

Polygenic Risk of Lung Cancer is Independent of Smoking Status: Insights from the Michigan Genomics Initiative

S. Pathania¹, A. Bierdz¹, **D. Fortman**¹, R. Talakanti¹, S.L. Kunkel¹, N. Ramnath^{1,2}, L.G. Fritsche¹

¹University of Michigan, Ann Arbor/MI/USA, ²VA Ann Arbor Healthcare System, Ann Arbor/MI/USA

Correspondence: dyfortma@med.umich.edu

OVERVIEW

Integrated framework for LCINS risk stratification

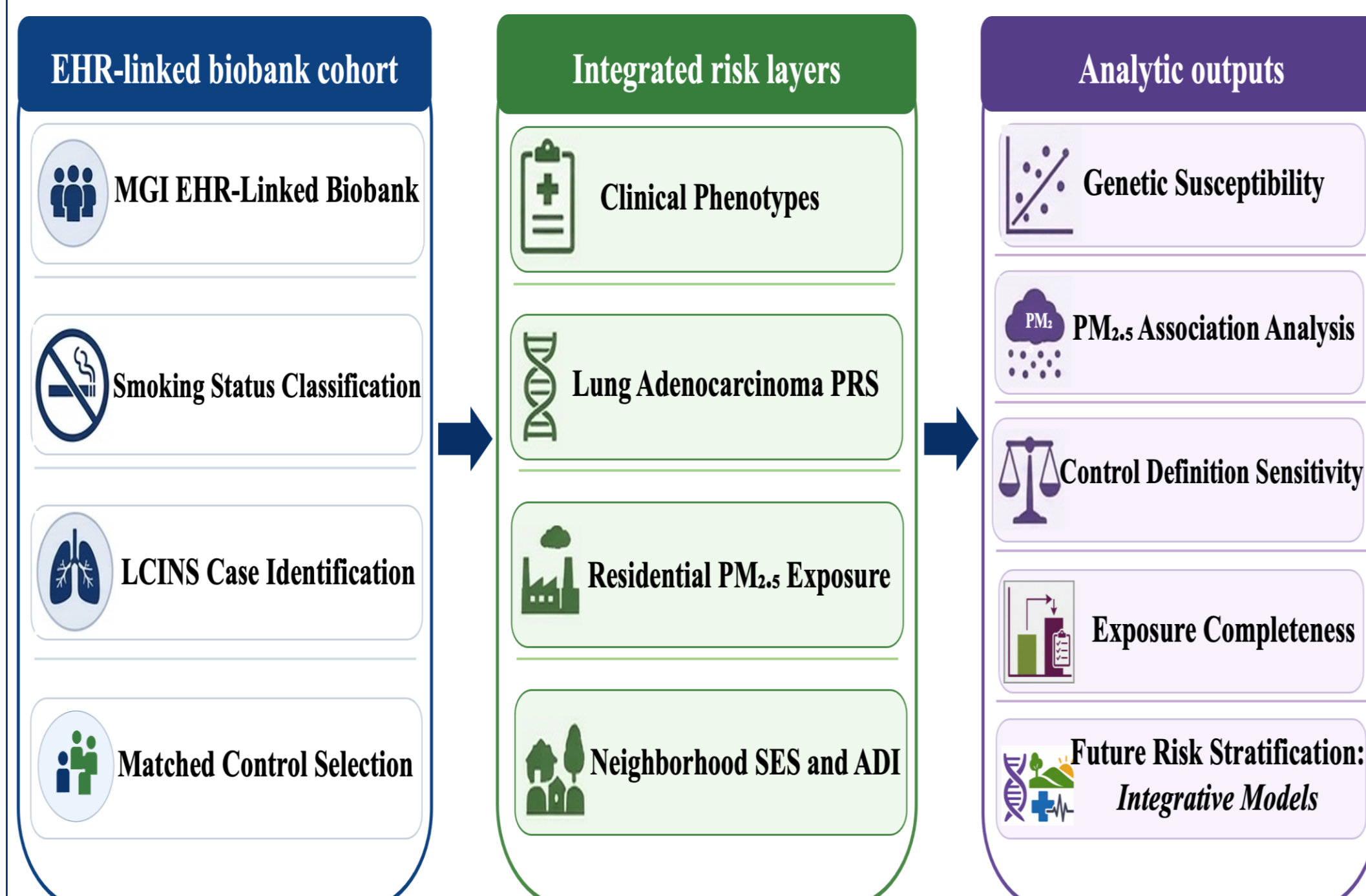


Figure 1. Integrated LCINS Risk Stratification Framework. Schema of cohort selection, multi-domain risk evaluation (PRS, PM_{2.5}, neighborhood socioeconomic factors), and analytic outputs using Michigan Genomics Initiative (MGI) EHR-linked data.

