Environmental Epidemiology

Mark Goldberg

INRS-Institut Armand-Frappier, University of Quebec,
and McGill University

October 1999

Environmental Epidemiology

- ✔ The study of the determinants of the distributions of disease that are exogenous to and nonessential for the normal functioning of human beings

Adapted from Hertz-Piccioto

Types of Environmental Exposures

- ✔ Point sources
  - – Pollution from a factory, municipal solid waste site
- ✔ Line sources
  - – EMF exposures from high tension power lines
  - – Combustion pollutants around high density motorways (TSP map)

- Area sources
  - Airborne combustion products from traffic and long-range transport
  - Volatile organic compounds contaminating underground water reservoirs

Example: Cancer Rates Living Near a Solid Waste Landfill Site

- ✔ Ecological Analysis (Goldberg et al., Arch Environ Health 1995;50:417-24)
  - – Landfill site opened in 1968
  - – 100,000 persons lived within 2 km of the site
  - – In 1993, it contained about $36x10^6$ T waste
  - – Rates for men and women living in zones around site 1981-1988
    - Zones defined by 3-character postal codes
    - Putative “upwind” and “downwind” zones
      - – Putative unexposed zone far from the site
    - Poisson regression adjusted for age and year, by sex

- Reference zones selected from the “unexposed” areas to ensure similarities for:
  - average household income
  - proportion of immigrants
  - proportion first language was French
  - unemployment and poverty rates
- Matching was not entirely successful, as some key factors were dissimilar (e.g., % of Italian)

- **Improved method for reference zones:**
  - Create a score $y_k = \Sigma_i\ c_j * \text{factor}_{jk}$

    where k = enumeration area and c_j is a scale factor that indicates the contribution of factor_i on the risk of developing the outcome of interest (from previous data) (normalized such that $\Sigma_i\ c_j = 1$)
  - conduct a cluster analyses on this variable and those enumeration areas that are "close" to the exposed area are used as the reference zone

- See Map
  - From paper “Risk of prematurity near a landfill site”

- ✔ Analytic study (Goldberg et al., Arch Environ Health 1999;54:291-6)
  - Multi-site cancer case-control study of occupation, men, 1979-85
  - Distance from site and by geographic zones (at time of interview)
  - Logistic regression for each site, adjusted for occupational and nonoccupational risk factors
    - Age, family income, cigarette smoking, alcohol consumption, ethnicity, place of birth, body mass index, consumption of vitamins, occupational “salubrity”

- See Tables
  - Comparison of Relative Risks for Ecological and Case-control Analyses
  - Relative Risks for Liver Cancer from the Case-control Analysis

Example: Cohort Studies - Harvard Six-cities Study

- ✔ Prospective cohort study of about 8,000 subjects selected randomly from 6 US cities with different levels of air pollution
- ✔ Originally designed as a longitudinal study of respiratory health
- ✔ Subjects followed every two years and lung function and questionnaires administered periodically
- ✔ Ambient air exposures assessed from special fixed-site monitoring stations (particles, sulfates, gaseous pollutants)

- Mortality analyses, comparing mean annual levels in each city for years near start of followup
- Assumed that subjects did not move during followup and that the rank ordering of cities for levels of air pollution was invariant of followup time
- Stratified Cox proportional hazards models to estimate cause-specific mortality relative risks

See Dockery et al., NEJM 1993;329:1753-9

- See Graph
  - Mortality rates by level of pollution by city and by pollutant (photocopy from NEJM article)
- See Table
  - Results from Harvard Six-cities Study: Estimates of Mortality Rates Comparing Most Exposed to Least Exposed City

Example: Time Series Studies

- **Objective:** To determine whether the daily number of deaths increases when air pollution increases on that day or on preceding days
- **Method:** Juxtapose a time series of deaths with a time series of air pollution.
- **Confounding factors:** Any factor that varies on short time scales and is associated with daily mortality (e.g., weather patterns, influenza epidemics). Smoking can not be a confounding variable unless patterns of consumption change on the scale of days.

