Designs for Developing and Evaluating Models of Absolute Risk

Mitchell H. Gail

NCI Division of Cancer Epidemiology and Genetics

NCI Conference on Risk Models

May 20-21,2004

Outline

- Definition of absolute risk
- Cohort design
- Combining case-control and registry data
- Kin-cohort and other family-based designs
- Combining various data sources
- Validation designs

Absolute Risk of Breast Cancer

age 40
nulliparous
menarche age 14

mother had breast cancer
no biopsies

What is the chance that she will be diagnosed with breast cancer between ages 40 and 70?

Absolute risk = 0.116 (11.6%)

Definition of Absolute Risk

$$\int_a^{a+\tau} h_1(t)r(t)\exp\left[-\int_a^t \{h_1(u)r(u)+h_2(u)\}\,du\right]dt$$

$h_1(t)$ is baseline hazard of breast cancer incidence

$h_2(t)$ is mortality hazard from competing risks

$r(t)=\exp\{\beta^T X(t)\}$ is relative risk of breast cancer

Cohort Study

Age	At Risk	Breast Cancers	Non-BC Deaths
30-39	1000	1	15
40-49	984	15	30
50-59	939	20	61

Absolute risk = (1+15+20)/1000=0.036

Individualized Absolute Risk from Cohort Studies

- **Cox proportional hazards**

$$h_1(t;x) = h_{10}(t)\exp(\beta x)$$

Benichou and Gail, Biometrics 1990
Anderson, Borgan, Gill, Keiding 1993

- **Cumulative incidence regression**

$$g\{\text{Prob}(\text{event 1 at } T \le t;x)\} = h_0(t) + \beta x$$

Fine and Gray, JASA 1999

Problems with Cohorts

- Non-representative absolute risks
- Prospective cohort study takes a long time
- Imprecise and unrepresentative data on competing causes of death
- Lack of detailed covariate data

Sampling a Cohort to Estimate Relative Risks and Cumulative Hazard under Cox PH Model

- Case-cohort design
  - Prentice and Self, Annals Stat, 1988
- Nested case-control design
  - Borgan, Goldstein, Langholz, Annals Stat, 1995

Combining Case-Control Data with Registry Data

Case Control Study

Relative Risk, r(t)

Attributable Risk, AR(t)

Registry

Composite age-specific hazard, $h_1^*(t)$

$$h_1(t)=\{1-AR(t)\}h_1^*(t)$$

Cornfield, JNCI, 1951; Gail et al, JNCI, 1989; Anderson et al, NSABP, 1992

Advantages of the Case-Control/Registry Approach

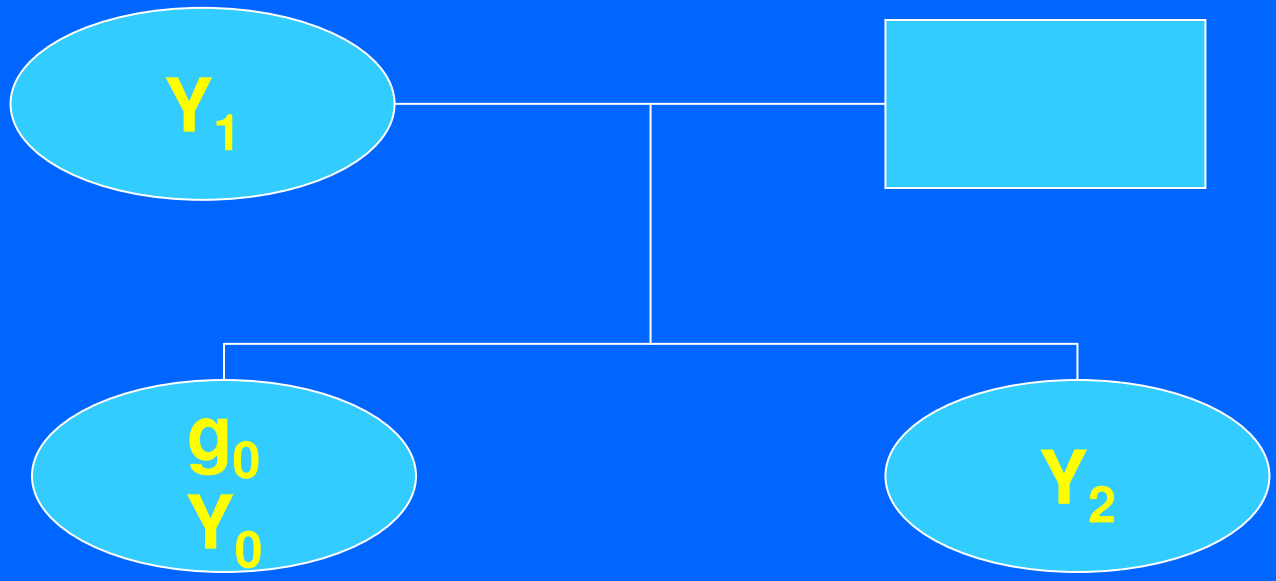
- Detailed information on covariates
- Study takes comparatively little time
- Composite age-specific rates from registry more precise and representative than from cohort
- Can combine several case-control studies to obtain relative risk model

Disadvantages

- Potential recall bias
- Either cases or controls must be representative of general population to estimate AR (unless separate survey of risk factors available)
- National registry data are not available for many endpoints such as stroke and myocardial infarction

Kin-Cohort Design

Struewing, Hartge, Wacholder et al, NEJM 1997



Proband

Gene Risk Estimates from Pedigrees with Many Affected Members

- **Maximize Prob(genetic markers|family phenotypes; θ, allele frequencies, age-specific incidence rates λ_i)**
  - **In theory, this adjusts for ascertainment**

- **Or look at prospective rates of contralateral cancer in mutation carriers**

  Easton et al, Am J Hum Genetics, 1995

Comments

- Ascertainment correction suspect if:
  - Criteria for ascertainment not clear
  - Residual familial correlation from other genes or shared environmental factors (leads to overestimates of penetrance)
- Hard to get covariate information
- Breast cancer risk to age 70 in BRCA carriers: 85% based on this method vs e.g. 56% based on kin-cohort method

Combining Data Sources Based on Modeling Assumptions
Tyrer, Duffy, Cuzick, Stat Med 2004

- National breast cancer rates
- Literature on BRCA1 and BRCA2 prevalences and penetrances
- Aggregation of breast cancer in a study of daughters of affected mothers
- Relative risks from other risk factors are from various studies, assumed to act multiplicatively
- Other assumptions such as:
  - Familial aggregation from a putative autosomal dominant gene
  - Other risk factors multiply the hazard for the mixed genetic survival distribution

Data Needed for Independent Validation

- **Relative risk features**
  - **Case-control data or cohort data**
- **Area under ROC curve (concordance)**
  - **Age-matched cases and controls**
- **Absolute risk calibration (i.e. whether observed events are close to expected events in various subgroups)**
  - **Cohort data needed (usually a large cohort)**

Summary

- Absolute risk is probability of an event in a defined interval before dying of competing causes
- Follow-up data in a cohort or registry is need to estimate absolute risk
- Various designs have different strengths and weakness
- Cohort needed to check calibration