From local to global: Epidemiologic studies of the health impacts of environmental exposures

Michael Brauer
School of Population and Public Health

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

PIMS Big Data in Environmental Science. UBC May 15, 2015
Overview

• Border Air Quality Study (BAQS)
• Canadian Census Health and Environment Cohort (CanCHEC)
• Global Burden of Disease (GBD)
Linked environmental and administrative health data to assess the health impacts of exposure to air pollution
If you build it...they will come
If you build it...they will come
If you build it…they will come

- Perinatal Database & Canada Census
- Physician Billing, Hospital Discharge, Vital Statistics Databases
- Residential history
- Geographic linkage
- Ambient Air Pollution Data
- Covariates
- Outcomes
- Air pollution exposure

BAQS
If you build it...they will come
If you build it...they will come
If you build it...they will come

- Perinatal Database & Canada Census
- Physician Billing, Hospital Discharge, Vital Statistics Databases
- Residential history
- Ambient Air Pollution Data
- Geographic linkage
- Covariates
- Outcomes
- Air pollution exposure
- Greenness
- Air Pollution (NO, NO₂, PM₂.5, BC)
- Noise (Traffic and all Sources)
- Walkability and Park Distance

and more...
Database linkage

- Perinatal Registry
- Canada Census
- Physician Billing
- Hospital Discharge
- Vital Statistics
- Prescriptions

Residential history

Environmental Data

Other risk factors

Health

Exposure

Geographic linkage
Birth Cohort
Birth Cohort

- **Birth cohort** identified 92,158 children born in the study area during a 4-year period (1999–2002).

- **Early childhood cohort** followed these newborns in 1999/2000 for 8 years.
Adult Cohort

- 678,361 adults 45–84 yrs (1999) area residents who lived in study area
- 5-year cohort definition period (1994–1998) to assess pre-existing diseases
- Follow-up to 2002
Residential Histories

- Provincial health plan registry files, hospital discharges, and physician billing records used to establish residential histories (postal code resolution).
  - Not straightforward: multiple databases of residential addresses (of varying qualities), updated at different times.
  - 34.9% of cohort had 2 addresses during pregnancy.
  - Opportunity to examine what happens when you change exposure (due to residential mobility).

89.2% of postal codes referenced to block face
Air pollution exposure

Typically,

• Measure a few key ingredients of complex mixture
• Measure background
• Capture time trends
Exposure Assessment Design

\[ \text{Exposure} = C_{\text{region}} + C_{\text{urban}} + C_{\text{traffic}} + C_{\text{woodburning}} + C_{\text{point source}} \]
Land Use Regression

Legend
- NOX Sampling
- NOX & PM Sampling
- NOX Colocation
- NOX & PM Colocation

Legend:
- Freeways
- Major Roads
- Open/Agricultural
- Residential
- Other

\[ y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + k \]

Map with various land use and sampling locations.

Sampling Location

100m
50m

Table with data:

- Toronto ID
- 50km
- 100km
- 200km
- Open/Agricultural
- Residential
- Commercial
- Industrial
- Environmental

Map with buffer zones and land use classification.
Woodsmoke

Mobile Monitoring on Cold, Clear Winter Evenings

(~ 12,000 points)
Spatially resolved independent variables (e.g., dwellings)

**Catchment area buffer** extracted independent variable (e.g., emissions, bldg. Age, SES, population)

Average of all measurements within buffer area (dependent variable)

Regression Model $R^2 = 0.6 - 0.7$
Evaluation studies: mobility
Evaluation studies: model stability

110/116 sites decreased

<table>
<thead>
<tr>
<th></th>
<th>Method 1 (temporal trend)</th>
<th>Method 2 (predictor values)</th>
<th>Method 3 (1&amp;2 joint)</th>
<th>Method 4 (calibrating coefficients)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forecast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.54</td>
<td>0.52</td>
<td>0.52</td>
<td>0.61</td>
</tr>
<tr>
<td>Error mean</td>
<td>1.57</td>
<td>4.62</td>
<td>1.15</td>
<td>0.00</td>
</tr>
<tr>
<td>SD</td>
<td>2.35</td>
<td>2.34</td>
<td>2.34</td>
<td>2.08</td>
</tr>
<tr>
<td><strong>Back-cast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.44</td>
<td>0.46</td>
<td>0.46</td>
<td>0.49</td>
</tr>
<tr>
<td>Error mean</td>
<td>-1.58</td>
<td>-5.16</td>
<td>-1.69</td>
<td>0.00</td>
</tr>
<tr>
<td>SD</td>
<td>3.08</td>
<td>3.05</td>
<td>3.05</td>
<td>2.96</td>
</tr>
</tbody>
</table>