- Target population consists of all persons living in a well-circumscribed geographical area
- Study is not entirely ecological in that there are no comparisons by place. There are, however, comparisons by time.
- There are no denominators.
- **Exposures:** Daily measurements from fixed-site monitors

- See graphs
  - Time Series Plot of All Cause Mortality excluding Accidents (mortality, COH plots)
  - Time Series Plot of All Cause Mortality excluding Accidents (mortality, temperature plots)

- **Analysis:** Must account for
  - non-independence of daily counts of death (serial autocorrelation)
  - overdispersion
- **Method of Analysis:** Poisson regression (using quasi-likelihood).
- **Statistical model:**
  - $E(\log(Y_i)) = \alpha + f_1(\text{time}_i) + f_2(\text{meteorology}_i) + f_3(\text{pollution}_i) + \ldots.$
  - Covariance corrected for non-Poisson variation

- See Graphs
  - Filtered Mortality by Time
  - Nonaccidental Deaths by Age Group (3-day Mean)

Types of Studies

- ✔ Ecological studies
  - Pure ecological studies
  - Mixed ecological/individual studies
- ✔ Cluster investigations
  - "*'Unusual'* aggregation in time, space, or both of occurrences of disease(s)"
- ✔ Case-control studies
- ✔ Cross-sectional studies
- ✔ Prospective and retrospective cohort studies

Time-Related Analyses

- ✔ Cyclical and other temporal patterns
- ✔ Clustering in time
  - Time series studies
- ✔ Longitudinal trends
  - Age-period-cohort models of rates

Spatially-Related Analyses

- ✔ Clustering in space
- ✔ Mapping of rates
  - – Definition of geographic regions
  - – Sparse data $\Rightarrow$ extreme values
    - Two-stage analyses (e.g., empirical Bayes)
  - – Errors in numerators and denominators
    - Migration to and from study regions
    - Incomplete ascertainment of cases
    - Conversion between different geographic identifiers
      - – e.g., 6-character postal codes and enumeration areas

- Tradeoffs in defining geographic areas
  - Large areas:
    - increased variability of exposure between subjects
    - fewer problems with mobility
    - reduced errors in estimating numerators and denominators
    - less extreme values
    - bias from aggregation of variables at smaller levels of geography (e.g., from enumeration areas to census tracts)

- Small areas
  - reduced variability of exposure between subjects
  - high variability and extreme values for outcomes
  - difficulties with mobility and estimating numerators and denominators

Example: Cluster Investigation in Reprocessed Textile Workers

- ✔ Observation of unusually high lung cancer mortality rates in 1979 in Prato, Italy
- ✔ High rates of malignant mesothelioma found among rag sorters in Prato
- ✔ A case-control study in Prato (1980-83) showed a 50% excess of lung cancer in textiles workers

See Quinn et al., Am J Ind Med 1987;11:255-66; Paci et al., Am J Ind Med 1987;11:267-73

- Major industry in Prato is recycling of old clothes
- Industrial hygiene survey of rag sorters working in small shops
- Clothing and rags from all over the world
- Clothes arrived in plastic bags or in bales
- Rags sorted by hand by men sitting on the floor
- Rags then baled and shipped to other processing plants in two (e.g., extraction of wool fibers, dyeing)

- It was found that bags from Canada, the US, the Soviet Union, South Africa, and Australia contained large quantities of asbestos
- These bags were ripped open by workers to be used as recycled bale covers
- Asbestos fibers identified in breathing air zones of these workers

Components of an Environmental Epidemiologic Study

- ✔ What is the problem?
  - – Accidents
  - – Perception of a hazard
    - • Clusters in space and time
  - – Investigator’s imagination
- ✔ Precise study objectives

- ✔ Precise definition of target population
  - Who is exposed?
  - Which population can serve as "unexposed" or reference group
  - Mobility patterns

