Probability of Causation: SYNERGY meta-analysis

Lung cancer due to occupational asbestos exposure

Javier Mancilla-Galindo Susan Peters Jenny Deng
Henk van der Molen Hans Kromhout Lützen Portengen
Roel Vermeulen Dick Heederik

2025-08-14

Table of contents

Descriptive characteristics of the SYNERGY study population

	Total (N=37866)	Controls (N=20965)	Lung cancer cases (N=16901)
sex			
Female	7810 (20.6%)	4514 (21.5%)	3296 (19.5%)
Male	30056 (79.4%)	16451 (78.5%)	13605 (80.5%)
Age			
Mean (SD)	61.7 (9.63)	61.5 (9.92)	62.0 (9.23)
Asbestos (ff/ml-years)			
Mean (SD)	2.42 (3.01)	2.22 (3.01)	2.60 (3.00)
Median (Q1, Q3)	1.33 (0.550, 3.22)	1.17 (0.476, 2.86)	1.54 (0.622, 3.53)
Min, Max	0.00214, 64.6	0.00229, 64.6	0.00214, 35.4
Never exposed	23114 (61.0%)	13653 (65.1%)	9461 (56.0%)
Exposure duration (years)			
Mean (SD)	18.4 (13.9)	17.9 (13.8)	18.9 (14.0)
Median (Q1, Q3)	15.0 (6.00, 30.0)	14.0 (5.00, 30.0)	15.5 (6.00, 31.0)
Min, Max	1.00, 63.0	1.00, 63.0	1.00, 62.0
Never exposed	23114 (61.0%)	13653 (65.1%)	9461 (56.0%)
Smoking			
Never smoker	8522 (22.5%)	7153 (34.1%)	1369 (8.1%)
Former smoker	13652 (36.1%)	8220 (39.2%)	5432 (32.1%)
Current smoker	15692 (41.4%)	5592 (26.7%)	10100 (59.8%)
Pack-years			
Median (Q1, Q3)	23.3 (1.60, 41.5)	9.75 (0, 29.1)	35.8 (21.0, 51.0)
Time since quitting smoking			
0-7 years	3448 (9.1%)	1422 (6.8%)	2026 (12.0%)
8-15 years	3429 (9.1%)	1898 (9.1%)	1531 (9.1%)
16-25 years	3517 (9.3%)	2346 (11.2%)	1171 (6.9%)
>25 years	3258 (8.6%)	2554 (12.2%)	704 (4.2%)

	Lung Cancer	
Exposed Asbestos	1	0
1	7440	7312
0	9461	13653

Exposure to asbestos (binary)

Contingency Table

- Exposure: Asbestos (`ever_asbestos0`). A total of 14752 were ever exposed.
- Outcome: Lung cancer (`status`) occurred in 16901 cases, out of which 13605 (80.5%) were male and 3296 (19.5%) female.

Odds Ratio (OR) and Attributable Fraction (AF).

The odds ratio (OR) is calculated from the contingency table as follows (Equation 1):

$$OR = \frac{a/b}{c/d} = \frac{ad}{bc} \quad (1)$$

where a is the number of cases with exposure, b is the number of controls with exposure, c is the number of cases without exposure, and d is the number of controls without exposure.

```
or <- (7440/7312)/(9461/13653)
```

The crude OR is **1.47**.

The attributable fraction (AF) is calculated from the OR as follows (Equation 2):

$$AF = \frac{OR - 1}{OR} \quad (2)$$

```
af <- (or-1)/or
```

The AF is **0.32**.

Here, the attributable fraction refers specifically to an approximation of the *excess fraction*, interpreted as the excess caseload due to exposure.[1]

Probability of Causation

PoC for an individual case is equal to the difference between risk under exposure (Risk_{exp}) minus baseline risk ($\text{Risk}_{\text{unexp}}$), divided by Risk_{exp} (Equation 3), which can be re-expressed as relative risks (RR) (Equation 4).[2]

$$PoC = \frac{Risk_{exp} - Risk_{unexp}}{Risk_{exp}} \quad (3)$$

$$PoC = \frac{RR - 1}{RR} \quad (4)$$

For quantitative exposures, PoC can be estimated using the RR at any given exposure value (x) from the exposure-response relation (Equation 5).[3, 4]

