Terms / Mesh Major Topic**

* “Workplace”[MeSH Terms]
* “Working Conditions”[MeSH Terms]
* “Employment”[MeSH Terms]
* “Occupational Health”[MeSH Terms]
* “Occupational Diseases”[MeSH Terms]
* “Occupational Exposure”[MeSH Terms]
* “Workers compensation”[MeSH Terms]

**Entry Terms**

* Work\*
* Occupat\*
* Manufactur\*
* Job exposure matrix
* JEM

### Concept

**Entry Terms**

* “Probability of causation”
* “Probabilistic causation”
* “Balance of probabilities”
* “Aetiological fraction”
* “Etiologic fraction”
* “Causal fraction”
* “Attributable risk”
* “Attributable proportion”
* “Attributable fraction”
* “Proportional liability”
* “Causation”
* “Risk model”
* “Stochastic model”
* “Work attribution”
* “Assigned share”
* “Causal exposure response”
* “Relative contribution”
* “Causal discovery”

### Context

**Mesh Terms / Mesh Major Topic**

* “Carcinogens”[MeSH]
* “Neoplasms”[MeSH]
* “Neoplasms, Squamous Cell” [MeSH]
* “Adenocarcinoma of Lung”[MeSH]
* “Small Cell Lung Carcinoma”[MeSH]
* “Carcinoma, Non-Small-Cell Lung”[MeSH]

**Entry Terms**

* Neoplasms
* Cancer
* Carcinogens
* Malignant

### Additional Context

**Mesh Terms / Mesh Major Topic**

* “Lung”[Mesh]
* “Lung Diseases”[Mesh]
* “Lung Neoplasms”[MeSH Major Topic]
* “Adenocarcinoma of Lung”[MeSH]
* “Small Cell Lung Carcinoma”[MeSH]
* “Carcinoma, Non-Small-Cell Lung”[MeSH]

**Entry Terms**

* Lung
* Lung Diseases
* Lung Cancer
* Pulmonary
* Bronchial
* Tracheal

### Filters (NOT)

* letter [Publication Type]23.82
* comment [Publication Type]
* news [Publication Type]
* case report [Publication Type]
* congress [Publication Type]
* clinical conference [Publication Type]

PubMed query:

[1] “(((("Probability of causation" OR "Probabilistic causation" OR "Balance of probabilities" OR "Aetiological fraction" OR "Etiologic fraction" OR "Causal fraction" OR "Attributable risk" OR "Attributable proportion" OR "Attributable fraction" OR "Proportional liability" OR "Causation" OR "Risk model" OR "Stochastic model" OR "Work attribution" OR "Assigned share" OR "Causal exposure response" OR "Relative contribution" OR "Causal discovery")) AND (("Workplace"[MeSH Terms] OR "Working Conditions"[MeSH Terms] OR "Employment"[MeSH Terms] OR "Occupational Health"[MeSH Terms] OR "Occupational Diseases"[MeSH Terms] OR "Occupational Exposure"[MeSH Terms] OR "Workers Compensation"[MeSH Terms] OR work\* OR occupat\* OR manufactur\* OR "Job exposure matrix" OR "JEM"))) AND (("Carcinogens"[MeSH] OR "Neoplasms"[MeSH] OR "Neoplasms, Squamous Cell"[MeSH] OR "Adenocarcinoma of Lung"[MeSH] OR "Small Cell Lung Carcinoma"[MeSH] OR "Carcinoma, Non-Small-Cell Lung"[MeSH] OR neoplasms OR cancer\* OR carcinogens OR malignant)) AND (("Lung"[MeSH] OR "Lung Diseases"[MeSH] OR "Lung Neoplasms"[MeSH Major Topic] OR "Adenocarcinoma of Lung"[MeSH] OR "Small Cell Lung Carcinoma"[MeSH] OR "Carcinoma, Non-Small-Cell Lung"[MeSH] OR lung OR "Lung Diseases" OR "Lung Cancer" OR pulmonar\* OR bronch\* OR trache\*))) NOT (letter[Publication Type] OR editorial[Publication Type] OR comment[Publication Type] OR news[Publication Type] OR "case reports"[Publication Type] OR congress[Publication Type] OR "clinical conference"[Publication Type])”

Embase query:

[1] “((((‘probability of causation’ OR ‘probabilistic causation’ OR ‘balance of probabilities’ OR ‘aetiological fraction’ OR ‘etiologic fraction’ OR ‘causal fraction’ OR ‘attributable risk’ OR ‘attributable proportion’ OR ‘attributable fraction’ OR ‘proportional liability’ OR ‘causation’ OR ‘risk model’ OR ‘stochastic model’ OR ‘work attribution’ OR ‘assigned share’ OR ‘causal exposure response’ OR ‘relative contribution’ OR ‘causal discovery’)) AND ((‘workplace’/exp OR ‘work environment’/exp OR ‘employment’/exp OR ‘occupational health’/exp OR ‘occupational disease’/exp OR ‘occupational exposure’/exp OR ‘workman compensation’/exp OR ‘work*’ OR ’occupat*’ OR ‘manufactur*’ OR ’job exposure matrix’ OR ’jem’))) AND ((’carcinogen’/exp OR ’neoplasm’/exp OR ’squamous cell carcinoma’/exp OR ’lung adenocarcinoma’/exp OR ’small cell lung cancer’/exp OR ’non small cell lung cancer’/exp OR ’neoplasms’ OR ’cancer*’ OR ‘carcinogens’ OR ‘malignant’)) AND ((‘lung’/exp OR ‘lung disease’/exp OR ‘lung tumor’/exp/mj OR ‘lung adenocarcinoma’/exp OR ‘small cell lung cancer’/exp OR ‘non small cell lung cancer’/exp OR ‘lung’ OR ‘lung diseases’ OR ‘lung cancer’ OR ‘pulmonar*’ OR ’bronch*’ OR ’trache\*‘))) NOT (’letter’:it OR ‘editorial’:it OR ‘comment’:it OR ‘note’:it OR ‘news’:it OR ‘case reports’:it OR ‘conference abstract’:it OR ‘conference paper’:it OR ‘clinical trial’:it)”

