Absolute Risk Models and Applications in Designing Intervention Studies

Ruth Pfeiffer
Pfeiffer@mail.nih.gov
Biostatistics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute, National Institutes of Health, USA
Outline

- Overview of absolute risk
- Examples of absolute risk models
  - Model for breast cancer incidence with modifiable risk factors
- Applications of models in public health interventions/prevention
“Risk”: Relative Risk

Ratio of probability of event in exposed versus a non-exposed group; etiologic studies
Breast Cancer Relative Risk (RR)

50 year old woman

- past diagnosis of benign breast disease
- body mass index (BMI)=30
- consumes 1 alcoholic drink/day
- other risk factors at lowest risk level

Relative risk compared to woman at lowest level of risk for ALL risk factors: RR=1.86
“Risk”: Absolute Risk

– crude risk, cumulative incidence

– \( P(\text{cancer occurs between age } a \text{ and } a+\tau | \text{ at risk at } a \text{ with risk factors } X \text{ in presence of competing risks}) \)

"Pure risk” (1-Kaplan Meier curve)

– \( P(\text{cancer occurs between age } a \text{ and } a+\tau | \text{ at risk at } a \text{ with risk factors } X \text{ and there are no competing risks}) \)
Absolute risk is probability diagnosed with breast cancer in the interval $a$ to $a + \tau$ among women with risk factors $X$.
20 Year Absolute Risk of Breast Cancer

Woman with past diagnosis of benign breast disease, BMI=30, who consumes 1 alcoholic drink/day, other factors at lowest level of risk: RR=1.86

<table>
<thead>
<tr>
<th>Current age</th>
<th>20-Year Absolute risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>6.4%</td>
</tr>
<tr>
<td>65</td>
<td>7.1%</td>
</tr>
</tbody>
</table>
Cause Specific Formulation of Absolute Risk for Person with Risk Factors X

\[ r(x, a, \tau) = P(T \leq a + \tau, \text{cause} = C \mid T > a; x) \]

\[ = \int_{a}^{a+\tau} \exp \left[ -\int_{a}^{t} \left( h_C(u, x) + h_D(u, x) \right) du \right] dt \]

- \( T \) - event time
- \( X \) - individual risk or protective factors
- \( a \) - age at start of projection
- \( \tau \) - length of projection
- \( h_C(t, x) \) - cancer hazard at age \( t \)
- \( h_D(t, x) \) - mortality hazard from competing risks
Absolute Risk Models for Cancer Incidence for the General Population Developed at NCI

- Breast cancer:
  - BC2013 (Pfeiffer et al, PLoS Medicine, 2013)


- Colorectal cancer
  [https://www.cancer.gov/colorectalcancerrisk/](https://www.cancer.gov/colorectalcancerrisk/) (Freedman, ….,Pfeiffer, JCO, 2009)
The Breast Cancer Risk Assessment Tool is an interactive tool designed by scientists at the National Cancer Institute (NCI) and the National Surgical Adjuvant Breast and Bowel Project (NSABP) to estimate a woman's risk of developing invasive breast cancer. See About the Tool for more information.

The Breast Cancer Risk Assessment Tool may be updated periodically as new data or research becomes available.

### Risk Tool

(Click a question number for a brief explanation, or read all explanations.)

1. Does the woman have a medical history of any breast cancer or of ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS) or has she received previous radiation therapy to the chest for treatment of Hodgkin lymphoma?

2. Does the woman have a mutation in either the BRCA1 or BRCA2 gene, or a diagnosis of a genetic syndrome that may be associated with elevated risk of breast cancer?

3. What is the woman's age?
   *This tool only calculates risk for women 35 years of age or older.*

4. What was the woman's age at the time of her first menstrual period?

5. What was the woman's age at the time of her first live birth of a child?

6. How many of the woman's first-degree relatives - mother, sisters, daughters - have had breast cancer?
CRC Risk Web Tool

Risk Calculator
About the Tool
Colorectal Cancer Risk Factors

Quick Links
Colorectal Cancer Risk Assessment Tool - Mozilla
Colorectal Cancer Risk Assessment Tool

Things to know before using this tool:

The Colorectal Cancer Risk Assessment Tool was designed for use by doctors and other health providers with their patients. If you are not a health provider, take these results to your doctor or other health provider to discuss your personal risk of colorectal cancer. (Colorectal cancer is another way to say colon and rectal cancer).

