Using linked data for advanced analytics in health research

Spotlight on the Canadian Census Health and Environment Cohorts (CanCHECs)

Tracey Bushnik (tracey.bushnik@statcan.gc.ca)

Delivering insight through data for a better Canada

Statistics Canada Statistique Canada

Canada

What are census health cohorts?

- Created by linking

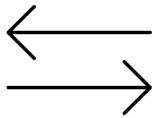
  long-form census records to administrative data

What are census health cohorts?

- Created by linking

long-form census records to administrative data

- Family status
- Education
- Income
- Ethnicity
- Nativity
- Language
- Labour market
- ...



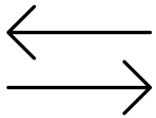
- Mortality
- Cancer
- Hospitalizations
- Ambulatory care
- ...

What are census health cohorts?

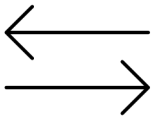
- Created by linking

long-form census records to administrative data

- Family status
- Education
- Income
- Ethnicity
- Nativity
- Language
- Labour market
- ...



- Mortality
- Cancer
- Hospitalizations
- Ambulatory care
- ...



- Tax
- Environmental exposures
- ...

What are census health cohorts?

- Created by linking

  long-form census records to administrative data

  - Family status
  - Education
  - Income
  - Ethnicity
  - Nativity
  - Language
  - Labour market
  - …

  - Mortality
  - Cancer
  - Hospitalizations
  - Ambulatory care
  - …

  - Tax
  - Environmental exposures
  - …

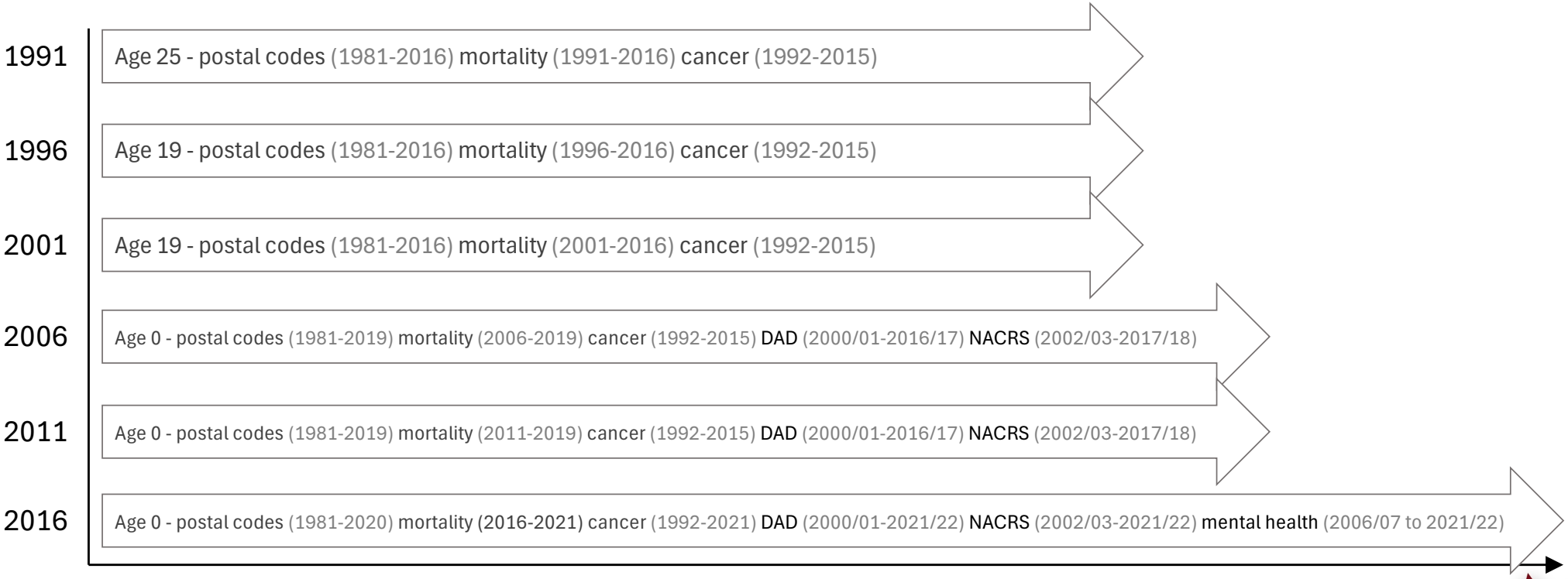
- administrative health outcome data generally lack detailed information about the characteristics of individuals experiencing the outcomes