OpenAlex

[1] “https://api.openalex.org/works?filter=title\_and\_abstract.search:((("workplace" OR "working conditions" OR "employment" OR "occupational" OR "occupation" OR "workers compensation" OR "work" OR "manufacture" OR "job exposure matrix" OR "JEM") AND ("Probability of causation" OR "Probabilistic causation" OR "Balance of probabilities" OR "Aetiological fraction" OR "Etiologic fraction" OR "Causal fraction" OR "Attributable risk" OR "Attributable proportion" OR "Attributable fraction" OR "Proportional liability" OR "Causation" OR "Risk model" OR "Stochastic model" OR "Work attribution" OR "Assigned share" OR "Causal exposure response" OR "Relative contribution" OR "Causal discovery") AND ("carcinogen" OR "neoplasms" OR "squamous cell carcinoma" OR "adenocarcinoma of lung" OR "lung adenocarcinoma" OR "small cell lung carcinoma" OR "non‑small‑cell lung carcinoma" OR "cancer" OR "malignant") AND ("lung" OR "lungs" OR "pulmonary" OR "bronchopulmonary" OR "bronchus" OR "bronchi" OR "trachea" OR "adenocarcinoma of lung" OR "lung adenocarcinoma" OR "small cell lung carcinoma" OR "non‑small‑cell lung carcinoma")))”

### Language restrictions

Studies will not be excluded due to language restrictions in the search strategy to avoid language publication bias.[12](#ref-pieper2021)

### Deduplication

The Automated Systematic Search Deduplicator (ASySD) will be used in R.[13](#ref-ASySD2023) Manual deduplication of remaining potential duplicates will be carried out in excel by first examining the repeated DOIs and then comparing the records. Empty abstract fields were searched in Google Scholar, WorldCat, and journal websites and added to the records when found. Reasons for manual removal of duplicates will documented and reported.

## Screening

Title and abstracts that are not in the English, Spanish, Dutch, or German (languages spoken at highly skilled level by researchers involved in the review) will be translated to English using Google Translate. Agreement in systematic reviews using translations from the more recent Google Translate models has been reported to be higher than 80%, with inter-rater interpretations being a greater source of disagreement than inexact translations.[14](#ref-Jackson2019) Potential inaccurate translations are not considered relevant for the purposes of this scoping review since the objective is to describe the applications of a the concept of the PoC and potential alternatives. Thus, attaining a large coverage of regions and languages is of higher relevance than highly precise data extraction.

The machine learning assisted learning tool ASReview[15](#ref-Vandeschoot2021) v.1.6.3 was used in the pilot screening stage, and v.2.0.2 or later[16](#ref-debruin2025) will be used for screening. The SAFE procedure will be followed as the stopping heuristic for screening.[17](#ref-Boetje2024) A full description of the screening procedures is provided in **Appendix 3**.

## Contacting Authors

No procedures for contacting authors are being considered at this moment because this is a scoping review with no plans to perform a meta-analysis.

## Data extraction

From each included source, the following entities will be extracted: descriptions of the study design, sample size, worker population characteristics, exposure types, and outcome definitions; exposure assessment methods and sources of exposure-response functions; statistical measures used for estimating excess risk and Probability of Causation (PoC), including any identified exposure thresholds for PoC ≥ 0.5; model assumptions and background risk estimates; descriptions and sources of uncertainty; and qualitative summaries of methodological features such as adjustment variables and analytical models. In addition, relevant metadata including authorship, publication year, and institutional affiliations will be recorded. Extraction instructions are provided in **Appendix 4**.

### Extraction stages

Pilot extraction (already conducted) and final extraction (yet to be conducted).

## Synthesis Plan

Descriptive analyses will be conducted to characterize the included studies in terms of their publication timeline and geographical distribution. The year of publication will be grouped by decade (1990–1999, 2000–2009, 2010–2019, and 2020–2025) and summarized in a frequency table and visualized with a bar chart. To explore the geographical distribution, the countries of the study populations will be extracted and visualized using a world map. This map will display the number of studies per country.

Study designs will be categorized as cohort studies, case-control studies, or case series, and their frequency will be summarized in a table. The specific design of each study will be cross-referenced with the exposures and populations studied to explore patterns in methodological approaches.

Occupational exposures will be classified by type (i.e., radiation, asbestos, coal tar pitch volatiles) and by the method used for exposure assessment (e.g., job-exposure matrices, direct measurements, etc). Outcome measures will be categorized according to whether they refer to lung cancer as a single entity or to specific subtypes. The methods used to ascertain outcomes (i.e., cancer registries, health records, panel assessment) will also be documented.

The calculation of the PoC will be described by identifying whether a threshold (e.g., PoC ≥ 0.5) was used, and whether confidence intervals were employed, as well as the assumptions and uncertainties described in each study. Causal assumptions, such as the assumed shape of the exposure-response curve, latency periods, or confounding control, will be reported. Sources of uncertainty described by authors will be grouped into major categories (i.e., statistical uncertainty, exposure misclassification, among others).

The review will identify methodological alternatives and improvements proposed in the literature. This may include Bayesian modeling approaches, attributable fraction estimates, mechanistic modeling strategies, or alternative outcomes such as loss of life expectancy, quality of life, among others. Where available, critiques of PoC and suggested alternatives for use in compensation schemes or legal settings will be summarized.

All analyses will be conducted in R (version 4.4.0), using the tidyverse suite for data wrangling and visualization, gt for formatted summary tables, and the maps package for generating geographical plots.

## Software

Open source documentation for all stages of the review will be conducted within an R project using quarto markdown (.qmd) documents in RStudio (v.2024.04.0) in Windows 11. R (v.4.4.0) packages will be used to refine the search strategy (litsearchr v.0.4.0) and deduplication with the Automated Systematic Search Deduplicator (ASySD v.0.4.3).[13](#ref-ASySD2023) The machine learning assisted learning tool ASReview v.1.6.3[15](#ref-Vandeschoot2021) will be used in the pilot screening stage, while ASReview v.2.0.2[16](#ref-debruin2025) will be used for the final study screening phase.

