Effects of Confounder Control and Causal Modelling in MAPLE, ELAPSE, and Medicare Cohorts

HEI Annual Conference

May 2021

Danielle Braun on behalf of the MAPLE, ELAPSE, and Harvard teams

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Introduction

Goal: all three studies assess health effects of long-term exposure to low levels of ambient air pollution, $PM_{2.5}$.

- Ideally you would randomize patients to low/high air pollution, but not feasible.
- Observational studies (e.g. claims data) have limitations due to lack of randomization.
  - Factors that are associated both with both exposures and health outcomes (e.g. SES-related factors) may **confound** exposure comparisons.

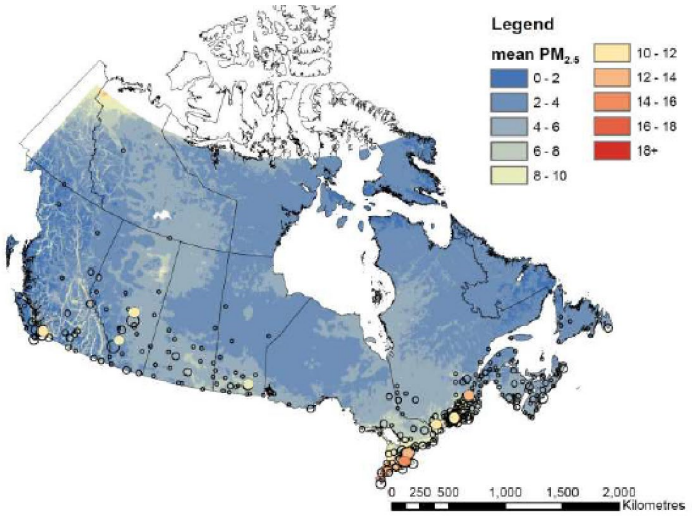
Challenges

1) Unmeasured confounders
   - Individual level covariates may confound exposure comparisons
   - Some individual level covariates are missing across all three studies
   - Each team approached this slightly differently

2) Adjusting for measured confounders
   - Confounders can be included as linear terms in statistical models→ could be sensitive to model misspecification

MAPLE Overview

- **Canadian Census Health and Environment Cohort (CanCHEC):** three census-based cohorts (1991, 1996, 2001) of 7.1 million respondents, 128 million person-years with individual- and contextual-level covariates
- **Canadian Community Health Survey (CCHS):** additionally has information about health behaviors (450,900 respondents and 4.4 million person-years)
- Age ≥25 y at baseline, censored at 89 y
- Non-accidental mortality



Individual covariates: income quintile, visible minority status, Indigenous identity, labor force status, marital status, educational attainment

Contextual covariates: community size, airshed, urban form, neighborhood marginalization (instability, deprivation, dependency ethnic concentration)

Behavioral covariates: **smoking behaviors, alcohol consumption, body mass index category, fruit and vegetable consumption, exercise behaviors**

MAPLE Covariate Adjustment

- Relationship between **$PM_{2.5}$** and **non-accidental mortality** was robust to the addition of behavioral covariates in the CCHS

CanCHEC (1991, 1996, 2001)	HR	Lower 95%CI	Upper 95%CI
Unadjusted model (stratified by age, sex, cohort)	1.20	1.013	1.027
+ Individual level covariates	1.063	1.055	1.070
+ Contextual covariates (Full model)	1.084	1.073	1.096
mCCHS	**HR**	**Lower 95%CI**	**Upper 95%CI**
Unadjusted model (stratified by age, sex, cohort)	1.002	0.966	1.039
+ Individual level covariates	1.034	0.997	1.073
+ Contextual covariates	1.123	1.056	1.194
+ Behavioural covariates (Full model)	1.086	1.021	1.155

HR estimates per 10 $\mu g/m^3$ increase in $PM_{2.5}$

MAPLE Indirect Adjustment

- Use an indirect adjustment via partitioned regression to adjust for unmeasured covariates (Shin et al., 2014).
- Method uses a representative ancillary dataset to estimate the association between variables missing in a primary dataset with the complete set of variables of the ancillary dataset to produce an adjusted risk estimate for the variable in question.
- Used the association between fine particle air pollution ($PM_{2.5}$) with mortality in the 2001 CanCHEC (N = 2.4 million, 10-years follow-up) as primary dataset.
- Used the 2001 cycle of the CCHS (N = 80,630) as the ancillary matching dataset that contained confounding risk factor information not available in CanCHEC (e.g., smoking).