The CanCHECs

- Canadian Census Health and Environment Cohorts (CanCHECs)
  - Linkage between long-form census respondents for census years 1991, 1996, 2001, 2006 and 2016 (National Household Survey respondents in 2011) with administrative health data and annual mailing address postal codes

The CanCHECs

The CanCHECs



DAD = Discharge Abstract Database, NACRS = National Ambulatory Care Reporting System

The CanCHECs

Mortality

(2023) Mortality inequalities of Black adults in Canada

(2024) Risk factors and inequities in transportation injury and mortality in the CanCHECs

(2022) Industrial air pollutant emissions and mortality from Alzheimer's disease in Canada.

Cancer

(2023) Site-specific cancer incidence by race and immigration status in Canada 2006–2015: a population-based data linkage study

(2022) Assessing geographic and industry-related trends in bladder cancer in Ontario: A population-based study.

Hospitalization

(2024) Creating an 11-year longitudinal substance use harm cohort from linked health and census data to analyze social drivers of health

(2020) Geographic variation in preventable hospitalisations across Canada: a cross-sectional study

Study

PNAS RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES SUSTAINABILITY SCIENCE OPEN ACCESS Check for updates

Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis

Hong Chen[a,b,c,d,1], Matthew Quick[e], Jay S. Kaufman[f,g], Chen Chen[h], Jeffrey C. Kwong[b,c,d,i], Aaron van Donkelaar[j], Jun Meng[j,k], Randall V. Martin[j], JinHee Kim[b,d], Eric Lavigne[l,m], Li Bai[c], Yi Li[f], Michael Tjepkema[e], Tarik Benmarhnia[h], and Richard T. Burnett[a]

Edited by Maureen Cropper, University of Maryland, College Park, College Park, MD; received June 3, 2022; accepted October 17, 2022

Study

- This study applied the g-formula — a method of causal analysis suitable for evaluating interventions using observational data with time-varying air pollution exposures, confounders, and health outcomes — to

Study

- This study applied the g-formula — a method of causal analysis suitable for evaluating interventions using observational data with time-varying air pollution exposures, confounders, and health outcomes — to
- produce nationally representative estimates from the 2006 CanCHEC of

Study

- This study applied the g-formula — a method of causal analysis suitable for evaluating interventions using observational data with time-varying air pollution exposures, confounders, and health outcomes — to
- produce nationally representative estimates from the 2006 CanCHEC of
- the health benefits resulting from ambient fine particulate matter ($PM_{2.5}$) reductions across multiple mitigation strategies, emission sources, and time periods over the course of a decade.

Study

To contextualize the findings, the study:

1. calculated the number of premature deaths that would have been averted if air pollution had decreased

Study

To contextualize the findings, the study:

1. calculated the number of premature deaths that would have been averted if air pollution had decreased
2. derived the number of years of life that would have been gained if air pollution had decreased