# References

1. Coenen P, Kezic S, Heederik DJJ, Peters S, Molen HF van der. Applying a ‘presumably plausible’ principle in a new one-time financial compensation system for occupational diseases in the Netherlands. *Occupational and Environmental Medicine*. Published online September 20, 2024. doi:[10.1136/oemed-2024-109533](https://doi.org/10.1136/oemed-2024-109533)

2. Moon J, Yoo H. Misdiagnosis in occupational and environmental medicine: a scoping review. *Journal of Occupational Medicine and Toxicology*. 2021;16(1):33. doi:[10.1186/s12995-021-00325-z](https://doi.org/10.1186/s12995-021-00325-z)

3. Toby Mathias CG. Contact dermatitis and workers’ compensation: Criteria for establishing occupational causation and aggravation. *Journal of the American Academy of Dermatology*. 1989;20(5, Part 1):842-848. doi:[10.1016/S0190-9622(89)70096-7](https://doi.org/10.1016/S0190-9622(89)70096-7)

4. Greenland S. Relation of probability of causation to relative risk and doubling dose: a methodologic error that has become a social problem. *American Journal of Public Health*. 1999;89(8):1166-1169. doi:[10.2105/AJPH.89.8.1166](https://doi.org/10.2105/AJPH.89.8.1166)

5. Siemiatycki J, Karp I, Sylvestre MP, Pintos J. Estimating the Proportion of Cases of Lung Cancer Legally Attributable to Smoking: A Novel Approach for Class Actions Against the Tobacco Industry. *American Journal of Public Health*. 2014;104(8):e60-e66. doi:[10.2105/AJPH.2014.302040](https://doi.org/10.2105/AJPH.2014.302040)

6. Wakeford R, Antell BA, Leigh WJ. A Compensation Scheme for Nuclear Industry Workers in the United Kingdom: *Health Physics*. 1998;74(1):1-9. doi:[10.1097/00004032-199801000-00001](https://doi.org/10.1097/00004032-199801000-00001)

7. Land C, Gilbert E, Smith JM. *Report of the NCI-CDC Working Group to Revise the 1985 NIH Radioepidemiological Tables*.; 2003:118. <https://www.govinfo.gov/app/details/GOVPUB-HE20-PURL-LPS47231>

8. Armstrong B, Tremblay C, Theriault G. Compensating Bladder Cancer Victims Employed in Aluminum Reduction Plants. *Journal of Occupational and Environmental Medicine*. 1988;30(10):771-775. doi:[10.1097/00043764-198810000-00004](https://doi.org/10.1097/00043764-198810000-00004)

9. Greenland S. Concepts and pitfalls in measuring and interpreting attributable fractions, prevented fractions, and causation probabilities. *Annals of Epidemiology*. 2015;25(3):155-161. doi:[10.1016/j.annepidem.2014.11.005](https://doi.org/10.1016/j.annepidem.2014.11.005)

10. Kim EA, Lee E, Kang SK, Jeong M. Probability of causation for occupational cancer after exposure to ionizing radiation. *Annals of Occupational and Environmental Medicine*. 2018;30(1):3. doi:[10.1186/s40557-018-0220-5](https://doi.org/10.1186/s40557-018-0220-5)

11. Grames EM, Stillman AN, Tingley MW, Elphick CS. An automated approach to identifying search terms for systematic reviews using keyword co‐occurrence networks. Freckleton R, ed. *Methods in Ecology and Evolution*. 2019;10(10):1645-1654. doi:[10.1111/2041-210X.13268](https://doi.org/10.1111/2041-210X.13268)

12. Pieper D, Puljak L. Language restrictions in systematic reviews should not be imposed in the search strategy but in the eligibility criteria if necessary. *Journal of Clinical Epidemiology*. 2021;132:146-147. doi:[10.1016/j.jclinepi.2020.12.027](https://doi.org/10.1016/j.jclinepi.2020.12.027)

13. Hair K, Bahor Z, Macleod M, Liao J, Sena ES. The automated systematic search deduplicator (ASySD): A rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. *BMC Biology*. 2023;21(1):189. doi:[10.1186/s12915-023-01686-z](https://doi.org/10.1186/s12915-023-01686-z)

14. Jackson JL, Kuriyama A, Anton A, et al. The Accuracy of Google Translate for Abstracting Data From NonEnglish-Language Trials for Systematic Reviews. *Annals of Internal Medicine*. 2019;171(9):677. doi:[10.7326/M19-0891](https://doi.org/10.7326/M19-0891)

15. Van De Schoot R, De Bruin J, Schram R, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*. 2021;3(2):125-133. doi:[10.1038/s42256-020-00287-7](https://doi.org/10.1038/s42256-020-00287-7)

16. De Bruin J, Lombaers P, Kaandorp C, et al. ASReview LAB v.2: Open-source text screening with multiple agents and a crowd of experts. *Patterns*. 2025;6(7):101318. doi:[10.1016/j.patter.2025.101318](https://doi.org/10.1016/j.patter.2025.101318)

17. Boetje J, Van De Schoot R. The SAFE procedure: a practical stopping heuristic for active learning-based screening in systematic reviews and meta-analyses. *Systematic Reviews*. 2024;13(1):81. doi:[10.1186/s13643-024-02502-7](https://doi.org/10.1186/s13643-024-02502-7)

# Appendix 1: Pilot Search Strategy (Lung Cancer)

An exploratory search of the literature was conducted in PubMed, Embase, OpenAlex, Policy Commons, and Overton. A summary of records is presented bellow:

Table 1: Summary of records for lung cancer

|  | Records | |
| --- | --- | --- |
| Database | Lung Cancer | Last 10 years |
| PubMed | 210 | 51 |
| Embase | 289 | 96 |
| OpenAlex | 299 | 124 |
| Policy Commons | 1867 | 974 |
| Overton | 7627 | 4617 |