MAPLE Indirect Adjustment: Results (Erickson et al, 2019)

- Results from the CCHS cohort analysis and the indirect adjustment suggest that behavioral covariates (e.g., smoking and diet) only slightly confounded the $PM_{2.5}$–mortality association.
- Based on sensitivity analyses in the 2001 CanCHEC cohort→ missing data on behavioral covariates were unlikely to significantly confound the $PM_{2.5}$–mortality relationship in the Canadian population.

Non-Accidental Mortality

	Unadjusted	Adjusted**	Weighted Adjustment
Model 1	1.103 (1.084 – 1.123)	1.135 (1.110 - 1.160)	1.131 (1.106 - 1.156)
mCCHS (Model 1)	1.177 (1.073 - 1.291)	1.202 (1.095 - 1.319)	
Model 2	1.201 (1.171 - 1.231)	1.161 (1.126 - 1.197)	1.178 (1.142 - 1.214)
mCCHS (Model 2)	1.163 (1.002 - 1.348)	1.136 (0.979 - 1.318)	

Supplemental Table s3: Indirect adjustment of missing risk factors on $PM_{2.5}$ hazard ratios (HR, 95% CI)* for different causes of mortality with comparison to the CCHS-mortality cohort

Model 1: Individual covariate model - stratified by 5-year age-sex groups, adjusted marital status, visible minority, Aboriginal identity, employment, income quintile, education, Model 2: Individual + ecological covariate model - Model 1, plus Can-Marg Index, Community Size, Airshed

mCCHS: equivalent model using CCHS-mortality linked cohort, adjusted model includes smoking, alcohol use, exercise, diet

* HR estimates per 10 µg/m^3 increase in $PM_{2.5}$ **Model 1 and 2 are indirectly adjusted for smoking, alcohol use, exercise, diet

Weighted adjustment: indirect adjustment uses sample weight (W-matrix)

ELAPSE Overview

- Large administrative cohorts from seven countries in Europe (N = 28 million)
- Age ≥30 y at baseline
- Non-accidental mortality

ELAPSE Covariate Adjustment

- Specified three models to evaluate association between $PM_{2.5}$ and non-accidental mortality;
  - **Model 1:** included only age (time axis), sex (as strata), and calendar year(s) of enrollment.
  - **Model 2:** added individual level variables available within each cohort.
  - **Model 3:** added to the model 2 area-level socio-economic status (SES) variables.

ELAPSE Covariate Adjustment

Non-Accidental Mortality

	Model 1	Model 2	Model 3
Pollutant	HR (95% CI)	HR (95% CI)	HR (95% CI)
Belgian cohort (N=5,474,548)			
PM2.5	0.960 (0.946, 0.974)	0.968 (0.956, 0.981)	1.023 (1.011, 1.035)
Danish cohort (N=3,083,235)			
PM2.5	1.221 (1.185, 1.257)	1.216 (1.190, 1.243)	1.141 (1.118, 1.164)
Dutch cohort (N=10,465,727)			
PM2.5	1.062 (1.034, 1.090)	1.036 (1.013, 1.059)	1.021 (0.999, 1.044)
English cohort (N=1,368,740)			
PM2.5	1.114 (1.090, 1.137)	1.082 (1.059, 1.105)	1.023 (1.001, 1.045)
Norwegian cohort (N=2,309,001)			
PM2.5	1.034 (1.018, 1.050)	1.071 (1.061, 1.082)	1.076 (1.066, 1.086)
Rome cohort (N=1,263,712)			
PM2.5	0.964 (0.888, 1.047)	1.019 (0.969, 1.071)	1.066 (1.033, 1.099)
Swiss cohort (N=4,188,175)			
PM2.5	0.979 (0.966, 0.993)	1.007 (0.997, 1.018)	1.026 (1.015, 1.038)

HR estimates per 5 μg/m^3 increase in $PM_{2.5}$

- In all cohorts, HRs were sensitive to more complete adjustment for potential confounders.
- Adjustment for individual and especially area-level confounders increased the HRs in some cohorts (e.g. Rome, Swiss, Norwegian) and decreased HRs in other cohorts (e.g. Dutch, English).

