

# Studies of rare cancers:

## Control selection (Case characterization)?

Sholom Wacholder, NCI  
Bethesda, May 11, 2007

## Methodologic Challenges in etiologic studies of rare cancer

- **Small numbers**
- **No economy of scale**
  - Fixed costs amortized over fewer cases
    - Qx development
    - Procedures
    - Consortia need to be larger to get to fixed # of cases
- **Case ascertainment**
  - Identify cases
  - Do tertiary referral centers find everyone
- **Given all the cases, where are controls?**

## Opportunities in rare cancers

- **Being first**
- **Detailed clinical characterization of disease available**
  - Tertiary hospitals
- **Motivated families**
- **Advocates can help**

## Standards for funding and publication

- Studies of rare cancers cannot meet same standards as studies of common cancers
  - Power too low
  - Control selection
  - Case ascertainment
  - No prior hypotheses based on epidemiology data

## Options for controls

- Global controls
- Sibling controls
- Sharing controls
- Hospital controls
- No controls

# Global controls

- Violate main principle of control selection
  - “*study base*” principle
- But ... for genetic factors
  - “Genomic control” methods may fix
    - Use other markers to correct
  - Will it work?
  - We are looking empirically to see
- For G-E interaction ...
  - look in Wacholder, CEBP, around 2002

## Sibling and related “controls”

- Especially for children’s diseases
- Take advantage of motivated families
  - Coercive?
- “Overmatched” on family, parental characteristics
- TDT methods (Spielman; Self)
  - Parents’ DNA ideal; sibs’ DNA works
- Also
  - Affected sib pair method if strong genetic determinant
    - >1 affected sibling in family

# Sharing controls

- If complete ascertainment of more than one case group in same study base ...
- ... economy of scale from sharing controls
  - E.g., NCI Black-White study from 1980's
- Biospecimen volume sufficient for assays needed with each case group
- For diseases with little prior knowledge
  - Qx's can be similar
- Otherwise: "partial qx"
  - Wacholder, 1994
  - Ask different subsets of controls to minimize burden on controls

## Hospital controls

- Use cases of other, specified disease(s) as controls
  - Same catchment areas
  - Unrelated to exposures of interest

# Case only

- **Not helpful for first study of a single case group**
  - No estimate of direction of effect
    - Only comparison of effect of X1 in presence, absence of X2
- **For multiple case groups**
  - Hospital controls are a variant of case-only
  - Cf. Exposure distribution in each case group vs. all or subset of others
    - Exclude case types likely to have similar etiology to case disease
    - Use empirical-Bayes compromise methods to choose which subset of other case groups to be used as controls for each exposure of interest
      - Exclude outliers
    - Someone needs to work theory out (2008)
      - (No guarantee it will work)

## Affirmative action

- Lower methodologic standards for rare diseases
- Is the 100<sup>th</sup> breast or prostate epidemiology cancer study more informative than the first study of a rare cancer?
  - Increase power of study of common cancer from 98% to 99%
  - or increase power of study of rare cancer from 0 to 50%?

## Pooling related case groups

- If appropriate controls available
- Motivation: sarcoma, where there are multiple subtypes
- Define each case by tumor characteristics
  - Clinical, pathologic, molecular
- Look for common characteristics across different disease types or disease subtypes
  - Don't impose standard nosological conventions

## Examples (some common cancers, sorry)

- **Breast cancer**
  - Cross
    - ER status
    - PR positive
- **Hematopoietic malignancies**
  - Leukemia vs. lymphoma
  - Lymphoid vs. myeloid cells
    - Further cell of origin distinction
  - Further pathology, molecular distinctions

## Sophistication on Exposure side, Simplicity on disease side

- Joint effects of 2 dichotomous exposures
  - Fit X1 + X2 in model
    - Not 4 levels
      - X1 pos; X2 pos
      - X1 pos, X2 negative
      - X1 negative, X2 pos
      - X1 neg, X2 pos
    - Identify whether X1, X2 have “independent effects”
      - Adjust for confounding
      - Effect modification
- We are not nearly as sophisticated on Disease side
  - At best, polytomous regression

# Modeling of Disease characteristics

- We can model  $Y1=ER$  and  $Y2=PR$ 
  - Chatterjee, JASA, 2004
  - Colditz et al., JNCI 2004 (?)
  - Adjust for confounding on Y side
- Tumors with common characteristics may have common etiology, even if nosological classification is different

## Control selection

- **Vexing problem**
  - Expensive to do properly
  - Challenging even to calculate response rate
    - Especially in real time
  - Response rates are low
    - Especially when biospecimens are requested

National Cancer Institute

## Bioinformatics

- Let ORD set up chat room/list-serve to facilitate connections of researchers who want to work together
- Data sharing would be facilitated if data management done right
  - Needs \$ up front