This tool can estimate the risk of colorectal cancer for men and women who are:
- Between the ages of 50 and 85
- African American
- Asian American/Pacific Islander
- Hispanic/Latino
- White

This tool does not yet apply to American Indians and Alaska Natives, but we are working to improve the tool for use by these groups of people. If you are African American, Asian American/Pacific Islander, or Hispanic/Latino, please click here for more information about race, ethnicity, and how we developed this tool.

This tool cannot accurately estimate risk of colorectal cancer for people who have the following problems:
- Ulcerative colitis
- Crohn disease
- Familial adenomatous polyposis (FAP)
- Hereditary Nonpolyposis Colorectal Cancer (HNPCC)
- Personal history of colorectal cancer

This Web site can help you learn more about cancer risk, including colorectal cancer risk. Cancer Risk: Understanding the Puzzle from NCI

It will take about 5 to 8 minutes to answer all the questions and obtain your risk estimate.

Risk Calculator

To estimate your risk, answer the following questions

Do you consider yourself to be Hispanic/Latino?
Other Absolute Risk Models

Cancer incidence models for special populations:

- Second primary thyroid cancer risk among 5-year survivors of childhood cancer  
  (Kovalchik, …., Pfeiffer, JCO, 2012)
- Breast cancer risk for young women treated for Hodgkin lymphoma  
  (Travis et al, JNCI 2005)

Mortality models

- Absolute risk of death from prostate cancer following diagnosis  
  (Albertson, Hanley, Fine, JAMA 2005)
- Models for cause specific mortality (cancer, cardiovascular)  
  (Kovalchik & Pfeiffer, Lifetime Data Analysis, 2014)
Building Absolute Risk Models

- **Outcome:** disease incidence, cause specific mortality
- **Target population**
- **Data sources used for model building**
  - Cohort
    - Prospective selected study population
    - Population-based (e.g. NHANES)
  - Sub-samples of a cohort or population base
    - Nested case-control or case-cohort
  - Population-based case-control combined with incidence data from registries (e.g. SEER)
Outline

- Overview of absolute risk
- Examples of absolute risk models
  - Model for breast cancer incidence with modifiable risk factors
- Applications of models in public health interventions/prevention
Absolute Risk Model for Breast Cancer with Modifyable Risk Factors “BC2013”

- Combined data on white non-Hispanic women ages 50+ from two cohorts (240,712 women; 7,695 breast cancer cases) to estimate relative risks for breast cancer

- Incorporated SEER cancer registry data on incidence

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family history of breast/ovarian cancer</td>
<td>1.39</td>
</tr>
<tr>
<td>Benign breast disease/biopsy</td>
<td>1.40</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>1.18</td>
</tr>
<tr>
<td>Age at first live birth</td>
<td>1.17</td>
</tr>
<tr>
<td>Parity</td>
<td>1.32</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>1.12</td>
</tr>
<tr>
<td>(0, &lt;1, 1+ drinks/day)</td>
<td></td>
</tr>
<tr>
<td>Body mass index (BMI)</td>
<td>1.09</td>
</tr>
<tr>
<td>(&lt;25, 25-30, 30-35, 35+)</td>
<td></td>
</tr>
<tr>
<td>Estrogen plus progestin HRT use</td>
<td>1.40</td>
</tr>
<tr>
<td>(never, 1-9, 10+ years)</td>
<td></td>
</tr>
<tr>
<td>Other HRT use (no/yes)</td>
<td>1.16</td>
</tr>
</tbody>
</table>
## Absolute Risk Estimates: Two 50 Year Old Women