The following terms were used for the search strategies and adapted for each database:

### Population

**Entry Terms**

* Neoplasms
* Cancer
* Carcinogens

**Mesh Terms / Mesh Major Topic**

* “Carcinogens”[MeSH Major Topic]
* “neoplasms/epidemiology”[MeSH Major Topic]
* “neoplasms/etiology”[MeSH Terms]
* “neoplasms/mortality”[MeSH Major Topic]

### Additional Context

**Entry Terms**

* Lung
* Lung Diseases
* Lung Cancer
* Pulmonary
* Bronchial
* Tracheal

**Mesh Terms / Mesh Major Topic**

* “Lung”[Mesh]
* “Lung Diseases”[Mesh]
* “Lung Neoplasms”[MeSH Major Topic]

### Concept

**Entry Terms**

* Probability of causation
* Probabilistic causation
* Balance of probabilities
* Assigned share
* Aetiological fraction
* Etiologic fraction
* Causal fraction
* Attributable risk
* Attributable proportion
* Attributable fraction
* Proportional liability
* Causation
* Risk model
* Stochastic model
* Work attribution
* Causal exposure response
* Dose response
* Excess risk
* Excess relative risk - to be included after pilot

### Context

**Mesh Terms / Mesh Major Topic**

* “Workplace”[MeSH Terms]
* “Working Conditions”[MeSH Terms]
* “Employment”[MeSH Terms]
* “Occupational Exposure”[MeSH Terms]
* “Occupational Diseases”[MeSH Terms]
* “occupational exposure/analysis”[MeSH Major Topic]
* “occupational diseases/chemically induced”[MeSH Terms]
* “occupational diseases/epidemiology”[MeSH Terms]
* “occupational exposure/adverse effects”[MeSH Major Topic]
* “workers compensation/economics”[MeSH Major Topic]
* “workers compensation/statistics and numerical data”[MeSH Major Topic]
* “workers compensation”[MeSH Terms]

#### PubMed

Population

("Workplace"[tiab] OR "Working Conditions"[tiab] OR "Employment"[tiab] OR "Occupational Exposure"[tiab] OR "Occupational Diseases"[tiab] OR "occupational exposure/analysis"[tiab] OR "occupational diseases/chemically induced"[tiab] OR "occupational diseases/epidemiology"[tiab] OR "occupational exposure/adverse effects"[tiab] OR "Workplace"[MeSH Terms] OR "Working Conditions"[MeSH Terms] OR "Employment"[MeSH Terms] OR "Occupational Exposure"[MeSH Terms] OR "Occupational Diseases"[MeSH Terms] OR "occupational exposure/analysis"[MeSH Major Topic] OR "occupational diseases/chemically induced"[MeSH Terms] OR "occupational diseases/epidemiology"[MeSH Terms] OR "occupational exposure/adverse effects"[MeSH Major Topic] OR "workers compensation/economics"[MeSH Major Topic] OR "workers compensation/statistics and numerical data"[MeSH Major Topic] OR "workers compensation"[MeSH Terms] OR "workers compensation"[tiab])

Concept

("Probability of causation"[tiab] OR "Probabilistic causation"[tiab] OR "Balance of probabilities"[tiab] OR "Aetiological fraction"[tiab] OR "Etiologic fraction"[tiab] OR "Causal fraction"[tiab] OR "Attributable risk"[tiab] OR "Attributable proportion"[tiab] OR "Attributable fraction"[tiab] OR "Proportional liability"[tiab] OR "Causation"[tiab] OR "Risk model"[tiab] OR "Stochastic model"[tiab] OR "Work attribution"[tiab] OR "Assigned share"[tiab])

Context

("neoplasms"[tiab] OR "Cancer"[tiab] OR "Carcinogens"[tiab] OR "neoplasms/epidemiology"[tiab] OR "neoplasms/etiology"[tiab] OR "neoplasms/mortality"[tiab] OR "Carcinogens"[MeSH Major Topic] OR "neoplasms/epidemiology"[MeSH Major Topic] OR "neoplasms/etiology"[MeSH Terms] OR "neoplasms/mortality"[MeSH Major Topic])

Context (Additional)

("Lung"[Mesh] OR "Lung Diseases"[Mesh] OR lung[tiab] OR lungs[tiab] OR pulmonar\*[tiab] OR bronchop\*[tiab] OR bronchu\*[tiab] OR bronchi\*[tiab] OR trache\*[tiab])

Full search string

(((("Workplace"[tiab] OR "Working Conditions"[tiab] OR "Employment"[tiab] OR "Occupational Exposure"[tiab] OR "Occupational Diseases"[tiab] OR "occupational exposure/analysis"[tiab] OR "occupational diseases/chemically induced"[tiab] OR "occupational diseases/epidemiology"[tiab] OR "occupational exposure/adverse effects"[tiab] OR "Workplace"[MeSH Terms] OR "Working Conditions"[MeSH Terms] OR "Employment"[MeSH Terms] OR "Occupational Exposure"[MeSH Terms] OR "Occupational Diseases"[MeSH Terms] OR "occupational exposure/analysis"[MeSH Major Topic] OR "occupational diseases/chemically induced"[MeSH Terms] OR "occupational diseases/epidemiology"[MeSH Terms] OR "occupational exposure/adverse effects"[MeSH Major Topic] OR "workers compensation/economics"[MeSH Major Topic] OR "workers compensation/statistics and numerical data"[MeSH Major Topic] OR "workers compensation"[MeSH Terms] OR "workers compensation"[tiab])) AND (("Probability of causation"[tiab] OR "Probabilistic causation"[tiab] OR "Balance of probabilities"[tiab] OR "Aetiological fraction"[tiab] OR "Etiologic fraction"[tiab] OR "Causal fraction"[tiab] OR "Attributable risk"[tiab] OR "Attributable proportion"[tiab] OR "Attributable fraction"[tiab] OR "Proportional liability"[tiab] OR "Causation"[tiab] OR "Risk model"[tiab] OR "Stochastic model"[tiab] OR "Work attribution"[tiab] OR "Assigned share"[tiab]))) AND (("neoplasms"[tiab] OR "Cancer"[tiab] OR "Carcinogens"[tiab] OR "neoplasms/epidemiology"[tiab] OR "neoplasms/etiology"[tiab] OR "neoplasms/mortality"[tiab] OR "Carcinogens"[MeSH Major Topic] OR "neoplasms/epidemiology"[MeSH Major Topic] OR "neoplasms/etiology"[MeSH Terms] OR "neoplasms/mortality"[MeSH Major Topic]))) AND (("Lung"[Mesh] OR "Lung Diseases"[Mesh] OR lung[tiab] OR lungs[tiab] OR pulmonar\*[tiab] OR bronchop\*[tiab] OR bronchu\*[tiab] OR bronchi\*[tiab] OR trache\*[tiab]))