- ✔ Outcomes
- ✔ Definition of potential confounding variables
- ✔ Definition of potential variables indicating biological interactions
- ✔ Statistical power
  - – Size of target population and expected level of effects

Key Issues

- ✔ Expected response rates
- ✔ Migration
- ✔ Measurement of exposures
- ✔ Measurement of potential confounders
- ✔ Interactions?
- ✔ Biases
- ✔ Pilot studies

Outcomes

- ✔ Acute versus chronic effects (latency)
- ✔ Precise definitions
  - Cancer
    - Histological confirmation
  - Respiratory
    - Chronic obstructive pulmonary diseases
      - ATS standardized questionnaire
    - Asthma
    - Lung function
      - Standardization to expected values (age, height, gender)

Confounding Variables

- ✔ Definitions and effects differ depending on whether study is ecological or individual-based
- ✔ Individual studies: causally associated with outcome and associated with exposure
  - Effects must be estimated on same scale (e.g., correlation coefficients do not reflect level of association in case-control studies (odds ratios))
  - Not in causal pathway
  - Variables can also be used to adjust for selection biases

Biological Interactions

- ✔ Susceptible subgroups
  - – Fewer subjects, perhaps greater effects (effect on power??)
  - – Gene-environment interactions

Environmental Exposure Assessment

- Exposure
  - Amount of a contaminant that a person may come into physical contact with over a specified period of time
- Dose
  - Amount of a contaminant that is absorbed or deposited in an organism over a specified period of time
  - Usually measured as mass per unit volume or per unit mass of affected tissue (e.g., blood lead levels in µgm per deci-liter)

- See Flow Chart
  - Environmental Exposure and Health Effects

- Exposures versus dose
  - Distribution in the body
  - Chemical and physical properties of agents (e.g., solubility in water, lipid tissues)
  - Metabolic processes, detoxification => metabolites
  - Body burdens (sojourn times, interactions with other organs, feedback mechanisms)

Action of chemicals

- ✔ Genotoxic
  - – Mutagens and carcinogens (e.g., ionizing radiation; benzene)
- ✔ Organ-specific toxicity
  - – Ethylene glycol (aircraft de-icing) causes kidney dysfunction and serious irreversible damage in sufficiently high doses
- ✔ Immunological/neurological effects
  - – E.g.; VOCs may induce neurogenic inflammation mediated through chemical receptors on slow velocity neural C-fibers.

See Meggs Environ Health Perspect 1993:101:234-238

Examples of Measurement of Dose

- ✔ Serum carboxyhemoglobin as a marker both for exposure to CO (for a study of cardiovascular diseases)
- ✔ Blood lead levels in children living near major traffic arteries (for a study of intellectual functioning)

Organ-specific Doses of Ionizing Radiation from Diagnostic X-rays

- ✔ Energy and distribution of flux of photons at skin estimated from:
  - Geometry of radiograph (view, distance to x-ray tube)
  - Parameters of x-ray tube (voltage, amperage, integrated time)
  - Shielding
  - Age and gender of subject

- Organ-specific doses estimated from Monte-Carlo calculations of photon flux through simulated body
  - Validated using standard phantom
- Doses estimated for members of a cohort of Adolescent Idiopathic Scoliosis
- Excess risks projected into time using dose-response models and lifetables

See Levy, *Health Physics*, 1994;66:621-33

Methods of Estimating Exposure

- Questionnaires
  - Information on physical properties of an environment
    - E.g., standardized questionnaire on indoor air quality
    - “Are you exposed to ….?”
    - “Do you use a wood stove…?”

- Simple categorization of potential exposure
  - "Do you smell odours around your home…?"
  - "How many members of your family smoke cigarettes in your home?" "Approximately how many per day?" etc….
- Activity patterns
  - "How much time do you spend doing…?"