ELAPSE Indirect Adjustment

- Applied the indirect adjustment method proposed by Shin et al., 2014.
- Applied the Shin method for natural cause mortality and the key potential confounders smoking status and BMI.
- In the Netherlands, Switzerland, Norway, Belgium and Denmark data from large population surveys on variables such as smoking, alcohol use, exercise and BMI were available.
  - English cohort has individual data on smoking and BMI in full cohort → no indirect adjustment needed.

ELAPSE Indirect Adjustment: Results

- In the Dutch, Swiss and Norwegian cohort, HRs were attenuated but remained (borderline) significant.
- In the Rome and Belgian cohort, HRs increased after indirect adjustment.

Table 22 Hazard ratios for associations between PM2.5 and NO_2 and **natural-cause mortality** in six administrative cohorts: indirect adjustment, adjustment for area-level disease

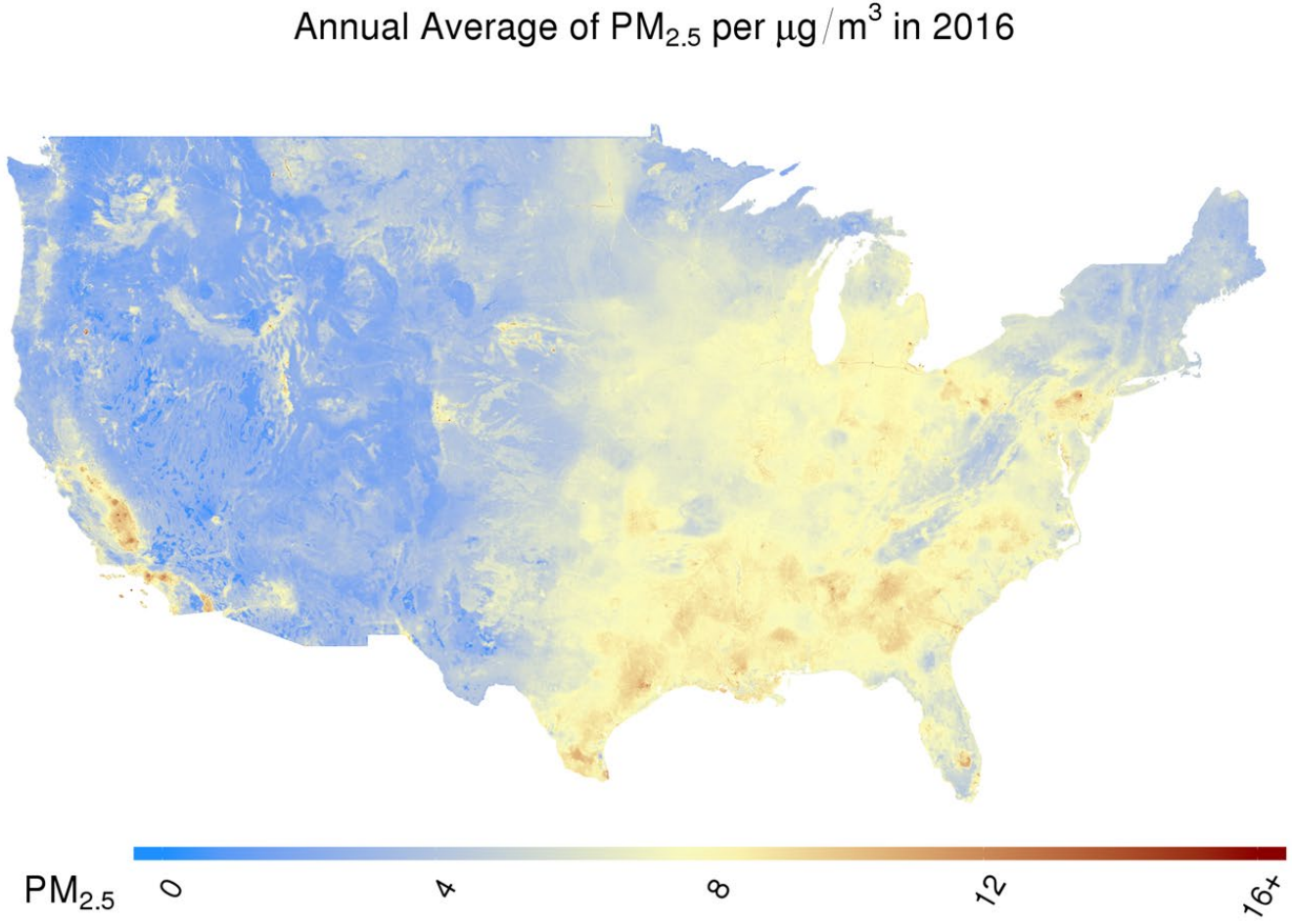
	Cohort	HR, main Model	HR with indirect adjustment	HR, main Model, reduced population area level adjustment	HR with area-level lung cancer, COPD, diabetes
PM2.5	Belgian	1.023 (1.011, 1.035)	1.049 (1.036, 1.062)	NA	1.029 (1.017, 1.041)[1]
	Danish	1.141 (1.118, 1.164)	1.118 (1.095, 1.140)	NA	1.143 (1.120, 1.167)[2]
	Dutch	1.021 (0.999, 1.044)	1.015 (0.993, 1.038)	NA	1.007 (0.985, 1.030)[3]
	Norwegian	1.076 (1.066, 1.086)	1.055 (1.045, 1.065)	1.095 (1.080, 1.109)[4]	1.081 (1.067, 1.095)[1]
	Rome	1.066 (1.033, 1.099)	1.111 (1.080, 1.142)	NA	1.041 (1.009, 1.075)[3]
	Swiss	1.026 (1.015, 1.038)	1.015 (1.003, 1.027)	NA	NA