Study

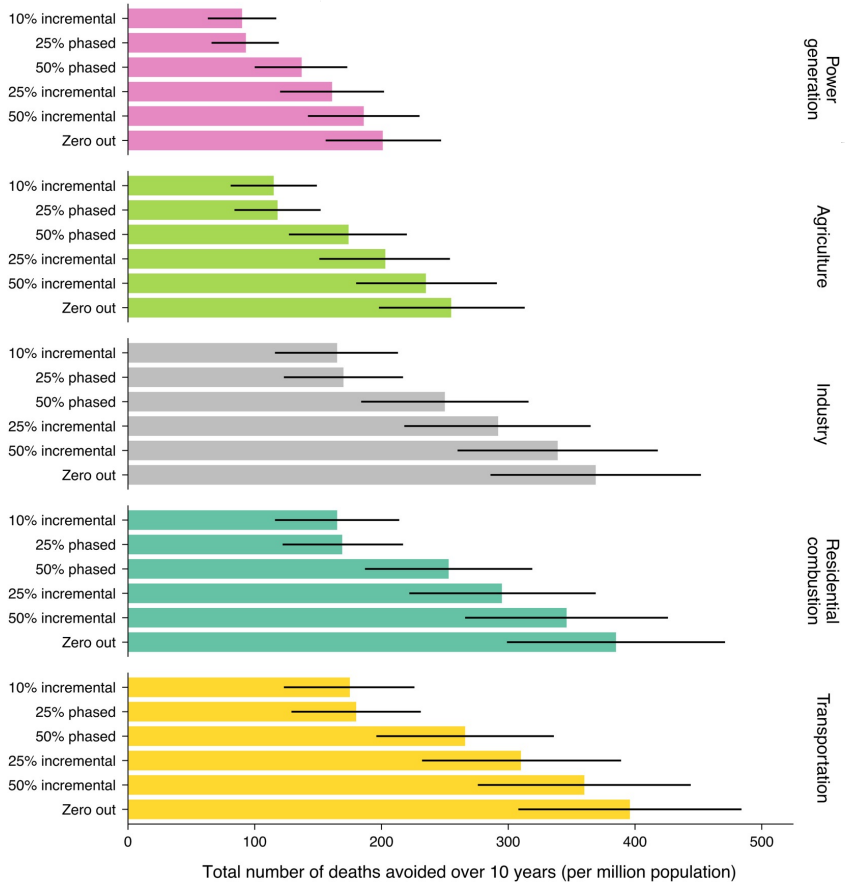
To contextualize the findings, the study:

1. calculated the number of premature deaths that would have been averted if air pollution had decreased
2. derived the number of years of life that would have been gained if air pollution had decreased
3. computed the economic benefit from these interventions using willingness-to-pay metrics that account for both direct and indirect costs

1. Deaths averted

If $PM_{2.5}$ from each source was reduced annually by 10% compared with the natural course:

- 90 fewer deaths (95%CI: 63–117) per million from power generation
- 115 fewer deaths per million from agriculture
- 165 fewer deaths per million from industry
- 165 fewer deaths per million from residential combustion
- 175 fewer deaths per million from transportation