NO in 2003

Change from 2003

Forecast, using 2003 model ($R^2 = 0.52$) to predict concentrations in 2010

Back-cast, using 2010 model ($R^2 = 0.63$) to predict concentrations in 2010
Evaluation studies: infiltration

Average outdoor PM$_{2.5}$ level $15 \, \mu g/m^3 \rightarrow$ indoor levels $5 - 9 \, \mu g/m^3$
Traffic-related air pollution (children)

- Increased low birthweight and pre-term birth
  - living <50m from provincial highway: 21% ↑ low birthweight

- Bronchiolitis
  - living <50m from provincial highway: 6% ↑

- Middle ear infections
  - 7% of cases attributable to traffic

- Asthma (early life exposure)
  - 13% of childhood asthma attributable to traffic

Woodsmoke

- 15% increase in SGA birth
- 32% increase in otitis media
- 8% increase in bronchiolitis
- 15% increase in COPD hospitalization
- No associations with:
  - pre-term birth
  - asthma incidence
  - cardiovascular, COPD mortality

++ > traffic pollution, +~traffic, - <traffic
Design “natural experiments”

- **Exposure Period**
  - Jan1994
  - Reference: Non-exposure
  - Constant exposure
  - Moved away
  - Moved close

- **Follow-up Period**
  - Baseline Jan1999
  - End Dec2002

**Design**
- “natural experiments”
Road proximity & cardiovascular death

Adjusted RR for CHD Mortality

≤ 50 m vs > 50 m Highway
≤ 150 m vs > 150 m Highway
≤ 150 m Highway or ≤ 50 m Major Road
≤ 50 m vs > 50 m Major Road

X-axis: Non-exposure to traffic

- Constant exposure to traffic
- Moved close to traffic
- Moved away from traffic

BAQS
Additional linkages to assess multiple built environment characteristics
Noise and air pollution joint effects

Impact of Noise and Air Pollution on Pregnancy Outcomes

Term birthweight

Joint associations of transportation noise and air pollution

Noise
Healthy neighborhoods – walkability and air pollution

HIGH walkability / LOW NO / LOW O₃

LOW walkability / HIGH NO / HIGH O₃
Joint Exposure – Birth Outcomes

Greenspace counteracts the negative impacts of air pollution and noise.
The Canadian Census Health and Environment Cohort

Rick Burnett, Health Canada
Dan Crouse, Health Canada -> University of New Brunswick
Michael Tjekema, Statistics Canada
Canadian Census Health and Environment Cohort (CanCHEC)

- 1991 long-form census
  - N = 2.7 million
  - age & sex
  - education
  - income
  - employment
  - immigration
  - ethnicity

- Residential mobility through time (Tax records)

- Canadian Mortality Database

- Canadian Cancer Registry

- Satellite-derived PM$_{2.5}$

- Land use regression models

- Observations from fixed-site stations

- Health Surveys (CCHS)
Structure of the 1991 Canadian Census Cohort

- Mobility: 1984-2011
- Cancer: 1969-2011
- Longitudinal Worker File: 1983-2011*

*Available for ~200,000 Cohort members
OR (95% CI) all lung cancer per 10 µg/m³ increase in PM$_{2.5}$
Spatiotemporal model: 1.29 (0.95 – 1.76)
Urban subset (monitors): 1.33 (0.82 – 2.15)

Indirect adjustment of health risk behaviours (e.g. smoking)

\[ HR_{adj} = \frac{HR_{unadj}}{HR_{smoking}^{pse - pe\cdot ps}}. \]

- \( p_e \) proportion of subjects exposed
- \( p_s \) proportion of subjects who smoke
- \( p_{se} \) proportion of subjects exposed, who smoke

Obtained from ancillary dataset (e.g. CCHS)
LUR: within-city variability in NO$_2$
## Partitioning exposure into within- and between-city contrasts (ppb)

<table>
<thead>
<tr>
<th>Year</th>
<th>City</th>
<th>Within-city</th>
<th>Between-city</th>
<th>Overall exposures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Subject-level LUR estimate – LUR mean</td>
<td>Citywide mean from the long-term trend</td>
<td>Sum of within &amp; between</td>
</tr>
<tr>
<td>2006</td>
<td>Montreal</td>
<td>15 - 11 = 4</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>2005</td>
<td>Montreal</td>
<td>15 - 11 = 4</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>2004</td>
<td>Windsor</td>
<td>18 - 10 = 8</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td>Windsor</td>
<td>18 - 10 = 8</td>
<td>27</td>
<td>35</td>
</tr>
</tbody>
</table>
Adjusted Mortality HRs (Personal: immigrant, visible minority, marital status, education, income, occupational class, employment status; Contextual: income, education, % recent immigrants, temperature) with city random effect and indirect adjustment for smoking and BMI. Crouse et al., 2015.