INTRODUCTION & AIM

- Lung cancer remains the leading cause of cancer-related mortality worldwide.
- Although smoking is the dominant risk factor, lung cancer in never-smokers (LCINS) is a clinically distinct subgroup often missed by smoking-based screening criteria.
- LCINS risk may reflect both inherited susceptibility and non-tobacco environmental exposures, including ambient particulate matter 2.5 (PM_{2.5}) and secondhand smoke.
- Polygenic risk scores (PRS) may capture inherited lung cancer susceptibility, but whether PRS associations differ by smoking status remains unclear.
- Figure 1** details the Michigan Genomics Initiative (MGI) strategy for LCINS risk stratification.

Why LCINS?

Lung cancer in never-smokers is an understudied clinical entity. Existing risk models were largely developed for smoking-associated lung cancer.

AIM:

To evaluate associations of PRS and 10-year pre-index PM_{2.5} exposure with lung cancer case status in MGI never- and ever-smoker analyses and to identify priorities for future LCINS risk-stratification models.

METHODS

- The MGI served as the primary EHR-linked case study of lung cancer in never smokers.
- Within smoking-status strata, cases were matched 3:1 to controls on age, sex, race/ethnicity, and index date (**Figure 2**).
- Two control groups were evaluated: lung cancer-free and cancer-free.
- The lung adenocarcinoma PRS was derived from McKay et al. (2017) GWAS summary statistics using MegaPRS.
- Conditional logistic regression evaluated PRS, clinical, and 10-year pre-index PM_{2.5} associations, including joint PM_{2.5} x PRS interaction models.
- Sensitivity analyses assessed restriction to European ancestry, exposure dataset, and exposure-window completeness.