- ✔ See Questionnaire Examples

Direct Measures of Exposure

- ✔ Personal monitoring
  - In breathing zone
  - Through dosimeters and other active or passive samplers worn by subjects
  - Response rates very important
  - Development of prediction models comparing personal measures with area measures
  - Be wary of longitudinal versus cross-sectional studies

Indirect Measures of Exposure

- Microenvironmental monitoring
  - Long-term samples or grab samples to determine spatial-temporal distribution (active or passive samplers)
  - Ambient air monitors for O_3, SO_2, particles
  - Measurement of methane & VOCs in air, soil, and ground water around solid waste landfill sites

- Statistical modelling (prediction models) using questionnaires, area measures, and personal monitoring
  - Interpolation techniques (Kriging)
- Other methods
  - Proximity to source
  - Distance from source

Biological Monitoring

- Cellular, biological or molecular measures obtained from biological media (human tissues, cells, fluids) that are indicative of exposure (referred to as markers for exposure)
  - Exogenous substance
  - Metabolite
  - Interaction of xenobiotic substance and molecule

- Measures of dose
  - Temporal issues:
    - from current or previous exposures?
    - from integrated or point exposures?
- Examples:
  - Exogenous agents (lead, asbestos, nicotine)
  - Metabolized chemicals (phenol, cotinine)
  - Endogenously produced (hydroxol radicals after ionizing radiation exposure)

- Molecular changes (benzo*a*pyrene DNA adducts)
- Cellular or tissue damage (sperm mobility)
- Pulmonary response (challenge tests)
- Skin response (chloracne after DDT exposure)
- Gastrointestinal response (diarrhea)
- Biological fluids (urinary cotinine after exposure to ETS)

- Issues
  - Response rates
  - Poorly understood relationships between biomarkers and exposures and outcome
  - Cost, resources
  - Between and within person variability
  - Reference period for exposure

Improving on Exposure

- ✔ Pharmacokinetic models
  - Used to calculate doses to target tissues from exposures
- ✔ Pharmacodynamic models
  - Used to describe dynamic processes that relate doses and effects in tissues to ultimate health effects

Example: Modelling Pulmonary Doses from Occupational Exposure to Silicon Carbide

- ✔ Cross-sectional study of pulmonary function and radiographs for pulmonary pneumoconioses
- ✔ Dust exposures assessed in detail by job
- ✔ Lung clearance of dust efficient at low doses
- ✔ "Overload" occurs when clearance mechanisms become saturated
  - Lung dose is assumed to be proportional to cumulative exposure

See Ballew et al., Am J Epidem 1995;141:690-6

- See Graph
  - “Lung fibrosis model…”

- Set of partial differential equations describe the model, which reduces to
  - Lung burden at time T is equal to the sum of:
    - lung burden at $(T-t_0)$ * exponential clearance term over time
    - volume of air inhaled * fraction insoluble particles/clearance rate * (1- exponential clearance term over time)

- See Table
  - “Comparison of models estimating the risk of radiographic opacities”

Issues in Exposure Assessment for Health Studies

- Objectives of health study
- Multiple contaminants
- Source of pollutant(s)
  - Multiple pollutants (complex mixtures)
- Route of exposure into the body
  - Respiratory, skin, ingestion

- Expected biological response
  - Chronic versus acute exposures
  - Temporal patterns of exposure
  - Latency
- Technology to estimate environmental and personal exposure
  - Biomarkers

- ✔ Exposure misclassification
  - Spatial and temporal variability
  - Between-person variability within “exposure areas”
- ✔ Sampling period in relation to expected biological effects
- ✔ Frequency and intensity of exposure

- ✔ Reliability and validity of estimates
  - Validation and reliability sub-studies
  - Within-person variability for biological markers
- ✔ Costs, resources
- ✔ Expected response rates

Ecological Studies

- ✔ **Definition:** An investigation of the distribution of health and its determinants between groups of individuals.

- ✔ The degree to which studies are purely ecological can vary considerably.