HR estimates per 5 $\mu g/m^3$ increase in $PM_{2.5}$

Harvard Overview

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

- Medicare enrollees (N = 68.5 million)
- Age ≥65 y
- All-cause mortality

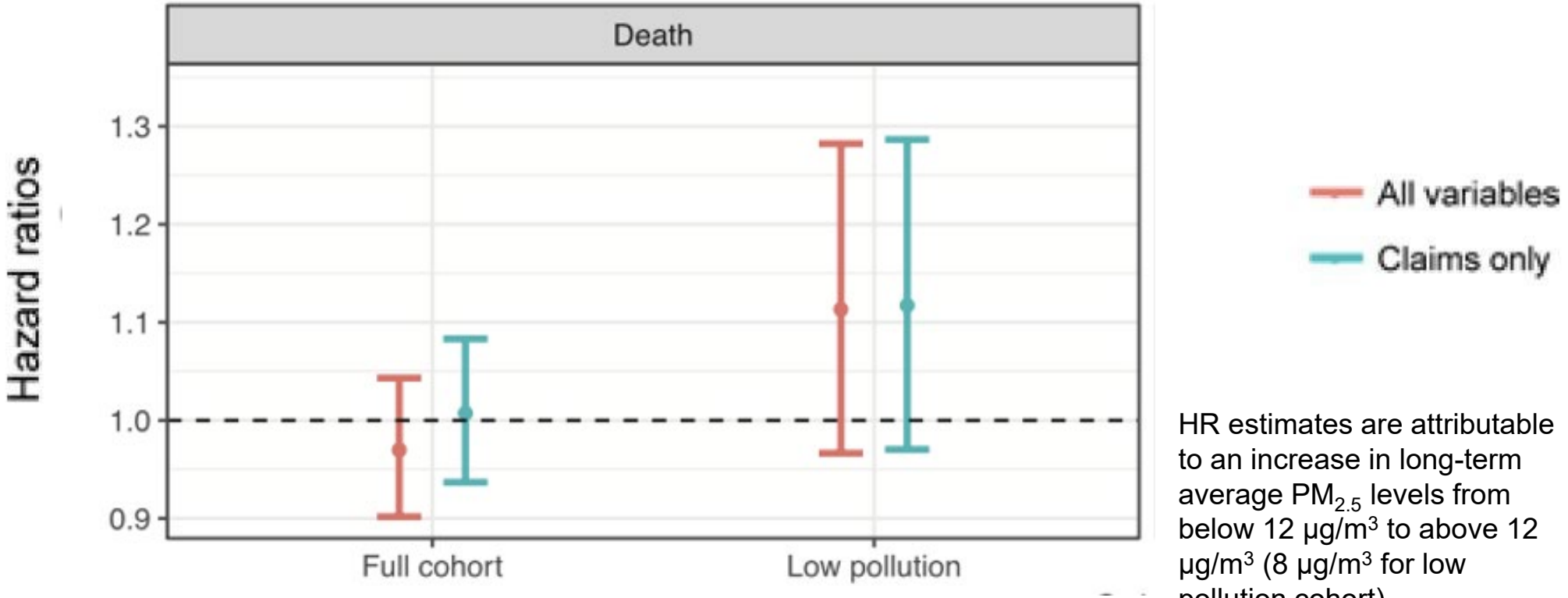


HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Harvard Covariate Adjustment

- Medicare data contain little information about individual level covariates.
- Obtained data from Medicare Current Beneficiary Survey (MCBS), which is an annual phone survey of a nationally representative sample of Medicare beneficiaries
  - contains information on more than 150 potential individual confounders: individual risk factors (e.g., smoking, body mass index [BMI], and income)
- Analyzed data for a cohort of ~32,000 beneficiaries from the MCBS-Medicare linked dataset.