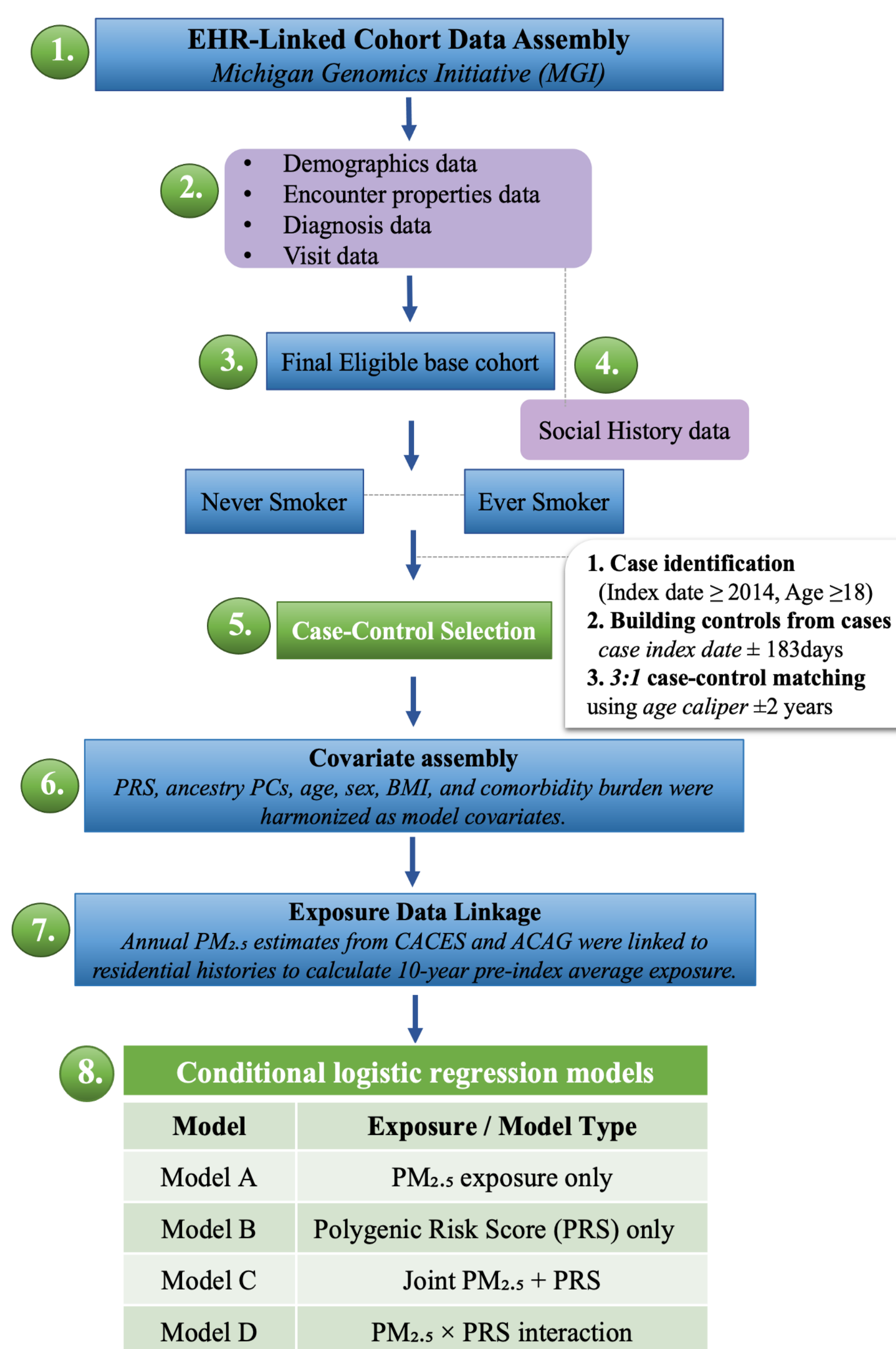


Figure 2. Analytic workflow for EHR-linked lung cancer cohort assembly, residential PM_{2.5} exposure linkage, and PRS-based LCINS risk modeling. Abbreviations: CACES (Center for Air, Climate, and Energy Solutions) and ACAG (Atmospheric Composition Analysis Group).