Reasons for Ecological Studies

- ✔ Data on the individual level not available
- ✔ Individual exposure measurements not available, but grouped level data are (e.g., mean radon gas levels from county-wide surveys)
- ✔ Comparison between large jurisdictional units (e.g., comparison of breast cancer rates with mean daily fat intake between countries)

- ✔ Quick and dirty
- ✔ Design limitations (e.g., Harvard Six-cities study)
- ✔ Interest in ecological effects (e.g., does increasing taxes on tobacco in different jurisdictions reduce consumption?)

Measurement variables

- ✔ **Aggregated measures:** summaries of attributes calculated from data on individuals for whole populations in well-defined geographic regions

  Examples: mean income; percentage of families below the poverty line; mean number of household members

- **Group level measures:** estimates of (environmental) attributes that have individual analogues. Usually obtained from surveys.

  Examples: maximum daily exposure to ozone; mean annual exposures to radon gas; daily mean levels of environmental tobacco smoke in public buildings

- **Global measures (contextual):** attributes that pertain to groups and do not have analogues at the individual level

  Examples: total area of green space; number of private medical clinics; population density

Types of studies

- ✔ **Individual level:** Well define target and study populations and data available on individuals for all (or most) covariates.

  Example: Cross-sectional study of respiratory symptoms and exposure to environmental tobacco smoke among children living in Montreal.

- **Purely ecological:** No data on individuals

  Example: Average per capita consumption of snuff and age-sex-race standardized mortality rates of oral cancer. Comparisons at the county level.

Partially ecological: Some individual data available.

Example: A study of low birth weight and environmental exposures to biogas from a landfill site.

- *Individual data:* age of mother, sex, birth weight, gestational age of baby, and geographic area of residence

- *Ecological:* geographic region of residence as a surrogate for exposure to biogas in the ambient air

Types of Ecological Studies

- Case-control
- Cohort and longitudinal
- Cross-sectional
- Time trend studies
- Immigrant studies

Levels of Inference

- ✔ Biological inferences on populations
  - Individual-level studies
  - Ecological-level studies
- ✔ In individual-level studies, inferences are made to the target populations using data collected from individuals

- In ecological-level studies, inferences are made strictly to the groups that are under investigation
- Ecological inferences usually refer to *contextual effects*

  <u>Example:</u> An ecological study investigating health care utilization for prenatal care between regions of Montreal as a function of number of clinics per region, etc...

- If a study is purely ecological, then biological inferences to target populations may be made as if the studies were conducted on individuals (referred to as *"cross-level inference")*
- Only under strict conditions will these inferences be correct

Ecological Fallacy

- Assumptions:
  - 1) that the effects estimated at the individual level are the relevant ones for making <u>biological inferences</u>
  - 2) that the effects are a **linear** function of the predictors; i.e. $E[y_i] = \alpha + \beta x_i$

- Assume the above relationship $\{E[y_i] = \alpha + \beta x_i\}$ to hold on an individual level and that the parameter of interest for the purposes of biological inference is β.
- Assume now that the population is segregated into groups and that the analysis proceeds by comparing the grouped mean between the k groups (no individual data available).

- ✔ The slope including group effects is:

  $\beta = \delta\beta_G + \gamma\,\beta_W$

  where β is the overall between-person slope (i.e., over all persons in all groups), β_G is the between-group slope (ecological effect), β_W is the within-group slope, and δ, γ are ratios of the between-group and within-group variances to the total variance of x ($\delta + \gamma = 1$).