Number of records: 210

Restricting to last 10 years

(((("Workplace"[tiab] OR "Working Conditions"[tiab] OR "Employment"[tiab] OR "Occupational Exposure"[tiab] OR "Occupational Diseases"[tiab] OR "occupational exposure/analysis"[tiab] OR "occupational diseases/chemically induced"[tiab] OR "occupational diseases/epidemiology"[tiab] OR "occupational exposure/adverse effects"[tiab] OR "Workplace"[MeSH Terms] OR "Working Conditions"[MeSH Terms] OR "Employment"[MeSH Terms] OR "Occupational Exposure"[MeSH Terms] OR "Occupational Diseases"[MeSH Terms] OR "occupational exposure/analysis"[MeSH Major Topic] OR "occupational diseases/chemically induced"[MeSH Terms] OR "occupational diseases/epidemiology"[MeSH Terms] OR "occupational exposure/adverse effects"[MeSH Major Topic] OR "workers compensation/economics"[MeSH Major Topic] OR "workers compensation/statistics and numerical data"[MeSH Major Topic] OR "workers compensation"[MeSH Terms] OR "workers compensation"[tiab])) AND (("Probability of causation"[tiab] OR "Probabilistic causation"[tiab] OR "Balance of probabilities"[tiab] OR "Aetiological fraction"[tiab] OR "Etiologic fraction"[tiab] OR "Causal fraction"[tiab] OR "Attributable risk"[tiab] OR "Attributable proportion"[tiab] OR "Attributable fraction"[tiab] OR "Proportional liability"[tiab] OR "Causation"[tiab] OR "Risk model"[tiab] OR "Stochastic model"[tiab] OR "Work attribution"[tiab] OR "Assigned share"[tiab]))) AND (("neoplasms"[tiab] OR "Cancer"[tiab] OR "Carcinogens"[tiab] OR "neoplasms/epidemiology"[tiab] OR "neoplasms/etiology"[tiab] OR "neoplasms/mortality"[tiab] OR "Carcinogens"[MeSH Major Topic] OR "neoplasms/epidemiology"[MeSH Major Topic] OR "neoplasms/etiology"[MeSH Terms] OR "neoplasms/mortality"[MeSH Major Topic]))) AND (("Lung"[Mesh] OR "Lung Diseases"[Mesh] OR lung[tiab] OR lungs[tiab] OR pulmonar\*[tiab] OR bronchop\*[tiab] OR bronchu\*[tiab] OR bronchi\*[tiab] OR trache\*[tiab])) AND ("2014"[dp] : "2024"[dp])

Number of records (Last 10 years): 51

#### Embase

(workplace:ti,ab OR 'working conditions':ti,ab OR employment:ti,ab OR 'occupational exposure':ti,ab OR 'occupational diseases':ti,ab OR 'occupational exposure'/exp OR 'occupational diseases'/exp OR 'occupational exposure/analysis' OR 'occupational diseases/chemically induced' OR 'occupational diseases/epidemiology' OR 'occupational exposure/adverse effects' OR 'workers compensation'/exp OR 'workers compensation/economics' OR 'workers compensation/statistics') AND ('probability of causation':ti,ab OR 'probabilistic causation':ti,ab OR 'balance of probabilities':ti,ab OR 'aetiological fraction':ti,ab OR 'etiologic fraction':ti,ab OR 'causal fraction':ti,ab OR 'attributable risk':ti,ab OR 'attributable proportion':ti,ab OR 'attributable fraction':ti,ab OR 'proportional liability':ti,ab OR causation:ti,ab OR 'risk model':ti,ab OR 'stochastic model':ti,ab OR 'work attribution':ti,ab OR 'assigned share':ti,ab) AND ('neoplasms':ti,ab OR cancer:ti,ab OR carcinogens:ti,ab OR 'neoplasm'/exp OR 'carcinogen'/exp OR 'neoplasms/epidemiology' OR 'neoplasms/etiology' OR 'neoplasms/mortality' OR 'carcinogen'/mj) AND ('lung'/exp OR 'lung diseases'/exp OR lung:ti,ab OR lungs:ti,ab OR pulmonar\*:ti,ab OR bronchop\*:ti,ab OR bronchu\*:ti,ab OR bronchi\*:ti,ab OR trache\*:ti,ab)