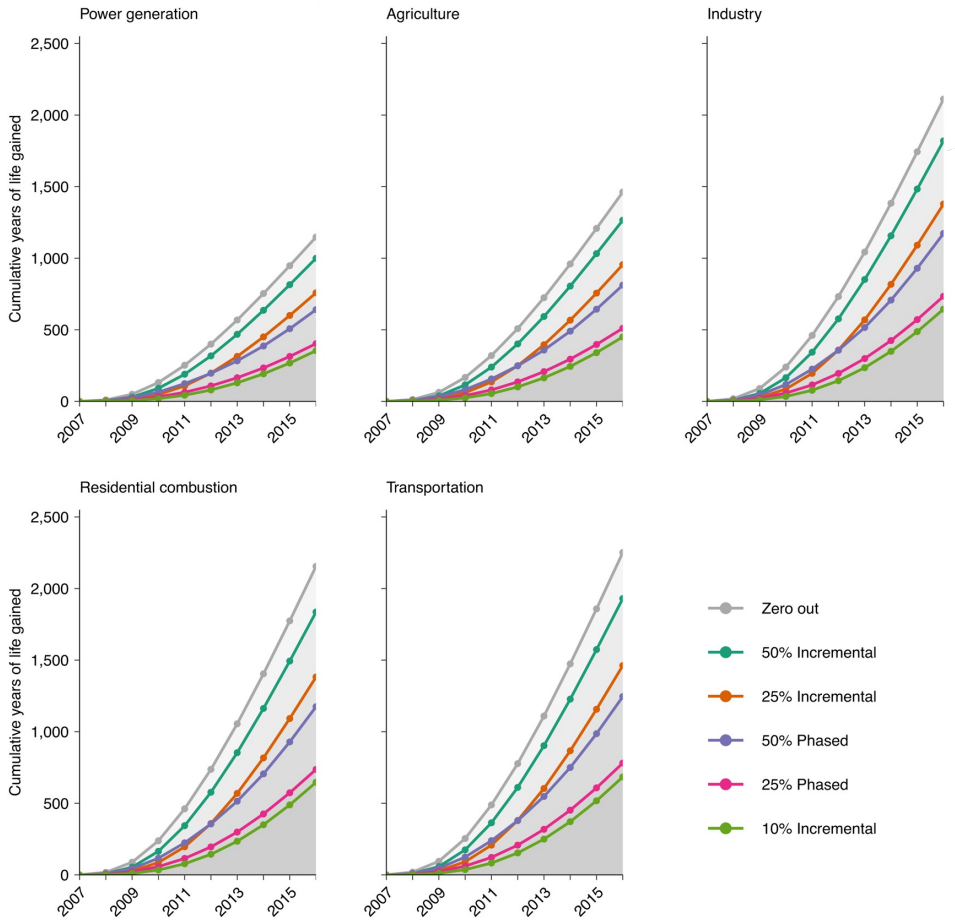


Source: Chen H, Quick M, Kaufman JS et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *PNAS* 2022; 119 (49): e2209490119. https://doi.org/10.1073/pnas.2209490119.

2. Years of life gained

Combined, reducing all five sources by X% would have saved Y life-years per million

X	Y
10%	2,783
25%	5,947
50%	7,865



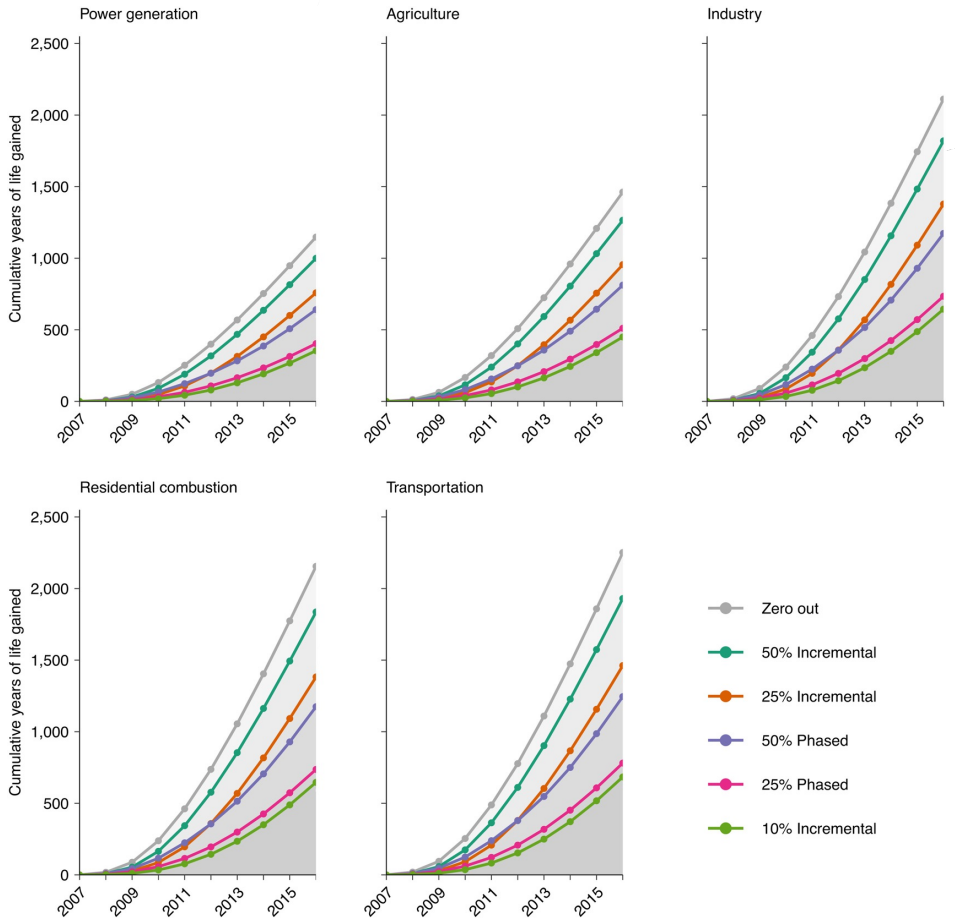
Source: Chen H, Quick M, Kaufman JS et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *PNAS* 2022; 119 (49): e2209490119. https://doi.org/10.1073/pnas.2209490119.