$$PoC(x) = \frac{RR(x) - 1}{RR(x)} \quad (5)$$

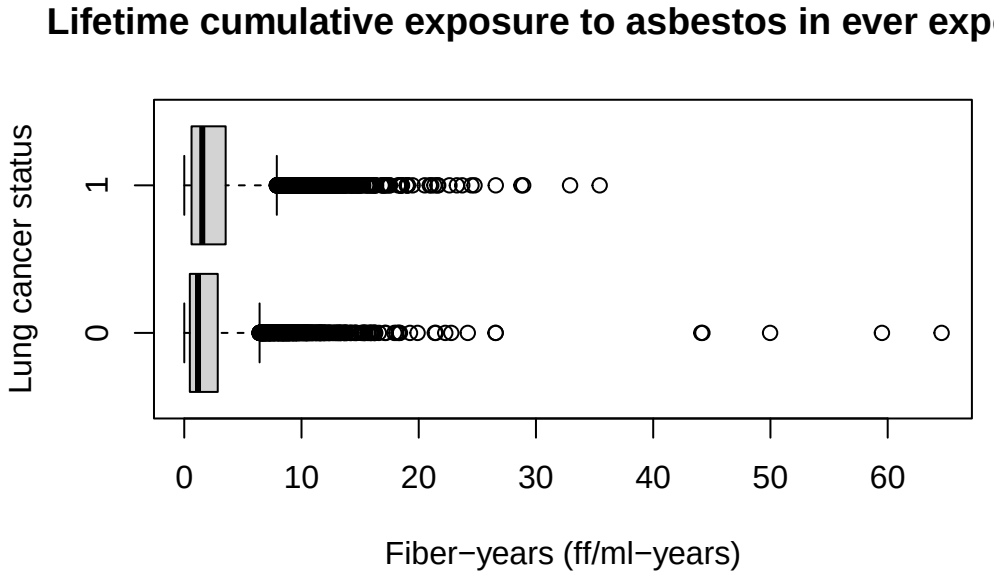
```r
PoCfun <- function(logor) {
  OR <- exp(logor)
  pmax((OR-1)/OR,0)
}
```

Exposure to asbestos as continuous variable

Two exposure measures were used for modelling the continuous exposure to asbestos.

Lifetime cumulative exposure to asbestos (`asbestos_cum0`)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.0000  0.0000  0.9409  0.8267 64.5996
```



Probability of Causation (PoC)

Mixed effects model - cum0

A mixed-effects logistic regression model is used to estimate the PoC, with `asbestos_cum0` as the main explanatory variable, and adjusted for:

- The study source of participants (`study_name`)
- Age category (`agegroup`). The age groups are: <45, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and > 74 years.
- Sex (`sex`), female and male.
- Smoking (`packyrs`), cigarette pack-years.
- Time since smoking cessation (`time_quit`). The categories are: current smokers; stopping smoking 2–7, 8–15, 16–25, and >= 26 years before interview/diagnosis; and never-smokers.

Adding random effects with a random intercept for each study source (`study_name`) and random slopes for the exposure (`asbestos_cum0`) within each study source.

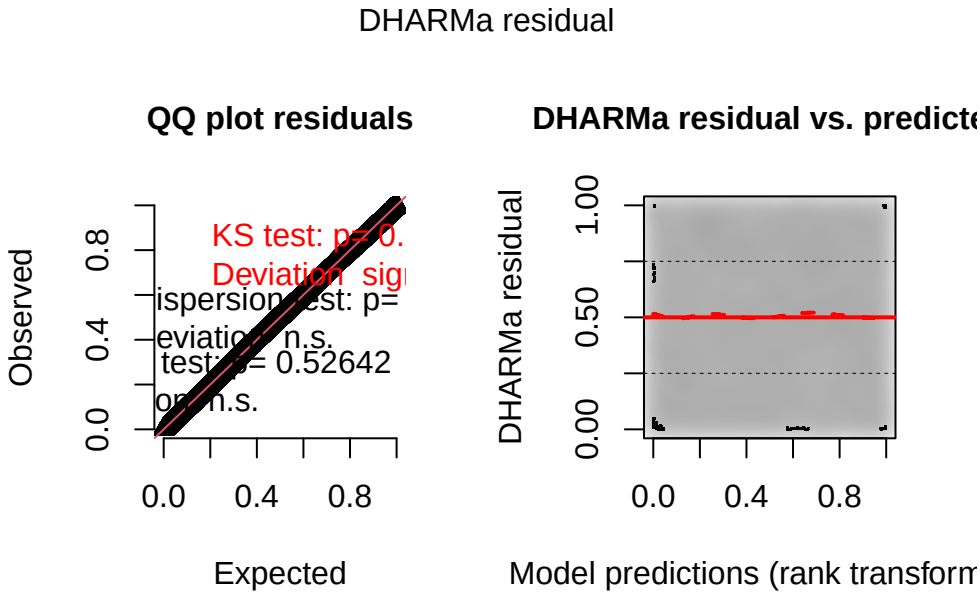
```
mixed_model_cum0 <- glmer(
  status ~ asbestos_cum0 + agegroup + sex + packyrs + time_quit +
    (1 + asbestos_cum0 | study_name),
  data = df,
  family = binomial,
  control = glmerControl(optimizer = "bobyqa"),
  nAGQ = 0
  )
```