<table>
<thead>
<tr>
<th></th>
<th>Woman 1</th>
<th>Woman 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at first birth</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td># of life births</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Family history breast/ovarian cancer</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Biopsy/benign breast disease</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Menopausal</td>
<td>yes (age 50)</td>
<td>yes (age 50)</td>
</tr>
<tr>
<td>Alcohol consumption (drinks/day)</td>
<td>0</td>
<td>&gt;1</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>24</td>
<td>35</td>
</tr>
<tr>
<td>E+P HRT use+duration</td>
<td>no</td>
<td>yes (5yrs)</td>
</tr>
<tr>
<td>10 year BC2013 absolute risk</td>
<td>1.1%</td>
<td>9.5%</td>
</tr>
</tbody>
</table>
Validated Model in Independent Cohort, Nurses’ Health Study

N=57,907 women ages 51-70 at baseline: Compute risk $r(X,a,T)$ for each woman, given baseline factors $X$, age $a$, and projection period $T$

**Calibration (bias):** Does the model correctly predict observed number of cancers that develop? Observed O=2934 cases, predicted E=2930 cases

$E/O = 1.00$ (95%CI: 0.96–1.04)
Validated Model in Independent Cohort, Nurses’ Health Study

N=57,907 women ages 51-70 at baseline: Compute risk $r(X,a,T)$ for each woman, given baseline factors $X$, age $a$, and projection period $T$

**Calibration (bias):** Does the model correctly predict observed number of cancers that develop?
Observed $O=2934$ cases, predicted $E=2930$ cases

$E/O = 1.00$ (95% CI: 0.96–1.04)

**Discriminatory accuracy:** AUC= probability that randomly selected case has larger predicted risk than randomly selected control: $\text{AUC}=0.58$
Outline

- Overview of absolute risk
- Examples of absolute risk models
  - Model for breast cancer incidence with modifiable risk factors
- Applications of models in public health interventions/prevention
Risk Models in Public Health Prevention

• Designing prevention trials
• Assessing disease burden and potential pay-off from prevention strategies
• “High risk” prevention strategy
• Allocation of preventive resources under cost constraints
Designing Prevention Trials

• Statistical power
  – Depends on the number of events
  – Number of events is proportional to average absolute risk of trial participants
Sample Size Calculation for Prevention Trial

Assume trial with 2 arms:
Baseline incidence $h_0(t)$
Incidence under intervention $h_0(t)\exp(\lambda)$

Test $H_0: \lambda=0$ using log-rank test

Given statistical power $\beta$ and type one error level $\alpha$, number of total events needed is

$$D = \frac{4(z_\alpha + z_\beta)^2}{\lambda^2}$$

Compute $D$ in given population based on absolute risk model for specified follow-up time
Designing Prevention Trials

• **Statistical power**
  – Number of events is proportional to average absolute risk of trial participants

• **Eligibility criteria**
  – Select subjects at high enough risk so they stand to benefit from intervention:
  – Increase efficiency of trial by including high risk subjects (more events for same trial size)

BCPT (P-1) trial for tamoxifen: women ages <60 years with BCRAT 5-year risk >1.66% were eligible

(Fisher et al, 98)
Assessing Burden of Disease from Modifiable Risk Factors

• Breast cancer model includes alcohol consumption, BMI, and hormone replacement therapy and other factors
• Calculate population absolute risk from modifiable factors
• If we eliminated modifiable risk factors, how much would population absolute risk decrease?
• Is this the same as attributable risk?
Assess Disease Preventable Due to Modifiable Risk Factors

Non-modifiable factors \((X)\): e.g. parity, age 1\(^{st}\) birth, benign breast disease/biopsy, …

**Modifiable factors \((Z)\):** BMI, alcohol, HRT use

Mean absolute risk in populations or in groups \(S\)

\[
E[r(a, \tau, X, Z) | (X, Z) \in S]
\]

Mean absolute risk reduction in groups \(S\)

\[
E[r(a, \tau, X, Z) - r(a, \tau, X, Z_0) | (X, Z) \in S]
\]

Fractional mean risk reduction

\[
\frac{E \left[ r(a, \tau, X, Z) \right] - E[ r(a, \tau, X, Z_0) ]}{E \left[ r(a, \tau, X, Z) \right]}
\]