Number of records: 289

Restricting to last 10 years

(workplace:ti,ab OR 'working conditions':ti,ab OR employment:ti,ab OR 'occupational exposure':ti,ab OR 'occupational diseases':ti,ab OR 'occupational exposure'/exp OR 'occupational diseases'/exp OR 'occupational exposure/analysis' OR 'occupational diseases/chemically induced' OR 'occupational diseases/epidemiology' OR 'occupational exposure/adverse effects' OR 'workers compensation'/exp OR 'workers compensation/economics' OR 'workers compensation/statistics') AND ('probability of causation':ti,ab OR 'probabilistic causation':ti,ab OR 'balance of probabilities':ti,ab OR 'aetiological fraction':ti,ab OR 'etiologic fraction':ti,ab OR 'causal fraction':ti,ab OR 'attributable risk':ti,ab OR 'attributable proportion':ti,ab OR 'attributable fraction':ti,ab OR 'proportional liability':ti,ab OR causation:ti,ab OR 'risk model':ti,ab OR 'stochastic model':ti,ab OR 'work attribution':ti,ab OR 'assigned share':ti,ab) AND ('neoplasms':ti,ab OR cancer:ti,ab OR carcinogens:ti,ab OR 'neoplasm'/exp OR 'carcinogen'/exp OR 'neoplasms/epidemiology' OR 'neoplasms/etiology' OR 'neoplasms/mortality' OR 'carcinogen'/mj) AND ('lung'/exp OR 'lung diseases'/exp OR lung:ti,ab OR lungs:ti,ab OR pulmonar\*:ti,ab OR bronchop\*:ti,ab OR bronchu\*:ti,ab OR bronchi\*:ti,ab OR trache\*:ti,ab) AND [2014-2024]/py

Number of records (last 10 years): 96

OpenAlex

https://api.openalex.org/works?filter=title\_and\_abstract.search:((("workplace" OR "working conditions" OR "employment" OR "occupational exposure" OR "occupational diseases" OR "workers compensation") AND ("probability of causation" OR "probabilistic causation" OR "balance of probabilities" OR "aetiological fraction" OR "etiologic fraction" OR "causal fraction" OR "attributable risk" OR "attributable proportion" OR "attributable fraction" OR "proportional liability" OR "causation" OR "risk model" OR "stochastic model" OR "work attribution" OR "assigned share") AND ("neoplasms" OR "cancer" OR "carcinogens") AND ("lung" OR "lungs" OR "pulmonary" OR "bronchopulmonary" OR "bronchus" OR "bronchi" OR "trachea")))

Number of records (title and abstract): 299

Number of records (title and abstract, last 10 years): 124

Number of records (full-text): 11290

Number of records (full-text, last 10 years): 3951

#### Policy Commons

Restricting to lung cancer only

("workplace" OR "working conditions" OR "employment" OR "occupational exposure" OR "occupational diseases" OR "workers compensation") AND ("probability of causation" OR "probabilistic causation" OR "balance of probabilities" OR "aetiological fraction" OR "etiologic fraction" OR "causal fraction" OR "attributable risk" OR "attributable proportion" OR "attributable fraction" OR "proportional liability" OR "causation" OR "risk model" OR "stochastic model" OR "work attribution" OR "assigned share") AND ("neoplasms" OR "cancer" OR "carcinogens") AND ("lung" OR "lungs" OR "pulmonary" OR "bronchopulmonary" OR "bronchus" OR "bronchi" OR "trachea")

Number of records: 1867

Number of records (last 10 years): 974

#### Overton

("workplace" OR "working conditions" OR "employment" OR "occupational exposure" OR "occupational diseases" OR "workers compensation")   
AND   
("probability of causation" OR "probabilistic causation" OR "balance of probabilities" OR "aetiological fraction" OR "etiologic fraction" OR "causal fraction" OR "attributable risk" OR "attributable proportion" OR "attributable fraction" OR "proportional liability" OR "causation" OR "risk model" OR "stochastic model" OR "work attribution" OR "assigned share")   
AND   
("neoplasms" OR "cancer" OR "carcinogens")   
AND   
("lung" OR "lungs" OR "pulmonary" OR "bronchopulmonary" OR "bronchus" OR "bronchi" OR "trachea")

Number of records: 7627

Number of records (last 10 years): 4617

# Appendix 2: Pilot Search Strategy (All types of cancer)

Table 2: Summary of records for all types of cancer

|  | Records | |
| --- | --- | --- |
| Database | All types of cancer | Last 10 years |
| PubMed | 412 | 120 |
| Embase | 559 | 209 |
| OpenAlex | 629 | 276 |
| Policy Commons | 3971 | 2155 |
| Overton | 13130 | 8018 |

#### PubMed

((("Workplace\*"[tiab] OR "Working Conditions"[tiab] OR "Employment"[tiab] OR "Occupational Exposure"[tiab] OR "Occupational Diseases"[tiab] OR "occupational exposure/analysis"[tiab] OR "occupational diseases/chemically induced"[tiab] OR "occupational diseases/epidemiology"[tiab] OR "occupational exposure/adverse effects"[tiab] OR "Workplace"[MeSH Terms] OR "Working Conditions"[MeSH Terms] OR "Employment"[MeSH Terms] OR "Occupational Exposure"[MeSH Terms] OR "Occupational Diseases"[MeSH Terms] OR "occupational exposure/analysis"[MeSH Major Topic] OR "occupational diseases/chemically induced"[MeSH Terms] OR "occupational diseases/epidemiology"[MeSH Terms] OR "occupational exposure/adverse effects"[MeSH Major Topic] OR "workers compensation/economics"[MeSH Major Topic] OR "workers compensation/statistics and numerical data"[MeSH Major Topic] OR "workers compensation"[MeSH Terms] OR "workers compensation"[tiab])) AND (("Probability of causation"[tiab] OR "Probabilistic causation"[tiab] OR "Balance of probabilities"[tiab] OR "Aetiological fraction"[tiab] OR "Etiologic fraction"[tiab] OR "Causal fraction"[tiab] OR "Attributable risk"[tiab] OR "Attributable proportion"[tiab] OR "Attributable fraction"[tiab] OR "Proportional liability"[tiab] OR "Causation"[tiab] OR "Risk model"[tiab] OR "Stochastic model"[tiab] OR "Work attribution"[tiab] OR "Assigned share"[tiab]))) AND (("neoplasms"[tiab] OR "Cancer"[tiab] OR "Carcinogens"[tiab] OR "neoplasms/epidemiology"[tiab] OR "neoplasms/etiology"[tiab] OR "neoplasms/mortality"[tiab] OR "Carcinogens"[MeSH Major Topic] OR "neoplasms/epidemiology"[MeSH Major Topic] OR "neoplasms/etiology"[MeSH Terms] OR "neoplasms/mortality"[MeSH Major Topic]))