RESULTS

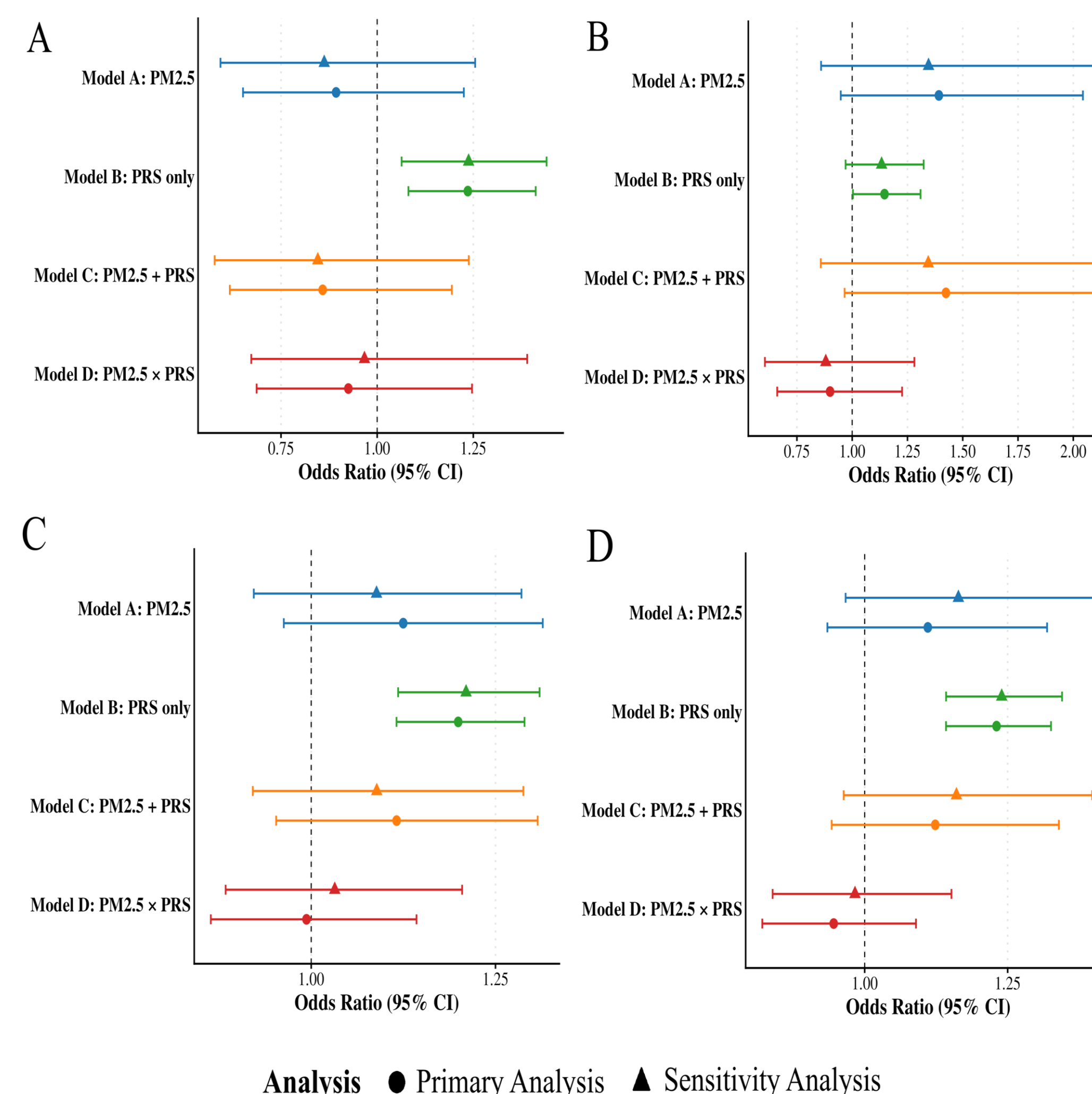


Figure 3. CACES-derived 10-year PM_{2.5}, PRS, and PM_{2.5}-PRS interaction associations with lung cancer risk. Forest plots show odds ratios and 95% confidence intervals from matched conditional logistic regression models in MGI analyses. Results are shown for never-smokers using lung cancer-free controls (A) and cancer-free controls (B), and for ever-smokers using lung cancer-free controls (C) and cancer-free controls (D).