```
                GVIF Df GVIF^(1/(2*Df))
asbestos_cum0 1.011428  1        1.005698
agegroup      1.188300  7        1.012399
sex           1.224947  1        1.106773
packyrs       5.535367  1        2.352736
time_quit     5.947026  5        1.195171
```

Asbestos Exposure-Response in SYNERGY

Model	Risk Increase per Fibre-year		Min Exposure for 50% PoC (fibre-years)[1]		Cases per 10,000 above 50% PoC in SYNERGY	
	Estimate	95% Prediction Interval	Point Estimate	Presumably Plausible	Point Estimate	Presumably Plausible
Population-average	6.8%	-3% ; 17.5%	10.59	4.31	115	845

[1]Probability of Causation (PoC)




```
[1] "I² for mixed_model_cum0 (intercept): 3.83 %"
```

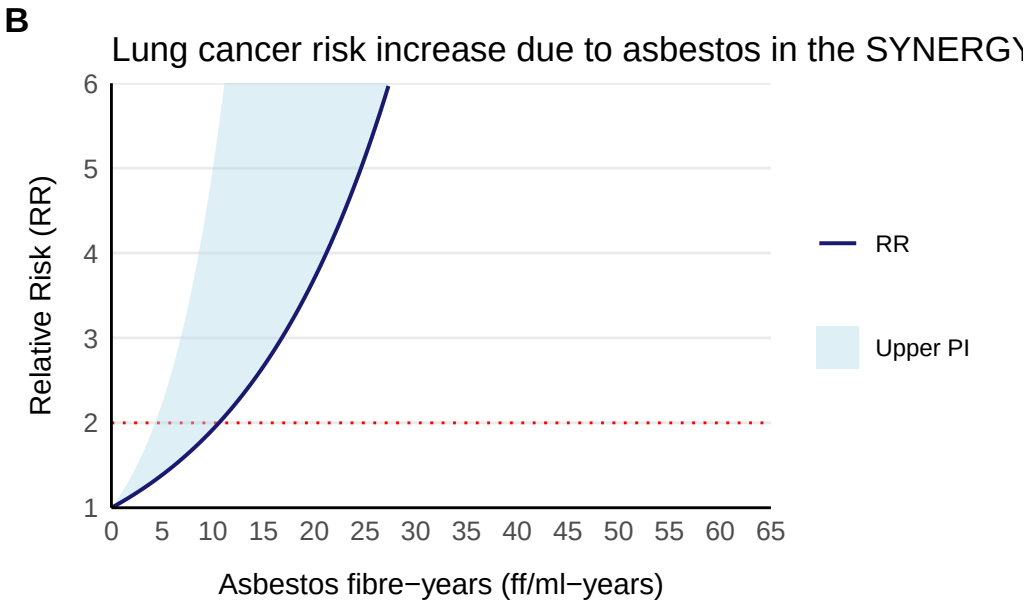
```
[1] "I² for asbestos slopes per study: 0.07 %"
```

The coefficient for exposure (b) is 0.0654 which corresponds to an increase in lung cancer risk of 6.76% per every additional fibre-years.

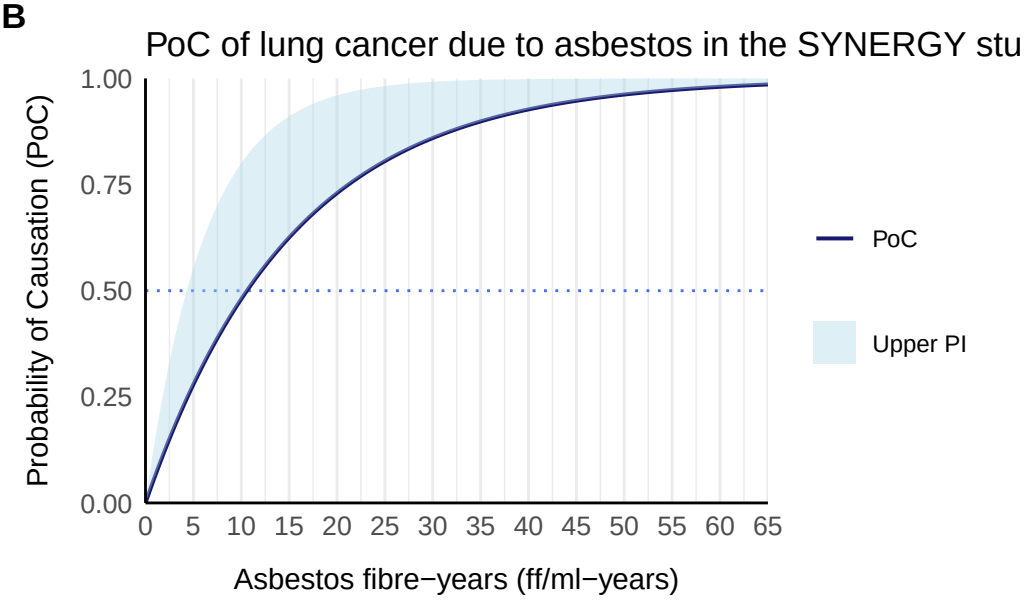
Out of those ever exposed to asbestos, the summary of PoC with this approach is as follows:

Scenario	N	mean	sd	min	q25	median	q75	max
PoC	14752	0.13	0.13	0	0.04	0.08	0.19	0.99
PoC_LB	14752	0.00	0.00	0	0.00	0.00	0.00	0.00
PoC_UB	14752	0.27	0.23	0	0.08	0.19	0.40	1.00

Plot of PoC values with the upper bound of the 95% prediction interval, according to cumulative asbestos exposure values.



B


PoC of lung cancer due to asbestos in the SYNERGY stu
Probability of Causation (PoC)
1.00
0.75
0.50
0.25
0.00
0 5 10 15 20 25 30 35 40 45 50 55 60 65
Asbestos fibre−years (ff/ml−years)
PoC
Upper PI

References

1 Rothman KJ, VanderWeele TJ, Lash TL. Measures of Effect and Measures of Association. In: Lash TL, VanderWeele TJ, Haneuse S, et al., eds. *Modern epidemiology* Fourth edition. Philadelphia: Wolters Kluwer; 2021. p. 79–104.

2 Tian J, Pearl J. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence* 2000; 28: 287–313.

3 Armstrong B, Thériault G. Compensating lung cancer patients occupationally exposed to coal tar pitch volatiles. *Occupational and Environmental Medicine* 1996; 53: 160–167.

4 Siemiatycki J, Karp I, Sylvestre M-P, *et al.* 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: e60–e66.

Package References

For specific information on the operating system, R version, and R package versions used, please refer to the **R/session** folder in the GitHub repository.