Number of records: 412

Restricting to last 10 years

((("Workplace\*"[tiab] OR "Working Conditions"[tiab] OR "Employment"[tiab] OR "Occupational Exposure"[tiab] OR "Occupational Diseases"[tiab] OR "occupational exposure/analysis"[tiab] OR "occupational diseases/chemically induced"[tiab] OR "occupational diseases/epidemiology"[tiab] OR "occupational exposure/adverse effects"[tiab] OR "Workplace"[MeSH Terms] OR "Working Conditions"[MeSH Terms] OR "Employment"[MeSH Terms] OR "Occupational Exposure"[MeSH Terms] OR "Occupational Diseases"[MeSH Terms] OR "occupational exposure/analysis"[MeSH Major Topic] OR "occupational diseases/chemically induced"[MeSH Terms] OR "occupational diseases/epidemiology"[MeSH Terms] OR "occupational exposure/adverse effects"[MeSH Major Topic] OR "workers compensation/economics"[MeSH Major Topic] OR "workers compensation/statistics and numerical data"[MeSH Major Topic] OR "workers compensation"[MeSH Terms] OR "workers compensation"[tiab])) AND (("Probability of causation"[tiab] OR "Probabilistic causation"[tiab] OR "Balance of probabilities"[tiab] OR "Aetiological fraction"[tiab] OR "Etiologic fraction"[tiab] OR "Causal fraction"[tiab] OR "Attributable risk"[tiab] OR "Attributable proportion"[tiab] OR "Attributable fraction"[tiab] OR "Proportional liability"[tiab] OR "Causation"[tiab] OR "Risk model"[tiab] OR "Stochastic model"[tiab] OR "Work attribution"[tiab] OR "Assigned share"[tiab]))) AND (("neoplasms"[tiab] OR "Cancer"[tiab] OR "Carcinogens"[tiab] OR "neoplasms/epidemiology"[tiab] OR "neoplasms/etiology"[tiab] OR "neoplasms/mortality"[tiab] OR "Carcinogens"[MeSH Major Topic] OR "neoplasms/epidemiology"[MeSH Major Topic] OR "neoplasms/etiology"[MeSH Terms] OR "neoplasms/mortality"[MeSH Major Topic])) AND ("2014"[dp] : "2024"[dp])

Number of records (last 10 years): 120

Embase

(workplace:ti,ab OR 'working conditions':ti,ab OR employment:ti,ab OR 'occupational exposure':ti,ab OR 'occupational diseases':ti,ab OR 'occupational exposure'/exp OR 'occupational diseases'/exp OR 'occupational exposure/analysis' OR 'occupational diseases/chemically induced' OR 'occupational diseases/epidemiology' OR 'occupational exposure/adverse effects' OR 'workers compensation'/exp OR 'workers compensation/economics' OR 'workers compensation/statistics') AND ('probability of causation':ti,ab OR 'probabilistic causation':ti,ab OR 'balance of probabilities':ti,ab OR 'aetiological fraction':ti,ab OR 'etiologic fraction':ti,ab OR 'causal fraction':ti,ab OR 'attributable risk':ti,ab OR 'attributable proportion':ti,ab OR 'attributable fraction':ti,ab OR 'proportional liability':ti,ab OR causation:ti,ab OR 'risk model':ti,ab OR 'stochastic model':ti,ab OR 'work attribution':ti,ab OR 'assigned share':ti,ab) AND ('neoplasms':ti,ab OR cancer:ti,ab OR carcinogens:ti,ab OR 'neoplasm'/exp OR 'carcinogen'/exp OR 'neoplasms/epidemiology' OR 'neoplasms/etiology' OR 'neoplasms/mortality' OR 'carcinogen'/mj)

Number of records: 559

Restricting to last 10 years

(workplace:ti,ab OR 'working conditions':ti,ab OR employment:ti,ab OR 'occupational exposure':ti,ab OR 'occupational diseases':ti,ab OR 'occupational exposure'/exp OR 'occupational diseases'/exp OR 'occupational exposure/analysis' OR 'occupational diseases/chemically induced' OR 'occupational diseases/epidemiology' OR 'occupational exposure/adverse effects' OR 'workers compensation'/exp OR 'workers compensation/economics' OR 'workers compensation/statistics') AND ('probability of causation':ti,ab OR 'probabilistic causation':ti,ab OR 'balance of probabilities':ti,ab OR 'aetiological fraction':ti,ab OR 'etiologic fraction':ti,ab OR 'causal fraction':ti,ab OR 'attributable risk':ti,ab OR 'attributable proportion':ti,ab OR 'attributable fraction':ti,ab OR 'proportional liability':ti,ab OR causation:ti,ab OR 'risk model':ti,ab OR 'stochastic model':ti,ab OR 'work attribution':ti,ab OR 'assigned share':ti,ab) AND ('neoplasms':ti,ab OR cancer:ti,ab OR carcinogens:ti,ab OR 'neoplasm'/exp OR 'carcinogen'/exp OR 'neoplasms/epidemiology' OR 'neoplasms/etiology' OR 'neoplasms/mortality' OR 'carcinogen'/mj) AND [2014-2024]/py

Number of records: 209

#### OpenAlex

https://api.openalex.org/works?filter=title\_and\_abstract.search:((("workplace" OR "working conditions" OR "employment" OR "occupational exposure" OR "occupational diseases" OR "workers compensation") AND ("probability of causation" OR "probabilistic causation" OR "balance of probabilities" OR "aetiological fraction" OR "etiologic fraction" OR "causal fraction" OR "attributable risk" OR "attributable proportion" OR "attributable fraction" OR "proportional liability" OR "causation" OR "risk model" OR "stochastic model" OR "work attribution" OR "assigned share") AND ("neoplasms" OR "cancer" OR "carcinogens")))

