Conceptual Issues in Risk Prediction

Colin B. Begg

Memorial Sloan-Kettering Cancer Center

Conceptual Issues in Risk Prediction

- What do we mean by the term “risk”?
- Model A – prevention – inherent risk
    - germline genetic susceptibility
    - environmental exposures
- Model B – early detection – probability that a cancer is in the process of development
    - factors reflecting causal events that have occurred: somatic changes, tumor markers, # breast biopsies

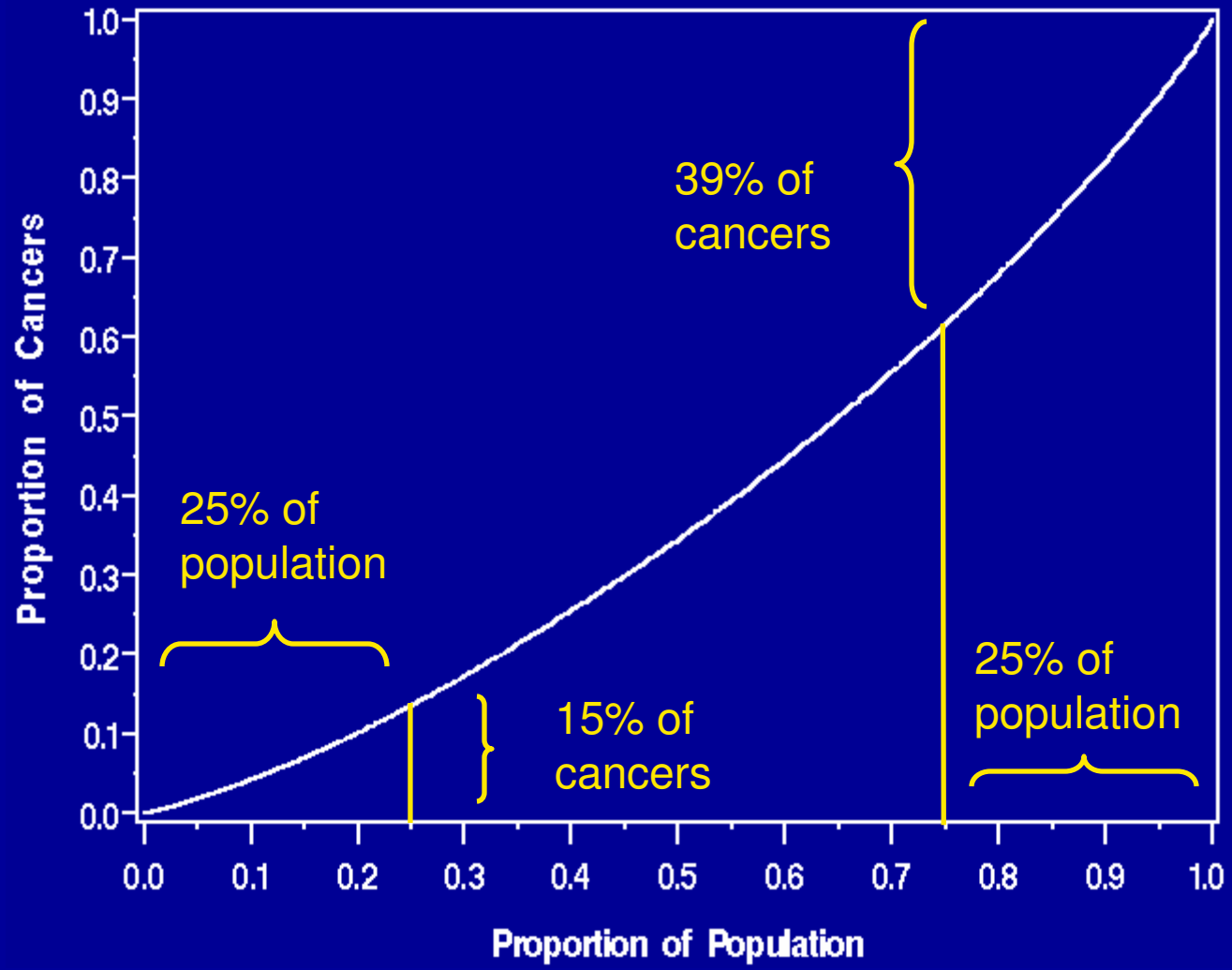
Predictability of a Cancer

- How predictable is cancer occurrence?
- How much does random variation limit prediction?
- Model A – The stochastic (unpredictable) aspects of future cancer events is inherent to the biology of the disease
- Model B – As more evidence of a latent cancer is identified its existence is increasingly predictable

Why Does Predictability Matter?

- Less predictability $\Rightarrow$ broad population-based prevention strategies are appropriate
- More predictability $\Rightarrow$ greater rationale for focussing prevention strategies on high risk individuals.