Harvard Covariate Adjustment: Results (Makar et al, 2017)



HR estimates are attributable to an increase in long-term average $PM_{2.5}$ levels from below 12 μg/m³ to above 12 μg/m³ (8 μg/m³ for low pollution cohort)

Figure 4.
Sensitivity to exclusion of MCBS variables: Hazard ratios and 95% confidence intervals based on robust, sandwich variance estimators computed including (red) and excluding (blue) MCBS variables.

Causal Inference Adjustment for Measured Confounders

Causal inference methods are designed to bridge the gap between observational studies and randomized experiments → are more robust to model misspecification

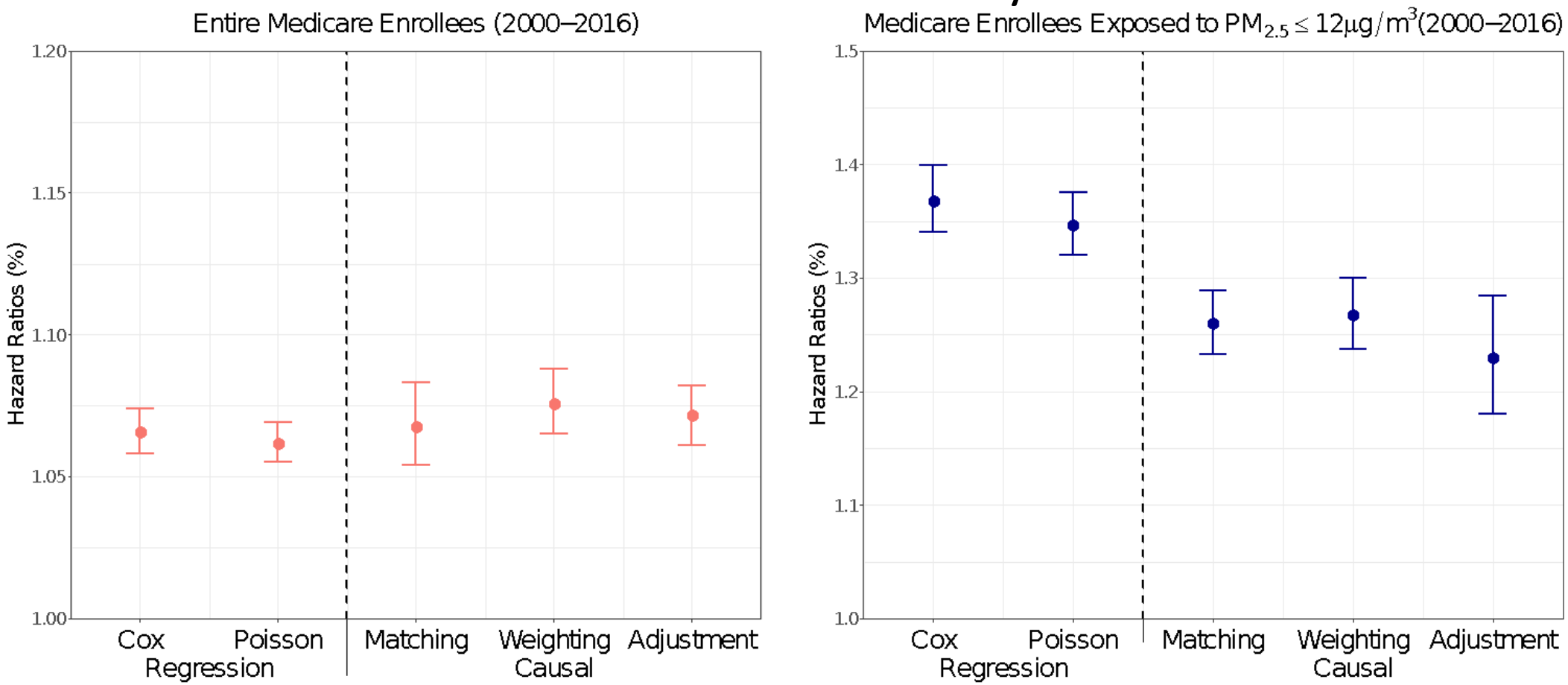
HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Harvard Causal Inference Modeling