- Auguie B (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. doi:10.32614/CRAN.package.gridExtra https://doi.org/10.32614/CRAN.package.gridExtra, R package version 2.3, https://CRAN.R-project.org/package=gridExtra.
- Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01 https://doi.org/10.18637/jss.v067.i01.
- Bates D, Maechler M, Jagan M (2025). *Matrix: Sparse and Dense Matrix Classes and Methods*. doi:10.32614/CRAN.package.Matrix https://doi.org/10.32614/CRAN.package.Matrix, R package version 1.7-3, https://CRAN.R-project.org/package=Matrix.
- Csárdi G, Hester J (2025). *pak: Another Approach to Package Installation*. doi:10.32614/CRAN.package.pak https://doi.org/10.32614/CRAN.package.pak, R package version 0.9.0, https://CRAN.R-project.org/package=pak.
- Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/.
- Fox J, Weisberg S, Price B (2022). *carData: Companion to Applied Regression Data Sets*. doi:10.32614/CRAN.package.carData https://doi.org/10.32614/CRAN.package.carData, R package version 3.0-5, https://CRAN.R-project.org/package=carData.
- Gohel D, Skintzos P (2025). *flextable: Functions for Tabular Reporting*. doi:10.32614/CRAN.package.flexta https://doi.org/10.32614/CRAN.package.flextable, R package version 0.9.9, https://CRAN.R-project.org/package=flextable.
- Grolemund G, Wickham H (2011). "Dates and Times Made Easy with lubridate." *Journal of Statistical Software*, *40*(3), 1-25. https://www.jstatsoft.org/v40/i03/.
- Hartig F (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. doi:10.32614/CRAN.package.DHARMa https://doi.org/10.32614/CRAN.package.DHARMa, R package version 0.4.7, https://CRAN.R-project.org/package=DHARMa.
- Iannone R, Cheng J, Schloerke B, Hughes E, Lauer A, Seo J, Brevoort K, Roy O (2025). *gt: Easily Create Presentation-Ready Display Tables*. doi:10.32614/CRAN.package.gt https://doi.org/10.32614/CRAN.package.gt, R package version 1.0.0, https://CRAN.R-project.org/package=gt.
- Lüdecke D (2025). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.9.0, https://CRAN.R-project.org/package=sjPlot.
- Makowski D, Lüdecke D, Patil I, Thériault R, Ben-Shachar M, Wiernik B (2023). "Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption." *CRAN*. https://easystats.github.io/report/.
- Müller K, Wickham H (2025). *tibble: Simple Data Frames*. doi:10.32614/CRAN.package.tibble https://doi.org/10.32614/CRAN.package.tibble, R package version 3.3.0, https:

//CRAN.R-project.org/package=tibble.
- R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
- Rich B (2023). *table1: Tables of Descriptive Statistics in HTML*. doi:10.32614/CRAN.package.table1 https://doi.org/10.32614/CRAN.package.table1, R package version 1.4.3, https://CRAN.R-project.org/package=table1.
- Rinker TW, Kurkiewicz D (2018). *pacman: Package Management for R*. version 0.5.0, http://github.com/trinker/pacman.
- Sera F, Armstrong B, Blangiardo M, Gasparrini A (2019). "An extended mixed-effects framework for meta-analysis." *Statistics in Medicine*, DOI: 10.1002/sim.8362. http://www.ag-myresearch.com/2019_sera_statmed.html.
- Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.
- Wickham H (2023). *conflicted: An Alternative Conflict Resolution Strategy*. doi:10.32614/CRAN.package.conflicted https://doi.org/10.32614/CRAN.package.conflicted, R package version 1.2.0, https://CRAN.R-project.org/package=conflicted.
- Wickham H (2023). *forcats: Tools for Working with Categorical Variables (Factors)*. doi:10.32614/CRAN.package.forcats https://doi.org/10.32614/CRAN.package.forcats, R package version 1.0.0, https://CRAN.R-project.org/package=forcats.
- Wickham H (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. doi:10.32614/CRAN.package.stringr https://doi.org/10.32614/CRAN.package.stringr, R package version 1.5.1, https://CRAN.R-project.org/package=stringr.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686 https://doi.org/10.21105/joss.01686.
- Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. doi:10.32614/CRAN.package.dplyr https://doi.org/10.32614/CRAN.package.dplyr, R package version 1.1.4, https://CRAN.R-project.org/package=dplyr.
- Wickham H, Henry L (2025). *purrr: Functional Programming Tools*. doi:10.32614/CRAN.package.purrr https://doi.org/10.32614/CRAN.package.purrr, R package version 1.1.0, https://CRAN.R-project.org/package=purrr.
- Wickham H, Hester J, Bryan J (2024). *readr: Read Rectangular Text Data*. doi:10.32614/CRAN.package.readr https://doi.org/10.32614/CRAN.package.readr, R package version 2.1.5, https://CRAN.R-project.org/package=readr.
- Wickham H, Vaughan D, Girlich M (2024). *tidyr: Tidy Messy Data*. doi:10.32614/CRAN.package.tidyr https://doi.org/10.32614/CRAN.package.tidyr, R package version 1.3.1, https://CRAN.R-project.org/package=tidyr.