2. Years of life gained

Combined, reducing all five sources by X% would have saved Y life-years per million

X	Y
10%	2,783
25%	5,947
50%	7,865

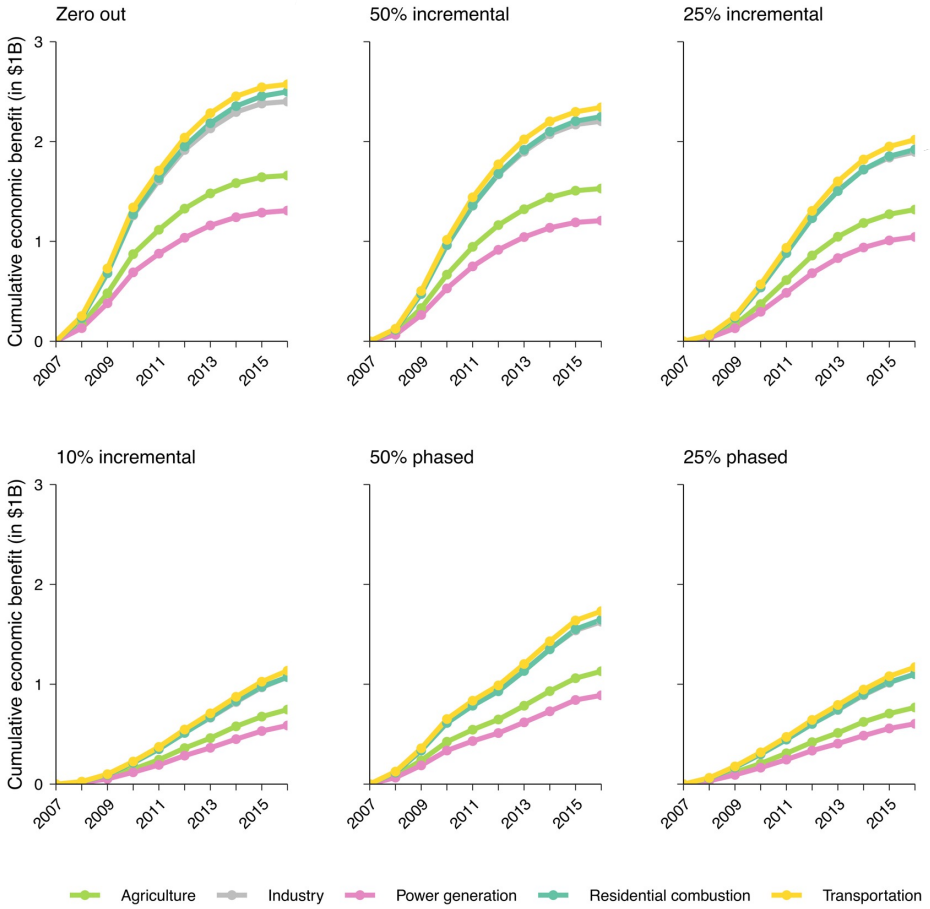
The bulk of life-years saved would have been experienced in the longer term e.g. more lives saved during the last few years of the period than within the first few years



Source: Chen H, Quick M, Kaufman JS et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *PNAS* 2022; 119 (49): e2209490119. https://doi.org/10.1073/pnas.2209490119.

3. Economic benefit

If the zero-out strategy had been deployed at baseline, per million population, the related health impacts by 2016 for all sources combined would have the economic valuation of $10.4B (95% CI: 8.1–12.7)