Modifying Alcohol Consumption, HRT use, BMI: Changes in 20-year BC2013 Risk for Women Aged 50 Years in NHS Validation Cohort

<table>
<thead>
<tr>
<th></th>
<th>Initial mean absolute risk (%)</th>
<th>Absolute mean risk reduction (%)</th>
<th>Fractional mean reduction in risk (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire population (N=2447)</td>
<td>5.9</td>
<td>1.0</td>
<td>18.6</td>
</tr>
<tr>
<td>Women with positive family history (N=295)</td>
<td>8.0</td>
<td>1.5</td>
<td>19.0</td>
</tr>
<tr>
<td>Women in top 10% of population risk (N=246)</td>
<td>9.2</td>
<td>2.5</td>
<td>27.2</td>
</tr>
</tbody>
</table>
Attributable Risk (AR)

Binary risk factor $X = 0$ (unexposed), $X = 1$ (unexposed)

$I$ - disease incidence

$$AR = \frac{I - I(X = 0)}{I} = 1 - \frac{I(X = 0)}{I}$$

- Attributable risk is a fractional quantity
- AR is computed using incidence
- Absolute risk is a probability over a time/age interval, e.g. 5 years
- Analogous to fractional mean risk reduction in absence of competing risks
The Strategy of Preventive Medicine¹

• The population strategy of prevention
• The “high-risk” strategy

¹Geoffrey Rose, Oxford University Press, 1992
Prevention Strategies for High Risk Subsets

Compute 5-year risk \( r \) from model for every woman at baseline.

Rank risks from lowest to highest risk: \( r_{(1)} \leq r_{(2)} \leq \ldots \leq r_{(N)} \)

Give prevention only to ‘high risk proportion \( q \)’ of population based on model risks \( r_{(1)} \leq r_{(2)} \leq \ldots \leq r_{(N)} \)
Distribution of Risk in Nurses’ Health Cohort

BC2013 estimates

q=10%
Distribution of Risk in Nurses’ Health Cohort

Only 17% of cases found in top 10% of Nurses’ Health Cohort based on risk estimated from BC2013

If we want to make sure cover 90% of cases we need to give intervention to 84% of population at highest risk from model

Pfeiffer and Gail, Biometrics 2011; Pfeiffer, Biostatistics 2013
Summary

- Well calibrated empirical models of absolute risk useful for:
  - sample size calculations, eligibility criteria for intervention trials
  - assessing potential impact of interventions on population risk

- When models have modest discriminatory ability, they are not useful for selecting high risk subgroups for intervention, as too many cases are missed

- Possible applications for risk models: cost constraints risk-based allocation can get more bang for the buck (better than age based alone)
Selected References

- Rose G. The Strategy of Preventive Medicine, Oxford University Press, 1992
- Freedman et al, JCO 2009;
- Gail MH et al, JNCI 1989;81:1879-1886 (“Gail” model)
- Kovalchik SA et al JCO; 2012.41. 8996, 2012
- Kovalchik SA, RM Pfeiffer, Lifetime data analysis 20 (2), 252-275, 2014
- Petracci, E. et al, JNCI 2011;103:1–12
- Pfeiffer RM et al, Risk Prediction for Breast, Endometrial, and Ovarian Cancer in White Women Aged 50 y or Older: Derivation and Validation from Population-Based Cohort Studies, PloS Medicine, 2013
- Pfeiffer RM, Biostatistics, 2013
- Pfeiffer RM, Gail MH, Biometrics,
- 2011Freedman et al, JCO 2011;19: 2327-2333
- Gail et al, JNCI 1999; 91: 1829-1846
- Garcia-Closas, M. et al, JNCI 2014;106 (prospects)
Collaborators

National Cancer Institute:
Mitchell Gail
Yikyung Park
Aimee Kreimer
Andrew Freedman
Patricia Hartge (ret.)

Harvard University:
Bernard Rosner

City of Hope: James Lacey

University of Utah: Saundra Buys