Distribution of Gail Model Predictors
Nurses Health Study



To What Extent are Future Cancers Inherently Predictable

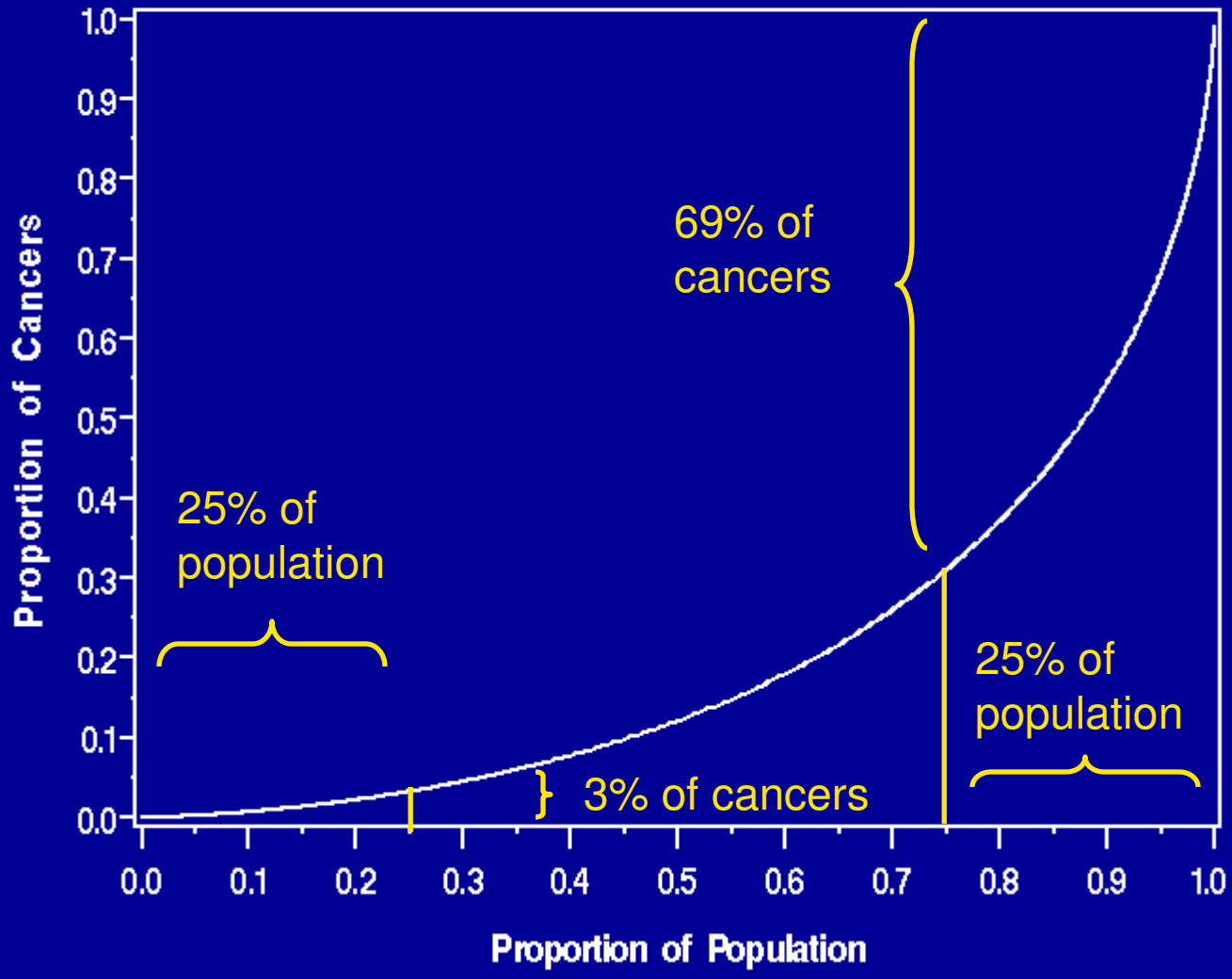
- Inferences on maximal attainable predictability are possible by studying incidence rates of second primary cancers (with some assumptions)

- Genetic predictability can be inferred from studies of familial aggregation (with some assumptions)

Maximum Attainable Predictability

- Each member of population has unique risk, r, density $p(r)$, mean μ
- Density of risks among incident cases = $q(r) = rp(r)/\mu$
- Incidence rate of second primaries = $\int r\, q(r) dr = \mu^{-1} E(r^2)$
- Standardized incidence ratio = $1 + \frac{\mathrm{var}(r)}{\mu^2}$
- Implication: standardized risk variation can be inferred from the standardized incidence ratio, an estimable quantity

Maximum Attainable Predictability
Breast Cancer



Data Sources for Risk Prediction Models

- In practice, models should use known risk factors
- Validation is a pivotal concern
- Issues
  - Absolute risks – generalizability in the context of geographic variations in incidence
  - Power/precision – studies with large number of "events" are ideal
- Implications
  - Aggregates of population-based case-control studies may be the ideal
  - In practice, datasets have been used opportunistically
    - BCDDP (volunteers) for the Gail model
    - CARET study (RCT) for lung prediction model (Bach)
    - BCLC (high risk families) for BRCA penetrance

Models for Enhancing Cancer Screening

- Tumor markers/ proteomics
  -- Can these be used to enhance screening strategies?

- Goal – identify patients who should be triaged to radiology (mammography) to localize/ identify a tumor.

- Predictability is not constrained by fundamental stochastic limitations.

- Just as for mammography, because of:-
  -- uncertain clinical impact
  -- lead time and length bias
  these strategies ultimately need to be evaluated in randomized trials with mortality as the endpoint.