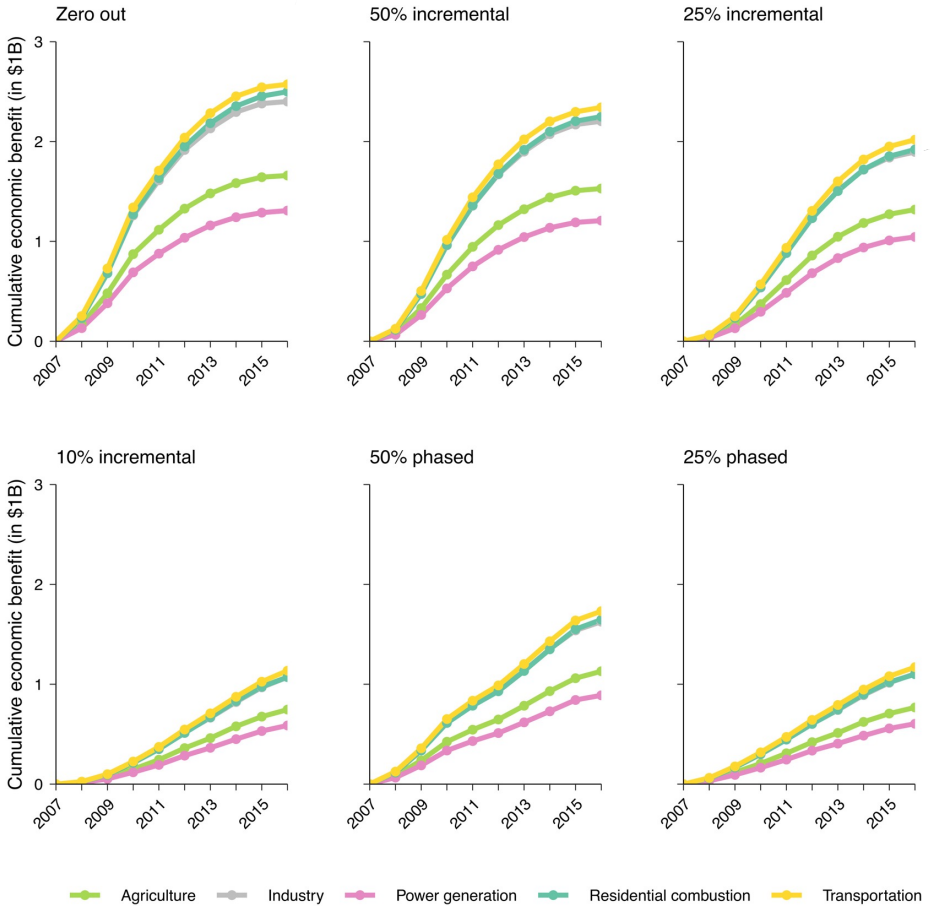


Source: Chen H, Quick M, Kaufman JS et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *PNAS* 2022; 119 (49): e2209490119. https://doi.org/10.1073/pnas.2209490119.

3. Economic benefit

If the zero-out strategy had been deployed at baseline, per million population, the related health impacts by 2016 for all sources combined would have the economic valuation of $10.4B (95% CI: 8.1–12.7)

Under a more modest 10% reduction per year, the corresponding health impacts would have an economic valuation of $4.6B/million for all five sources combined.



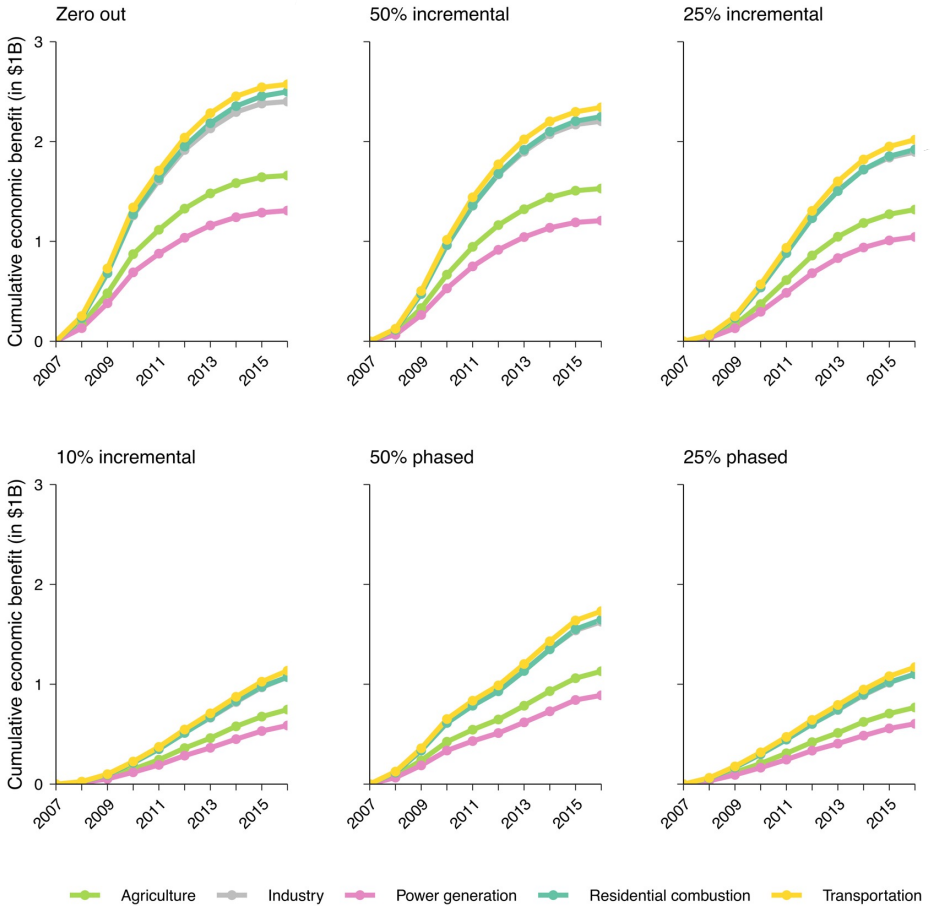
Source: Chen H, Quick M, Kaufman JS et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *PNAS* 2022; 119 (49): e2209490119. https://doi.org/10.1073/pnas.2209490119.