Summary
Quantification of the disease burden caused by different risk inform prevention by providing an account of health loss attributable to each disease, injury, and other noncommunicable health problems. The burden of disease attributable to all risk factors in 2010

Methods
We estimated deaths and disability-adjusted life years (DALYs) sum of years lived with disability (YLD) and years of life lost (YLL) attributable to the independent effects of each risk factor and clusters of risk factors in 21 regions in 1990 and 2010. We estimated exposure distribution for each year, region, sex, and age group, and relative risks for all risk factors, as well as the combined impact of risk factors. We used these estimates, together with estimates of cause-specific deaths and DALYs from the Global Burden of Disease Study 2010, to calculate the burden attributable to each risk factor. We estimated the burden attributable to each risk factor compared with the theoretical-risk minimum-risk exposure. We incorporated uncertainty in disease burden, relative risks, and exposures into our estimates of attributable burden.

Findings
In 2010, the three leading risk factors for global disease burden were high blood pressure (7.9% 95% uncertainty interval 6.7–9.1% of global DALYs), tobacco smoking including second-hand smoke (6.1% 5.3–6.9% of global DALYs), and alcohol use (5.1% 4.7–5.5% of global DALYs). In 1990, the leading risk factors were childhood overweight (7.9% 6.8–9.1% of global DALYs), household air pollution from solid fuels (7.9% 6.8–8.5% of global DALYs), and tobacco smoking including second-hand smoke (6.1% 5.4–6.8% of global DALYs). Dietary risk factors and physical activity collectively accounted for 10.9% 95% uncertainty interval 9.7–12.1% of global DALYs in 2010. With the most prominent dietary risks being low in fruits and vegetables high in sodium. Several risks that primarily affect child health and communicable diseases, including unimproved water and sanitation and climate microclimate deficiencies, fell in rank between 1990 and 2010, with unimproved water access falling from 17th to 19th place.

What is the Global Burden of Disease?

• Systematic quantification of magnitude of health loss due to diseases, injuries and risk factors

• Global disease, injury, & risk burden estimates for 1990 – 2013 (5 yr intervals) using comparable methods for 188 countries (+ sub-country analyses)
  
  – incidence and prevalence of 301 diseases and injuries and 2,337 relevant disabling sequelae, stratified by sex and 20 age groups
  
  – **Role of 76 modifiable risk factors in burden of disease**

• Collaborative effort coordinated by (Gates-funded) Institute for Health Metrics and Evaluation (UW), [WHO] + ~1000 volunteers....

• Annual updates beginning in 2015

http://www.healthdata.org/gbd
Mortality and Burden of Disease

- Mortality = Numbers of Deaths
- Burden = Disability Adjusted Life Year (DALY)

\[
\text{DALY} = \text{YLL} + \text{YLD}
\]

- years of life lost due to premature death (YLLs)
- years of life lived with disability (YLDs)

one DALY = one lost year of healthy life
Exposure to Outdoor Air Pollution

PM$_{2.5}$ Ozone

General approach

Worldwide Health Evidence

Concentration–Response Relationships

Country-Specific Mortality, Disease

Baseline Incidence

Global Burden, DALYs, Mortality

Population Attributable fraction $\times$ Deaths (cause-specific)
Population Attributable fraction $\times$ DALYs (cause specific)
Population attributable fraction

\[ PAF = \frac{\int_{x=0}^{m} RR(x)P(x)dx - \int_{x=0}^{m} RR(x)P'(x)dx}{\int_{x=0}^{m} RR(x)P(x)dx} \]

Population Attributable Fraction: sex, age, country, time
Measurements

Annual average PM2.5 (µg/m3)

- 0.0 - 8.4
- 8.4 - 13.0
- 13.0 - 18.1
- 18.1 - 25.3
- 25.3 - 34.1
- 34.1 - 46.2
- 46.2 - 61.0
- 61.0 - 75.6
- 75.6 - 141.5
- 141.5 - 193.9
• Final estimates based on average of (1.4 million) grid cell values (SAT, TM5) and calibrated (regression model) with measurements
  - 0.1° x 0.1° resolution
• Incorporate variance between two estimates and measurements in uncertainty assessment
• Unique contributions from each approach
Integrating risk from multiple sources to estimate risk due to ambient PM$_{2.5}$
Integrated Exposure-Response functions (IER)

Key assumption
Risk is function of PM$_{2.5}$ inhaled dose regardless of source

Extrapolation model
• reflect change in risk observed in cohort studies at low concentrations
• near-linear at low concentrations
• predict risk for highest PM$_{2.5}$ consistent with risks from smoking (Pope et al. 2011)

Burnett et al. 2014