

- Use a GAM to find the best parametric form (e.g., x^2). This can be achieved by inspecting the plots of $f(x)$ versus x (with the residuals) and by comparing the AICs between the GAM and the parametric versions.
- Then use the semi-parametric GAM to provide an adjusted estimate of x :
 - $E(y) = \beta_0 + \beta_1 x^2 + f(z)$

- Another way to use the GAMs is not to try to determine the most parsimonious model using statistical inference but to use it as a mean to find models that make epidemiological sense.
- For example, assume that the variable of interest is x and the covariate is z , and both are measured continuously.

The Montreal Occupational Breast Cancer Study

- **Objective:** To determine whether past occupational exposures are associated with increased rates of postmenopausal breast cancer.

Design

- Population-based case-control study
- 18 hospitals in Montreal
- *Cases:* Incident breast cancer, ≥ 50 y
- *Controls:* other selected sites of cancer
 - ~ frequency-matched to cases by age and hospital

Fieldwork

- Interviews by telephone or face-to-face
- Detailed questionnaire on nonoccupational risk factors
- Probing questionnaire regarding details of each occupation
- Interviews given to a team of chemists and industrial hygienist who translate job descriptions into a set of profiles of exposure for about 300 agents

- **Response rates:**
 - 81% cases
 - 73% controls
- **Quality of data:**
 - few missing data
- **Validity of data:**
 - control population consists of other sites of cancer
 - do we find the same risk factors usually observed in c-c studies of breast cancer?

Statistical Analyses

- **Outcome:**
 - breast cancer, binary response => logistic model
- **Explanatory Variables:**
 - Age at menarche
 - Number of full-term births
 - Age at 1st full-term birth

Confounding Variables

- Age at diagnosis
- Age at menopause
- Previous breast disease
- Alcohol consumption
- Education
- Family history of breast cancer
- Language
- "Ethnicity"

Questions

- Which covariates are associated with the risk of developing breast cancer?
 - What is the best functional form to use for each?
- Are the three reproductive variables independent risk factors?
 - Can they be modeled together (collinearity)?
 - What is the best functional form?