- When there are no group effects then $\beta = \beta_W$, so β_W is the correct regression coefficient
- When there are group effects $\beta \neq \beta_W$, so that $\beta_G \neq \beta_W$
- Ecological bias or “cross-level bias” occurs when $\beta_G \neq \beta_W$

See Piantadosi, AJE 1988;127:893-904

Conditions for No Ecological Bias

- ✔ Background rate of disease (in the unexposed) does not vary across groups
  - background rates may vary, apart from statistical variation, due to unequal distributions of risk factors across groups

  AND
- ✔ These is no confounding within groups

  AND
- ✔ There is no effect modification by group

- In general, the ecological linear regression model will estimate the difference in rates between groups.
- The ecological regression coefficient is equal to the sum of:
  - difference in rates at the individual level
  - bias from the association between the confounding factor and group
  - bias from the interaction between a factor and group (only if the difference in rates does not vary by group will there be no interaction)

Examples of Ecological Bias

- ✔ Group is an effect modifier
  - – i.e., effect of exposure varies across groups
  - – can arise from differential distribution of effect modifiers across groups
  - – can occur even if after control for ecological covariates

- See Table:
  - Ecological Bias: Effect Modification by Group, No Confounding

Confounding by Non-Confounders

- Variable is not a confounder on the individual level
  - may occur if background rates vary by group
  - if rate differences between groups not constant

- See Table:
  - Ecological Bias: Ecological Confounding of a non-Confounder on the Individual Level

Adjustment for Ecological Confounder Increases Bias

- Variable is not a confounder on the individual level (factors not associated)
  - background rates differ by group
  - rate differences vary by group

- See Table:
  - Ecological Bias: Ecological Confounder Increases Bias

Nondifferential Misclassification of Exposure

- ✔ For both linear and log-linear models nondifferential misclassification of exposure (binary variable) leads to an overestimation of effect in ecological studies, even if there are no other sources of ecological bias

See Brenner AJE 1992;135:85-95

Non-Linear Effects of Covariates

- If there is a nonlinear association between the outcome, the exposure and the covariate, ecological bias may occur
  - due to the linear ecological model not holding in the underlying population (e.g., Risk(x,c) = (1 + ßx) exp(γc))
  - not correctly summarizing the ecological covariates across groups (using just means instead of other more complex summaries)

See Greenland and Robbins, AJE 1994;139:747-760

Possible Solutions

- ✔ Obtain detailed information on covariates so that not just mean levels are used in the analysis
- ✔ Obtain joint distributions of covariates and exposures
- ✔ Use another analytic approach (individual-level or semi-individual-level studies)

Example: Association between Radon in Homes and Lung Cancer

- Studies of uranium miners and smelters have shown strong <u>positive</u> exposure-response relationships between level of radon gas and lung cancer
- Ecological studies of lung cancer rates and mean level of radon by county in the US and elsewhere have shown strong <u>negative</u> correlations

Case-control Study in Sweden

- 1360 cases and 2847 controls
- Age 35-74 years, 1980-84, living in 109 municipalities
- Radon monitored in 9000 homes occupied by subjects since 1947 for > 2 years
- Time-weighted concentrations estimated per subject
- Carried out an analysis of individual data and ecological data

- ✔ **Ecological radon levels:** Average radon exposure aggregated in each municipality from controls living there
- ✔ **Ecological analysis:** Odds ratios per county (only males with >10 cases per county)

See Lagarde and Pershagen AJE 1999;149:268-74)

- See Graph:
  - "Ecological Association of Lung Cancer and Radon by County, Sweden"

- See table:
  - “Association of Radon and Lung Cancer Risk: Comparison of Individual and Ecological Estimates”

References

ENVIRONMENTAL EPIDEMIOLOGY

Chapter 28, “Environmental Epidemiology”, Irva Hertz-Piccioto, in Rothman and Greenland

“Human Exposure Assessment for Airborne Pollutants: Advances and Opportunities”. National Academy of Sciences, Washington, DC 1991

ECOLOGICAL STUDIES

Chapter 23, “Ecological Studies”, Hal Morgenstern, in Rothman and Greenland

Richardson et al., Int J Epidem 1987;16:111-120.

Piantadosi et al., Am J Epidem 1988;127:893-904.

Greenland and Morgenstern, Int J Epidem 1989;18:269-274.

Brenner et al., Am J Epidem 1992;135:85-95.

Brenner et al., Epidemiology 1992;3:456-9.

Greenland and Robbins, Am J Epidem 1994;139:747-760.