3. Economic benefit

If the zero-out strategy had been deployed at baseline, per million population, the related health impacts by 2016 for all sources combined would have the economic valuation of $10.4B (95% CI: 8.1–12.7)

Under a more modest 10% reduction per year, the corresponding health impacts would have an economic valuation of $4.6B/million for all five sources combined.

The bulk of economic benefit occurred over the long term



Source: Chen H, Quick M, Kaufman JS et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *PNAS* 2022; 119 (49): e2209490119. https://doi.org/10.1073/pnas.2209490119.

Study

Advantages of doing advanced analytics with census cohort data

- Large sample with many individual-level characteristics
- Diverse sample (geography, exposures)
- Long follow-up time and can assess cumulative impacts

Study

Advantages of doing advanced analytics with census cohort data

- Large sample with many individual-level characteristics
- Diverse sample (geography, exposures)
- Long follow-up time and can assess cumulative impacts

Challenges of doing advanced analytics with census cohort data

- Big datasets, computational needs, analysis speed
- Hard to implement machine learning approaches

More information about census health cohorts

https://www.statcan.gc.ca/en/microdata/data-centres/data/canchec

https://crdcn.ca/data/canadian-birth-census-cohort/

References

Branion-Calles M, Winters M, Rothman L, Harris MA. Risk factors and inequities in transportation injury and mortality in the Canadian Census Health and Environment Cohorts (CanCHECs). *Epidemiology*. 2024 Mar 1;35(2):252-62.

Cakmak S, Toyib O, Hebbern C et al. Industrial air pollutant emissions and mortality from Alzheimer's disease in Canada. *Hygiene and Environmental Health Advances*. 2022 Dec 1;4:100019.

Chen H, Quick M, Kaufman JS et al. Impact of lowering fine particulate matter from major emission sources on mortality in Canada: A nationwide causal analysis. *PNAS* 2022; 119 (49): e2209490119. https://doi.org/10.1073/pnas.2209490119.

Malagón T, Morais S, Tope P et al. Site-specific cancer incidence by race and immigration status in Canada 2006–2015: a population-based data linkage study. *Cancer Epidemiology, Biomarkers & Prevention*. 2023 Jul 5;32(7):906-18.

Marouzi A, Plante C, Fornssler B. Creating an 11-year longitudinal substance use harm cohort from linked health and census data to analyze social drivers of health. *medRxiv*. 2024 Feb 15:2024-02.

Stringer L, Ly TL, Moreno NV et al. Assessing geographic and industry-related trends in bladder cancer in Ontario: A population-based study. *Canadian Urological Association Journal*. 2022 Feb;16(2):E82.

Tjepkema M, Christidis T, Olaniyan T, Hwee J. Mortality inequalities of Black adults in Canada. *Health Reports*. 2023 Feb 1;34(2):3-16.

Wilk P, Ali S, Anderson KK, Clark AF et al. Geographic variation in preventable hospitalisations across Canada: a cross-sectional study. *BMJ open*. 2020 May 1;10(5):e037195.