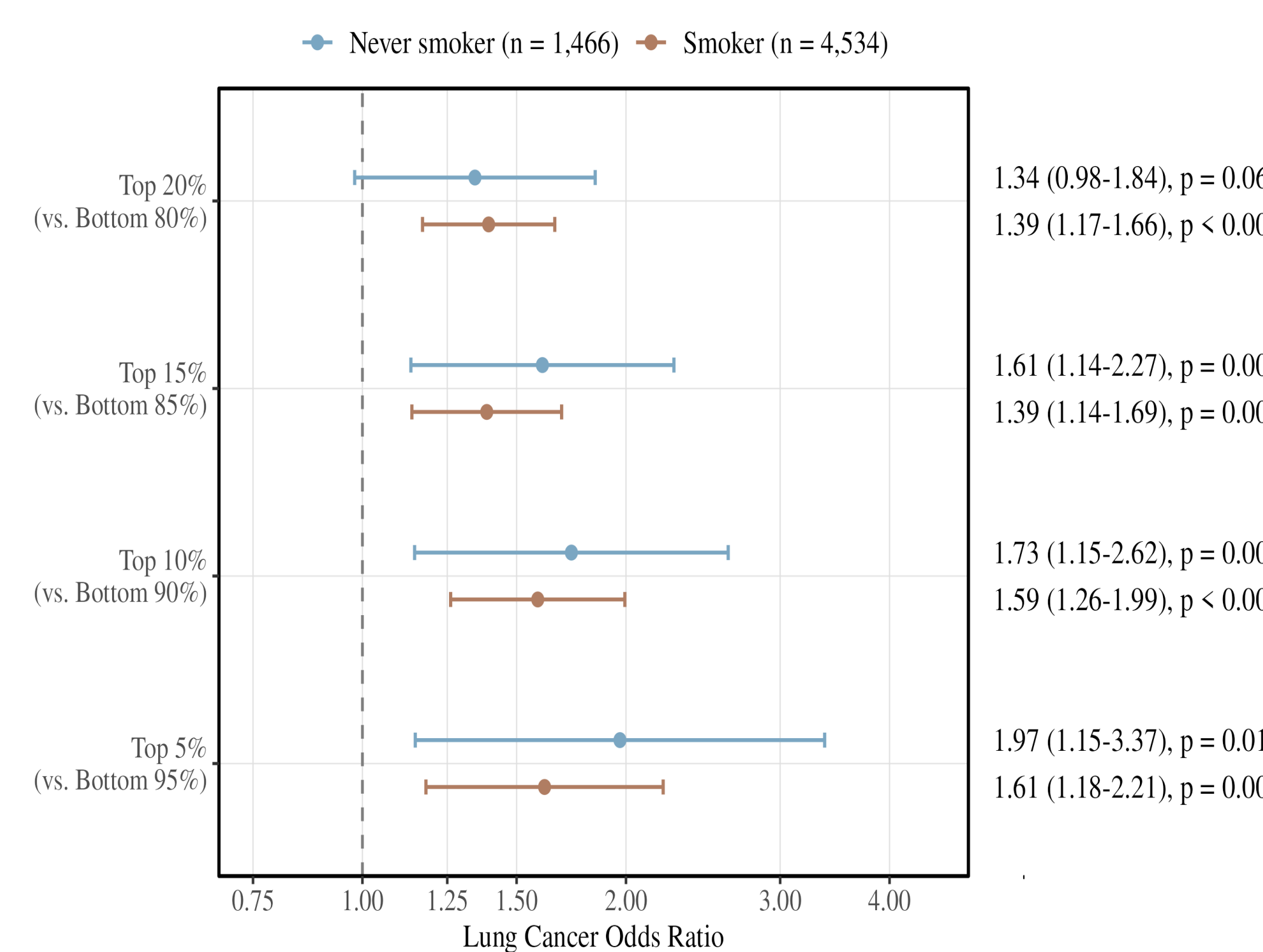


Figure 4. PRS risk stratification by top percentiles. Associations between PRS and lung cancer risk in ever-smokers and never-smokers from matched case-control analyses using cancer-free controls. ORs and 95% CIs were estimated using conditional logistic regression stratified by matched set. Top 20%, 15%, 10%, and 5% PRS groups were compared with the remaining lower-risk participants.

DISCUSSION

- PRS associations were consistent across control definitions and sensitivity analyses, independent of smoking status (**Figures 3 – 4**).
- PM_{2.5} estimates were more variable, reflecting sensitivity to control definitions, calendar eligibility, and person-year completeness (**Figure 3**).
- Exposure product affected results: ACAG provided global coverage, whereas CACES offered more US-specific exposure estimates.
- Control definitions were clinically distinct and appeared to affect PM_{2.5} estimates more than PRS estimates.

CONCLUSION

- PRS showed a consistent, smoking-independent association with LCINS, supporting its integration into multidomain lung cancer risk models.
- PM_{2.5} estimates were more sensitive to exposure-data gaps, supporting the need for improved exposure infrastructure and larger validation cohorts.

FUTURE DIRECTIONS

- Develop integrated lung cancer risk models combining genetic, environmental, clinical, family history, neighborhood factors.
- Improve PM_{2.5} coverage, residential histories, geocoding, and completeness reporting.
- Pool multi-institutional LCINS cohorts to increase statistical power, enable subgroup analyses, and externally validate findings.

References

- Sigel, et al. *CA Cancer J Clin*, 2024.
- LoPiccolo, et al. *Nat Rev Clin Oncol*, 2024.
- Sun, et al. *Nat Rev Cancer*, 2007.
- Arrieta, et al. *Crit Rev Oncol Hematol*, 2024.
- McKay, et al. *Nat Genet*, 2017.
- Wang, et al. *Math Prob Engin*, 2020.
- Shen, et al. *American Chem Society*, 2024
- Zhang, et al. *Nat Commun*, 2021