Number of records (title and abstract): 629

Number of records (title and abstract, last 10 years): 276

Number of records (full-text): 22992

Number of records (full-text, last 10 years): 8835

#### Policy Commons

("workplace" OR "working conditions" OR "employment" OR "occupational exposure" OR "occupational diseases" OR "workers compensation") AND ("probability of causation" OR "probabilistic causation" OR "balance of probabilities" OR "aetiological fraction" OR "etiologic fraction" OR "causal fraction" OR "attributable risk" OR "attributable proportion" OR "attributable fraction" OR "proportional liability" OR "causation" OR "risk model" OR "stochastic model" OR "work attribution" OR "assigned share") AND ("neoplasms" OR "cancer" OR "carcinogens") AND ("lung" OR "lungs" OR "pulmonary" OR "bronchopulmonary" OR "bronchus" OR "bronchi" OR "trachea")

Number of records: 3971

Number of records (last 10 years): 2155

#### Overton

("workplace" OR "working conditions" OR "employment" OR "occupational exposure" OR "occupational diseases" OR "workers compensation")   
AND   
("probability of causation" OR "probabilistic causation" OR "balance of probabilities" OR "aetiological fraction" OR "etiologic fraction" OR "causal fraction" OR "attributable risk" OR "attributable proportion" OR "attributable fraction" OR "proportional liability" OR "causation" OR "risk model" OR "stochastic model" OR "work attribution" OR "assigned share")   
AND   
("neoplasms" OR "cancer" OR "carcinogens")

Number of records: 13130

Number of records (last 10 years): 8018

# Appendix 3: Screening Procedures

## Full-length review screening

Three percent of the records were randomly screened to determine the number of relevant records expected in the whole dataset, according to the SAFE procedure. The size of the training set ( ) is:

Thus, **34 records** were randomly reviewed.

There was 1 relevant record () found. Thus, the number of relevant records () expected in the whole dataset () is:

A total of **34 relevant records** are expected.

According to the SAFE procedure, active learning with an initial lightweight model (ELAS u4, default configuration) will be used according to the following stopping heuristics:

* At least twice the () should be screened: **68 records**
* A minimum of 10% of the records should be screened: **113 records**
* No extra relevant records have been identified in the last 25 records.

When these criteria are met, the model will be changed to the ELAS h3 deep learning model (default configuration), to allow the identification of more complex semantic context that can uncover difficult to find records.

The stopping heuristic will be:

* No extra relevant records have been identified in the last 25 records.

## Pilot Screening

For the pilot screening phase, 5% of the records from the lung cancer search were be randomly screened to determine the number of relevant records expected in the whole dataset, according to the SAFE procedure. The size of the training set ( ) is:

Thus, **23 records** were randomly reviewed.

There was 1 relevant record () found. Thus, the number of relevant records () expected in the whole dataset () is:

A total of **20 relevant records** were expected, but only 10 were selected chosen for pilot data extraction. These records were later incorporated as prior knowledge in the full dataset used in the scoping review.

# Appendix 4. Extraction Instructions

As part of the data extraction phase of this scoping review, a structured dataset was generated based. Reviewers are requested to verify, annotate, and refine entries for each study beginning from the column labeled type, and extending across all subsequent variables in the dataset. Stated differently, colored columns should not be modified, except for reclassification of the “included” columns as explained below.

* Included: The included variable indicates whether a study has met all eligibility criteria for inclusion in the final data synthesis. If, upon full-text review, a study is found not to satisfy the pre-specified selection criteria, screeners may manually update this column to reflect its exclusion.
* Study Type: Register the study design classification (i.e., “cohort”, “case-control”, “case-series”).
* Exposure: Confirm the primary occupational exposure(s) evaluated in the study (i.e., “radiation”, “asbestos”, “coal tar pitch volatiles”). If multiple exposures are discussed, retain only those for whic PoC was estimated.
* Worker Population: Assess the description of the study population, including industry or occupational group (i.e., “uranium miners”, “nuclear plant workers”, “aluminum smelter workers”). This should reflect the primary occupational group under investigation for PoC estimation.
* Outcome: Review the reported health outcome(s) for which PoC was calculated. Note whether lung cancer was assessed as a single entity or by histologic subtype.
* Exposure Assessment: Confirm the method by which exposure was estimated or reconstructed. Common methods include job-exposure matrices (JEMs), biomonitoring, registry data, or self-reported histories.
* Source of Exposure-Response Relationship: Record the population or study from which the exposure-response function was derived.
* Measure for Excess Risk: Identify the statistical measure used to estimate excess relative risk (ERR) or equivalent (i.e., “odds ratio”, “incidence rate ratio”).
* Threshold for PoC ≥ 0.5: Indicate whether a quantitative threshold of exposure was identified at which the PoC is greater than or equal to 0.5.
* Exposure Threshold Level: Where applicable, record the specific exposure level (i.e., “4.3 fiber-years”, “100 µg/m³-years benzo[a]pyrene”) associated with PoC ≥ 0.5.
* Uncertainty Characterization: Summarize the types of uncertainty discussed in the study, such as statistical error, measurement error, population transferability, or model mis-specification.
* Uncertainty Description: Where provided, include more detailed explanations of uncertainty, such as the use of upper-bound estimates, Monte Carlo simulation, or Bayesian interval estimation.
* Background Risk: Clarify the assumed or modeled background risk of lung cancer in the reference (non-exposed) population. This may include population incidence rates, comparator worker groups, or transferred baseline risk from external cohorts.
* Model Assumptions: Identify and list the model assumptions explicitly stated in the study, particularly related to confounding adjustment (i.e., age, sex, smoking), latency, or dose-response form.
* Comments: This column should be used to record any clarifications, methodological notes, or supplementary information not captured in the other variables. For example, mention if a study includes additional datasets, software tools used for PoC implementation, or correspondence with authors.

For transparency and reproducibility, please maintain a log of any substantive reclassifications made during your review of the dataset.