- Applied various traditional and causal inference approaches, conducted extensive sensitivity analyses.
- Relied on publicly available data and provide code that allows for reproducibility https://github.com/wxwx1993/National_Causal.

Harvard Causal Inference Modeling: Results (Wu et al, 2020)

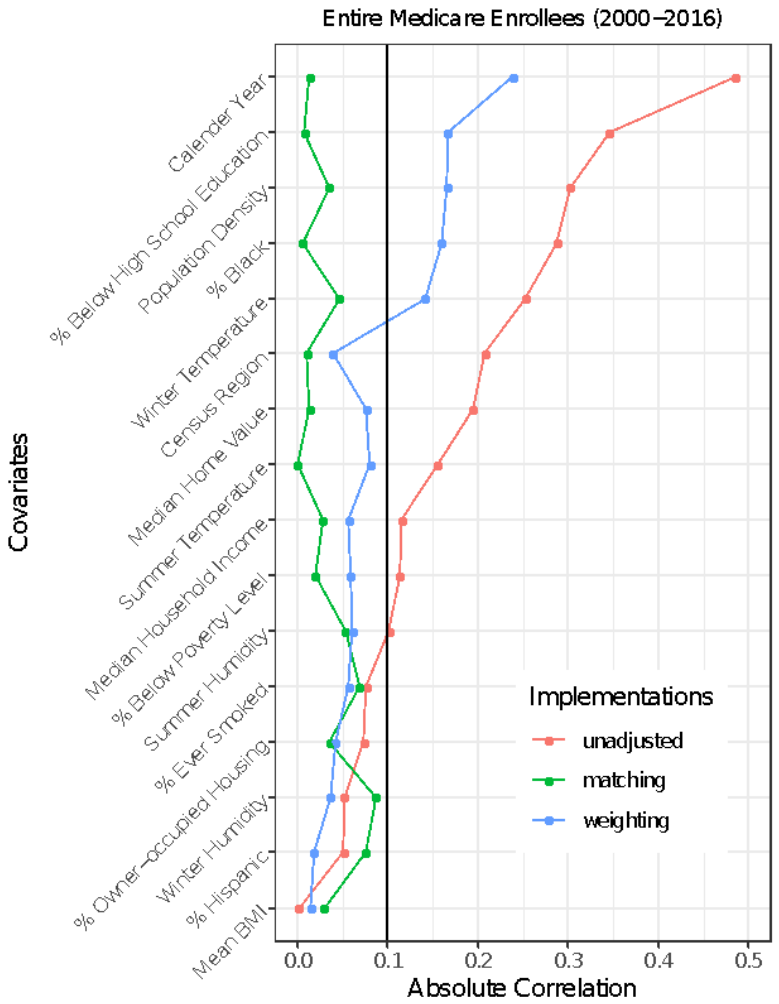
All-cause mortality

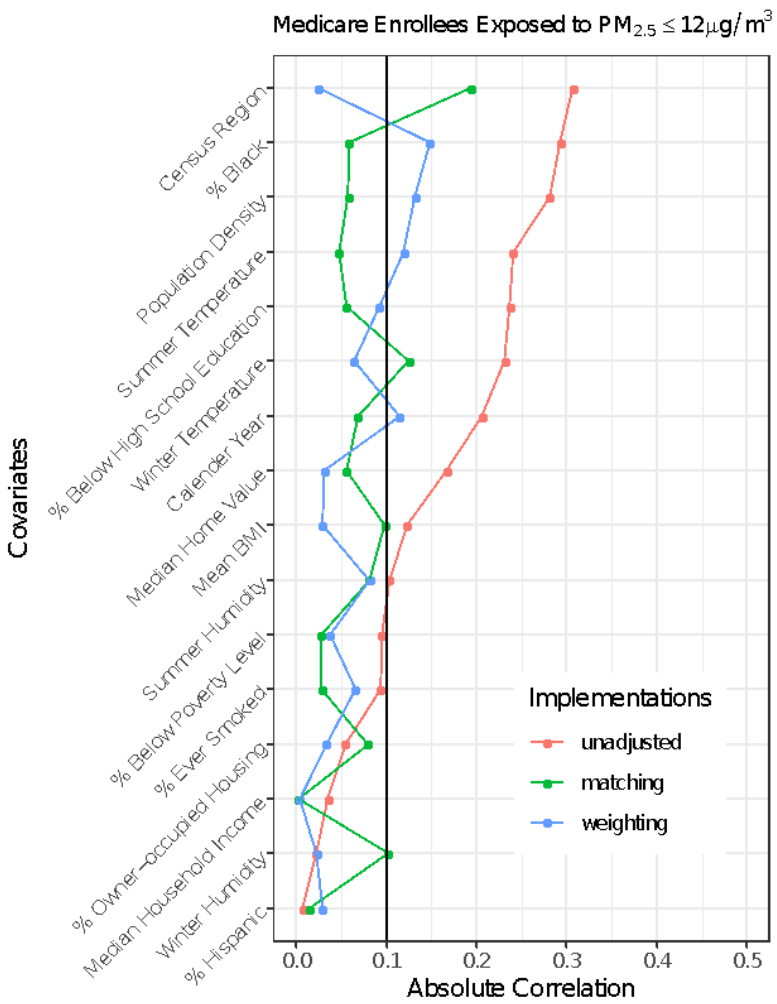


HR estimates per 10 μg/m³ increase in $PM_{2.5}$

Harvard Causal Inference Modeling: Results (Wu et al, 2020)

- Covariate balance used to assess the quality of the causal inference approaches at recovering randomized experiments.
- Evaluate covariate balances for each confounder.





Conclusions

- All three studies; missing data on covariates were unlikely to significantly confound the $PM_{2.5}$–mortality relationship
- In Medicare, five approaches (traditional and casual) lead to similar results→ robust and reproducible evidence on causal link between PM and mortality

MAPLE

ELAPSE

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Ongoing Work

- All three teams are working on harmonizing confounders and models across the datasets
- ELAPSE and MAPLE working on applying causal inference approaches

MAPLE

ELAPSE

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Acknowledgements and Disclaimers

- This work was made possible by the support from the Health Effects Institute
- Research described in this presentation was conducted under contract to the Health Effects Institute (HEI), an organization jointly funded by the United States Environmental Protection Agency (EPA) (Assistance Award No.CR-83467701) and certain motor vehicle and engine manufacturers.

- The contents of this presentation do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers.

Additional Slides