Towards institution- and investigator-specific self-updating risk calculators for prostate cancer

Donna Pauler Ankerst
Mathematics, Technische Universitaet Muenchen, Urology and Epidemiology/Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, Texas
The Prostate Cancer Prevention Trial (PCPT) Risk Calculator

Thompson, Ankerst et al, NEJM 2004; JNCI 2006

www.prostate-cancer-risk-calculator.com
Based on the provided risk factors a prostate biopsy performed would have a:

- 1% chance of high-grade prostate cancer,
- 8% chance of low-grade cancer,
- 91% chance that the biopsy is negative for cancer.

About 2 to 4% of men undergoing biopsy will have an infection that may require hospitalization.

Please consult your physician concerning these results. Click [here](#) to watch a video overview of these results.
Nominal logistic regression/standard risk factors

PSA: enter prostate-specific antigen in ng/mL
DRE: enter 1 if digital rectal examination is abnormal, 0 otherwise
FAMHIST: enter 1 if there is a first-degree family history of prostate cancer, 0 otherwise
PRIORBIOP: enter 1 if there has been one or more prior biopsies performed (all negative for prostate cancer), 0 otherwise
AA: enter 1 for African American, 0 otherwise
AGE: enter age in years

S1 = -3.002 + 0.256L2PSA + 0.016Age + 0.122AA - 0.455PriorBiop - 0.039DRE + 0.272FamHist

S2 = -7.053 + 0.705L2PSA + 0.048Age + 1.042AA - 0.214PriorBiop + 0.401DRE + 0.225FamHist

Risk of no cancer = 1/[1 + exp(S1) + exp(S2)]
Risk of low-grade cancer = exp(S1)/[1 + exp(S1) + exp(S2)]
Risk of high-grade cancer = exp(S2)/[1 + exp(S1) + exp(S2)]
US National Cancer Institute collection of Cancer Risk Calculators

http://epi.grants.cancer.gov/cancer_risk_prediction/
Risk Calculators

This information is provided by Cleveland Clinic as a convenience service only to physicians and is not intended to replace the physicians' medical advice. Please remember that this information, in the absence of a visit with a physician's patient, must always be considered as an educational service only and are not designed to replace a physician's independent judgment about the appropriateness or risks of a procedure or recommendations for a given patient. CCF makes no representation or warranty concerning the accuracy or reliability of this information and does not warrant the results of using this tool. In no event shall CCF be liable for any damages, direct, indirect, consequential or otherwise, relating to the use of this information or this tool.

Benign Prostatic Hyperplasia:

- Predicting Acute Urinary Retention or Surgical Intervention within 2 Years (with or without Dutasteride)

Bladder Cancer:

- 5-Year Recurrence-Free Survival

Brain Cancer:

- Predicting 6 and 12 month Survival from Brain Metastases

Breast Cancer:

- Predicting Positive Non-Sentinal Lymph Node in Patients with Positive SLN (without Frozen Section Info)
- Predicting Positive Non-Sentinal Lymph Node in Patients with Positive SLN (with Frozen Section Info)
- Predicting Sentinel Lymph Node Metastasis (without Pre-Operative Information)
Completion of randomized trials and studies have brought about a change in the clinical landscape since 2006

Different case-mixes of hospital settings + changes in clinical practice imply constant updates to calculators are necessary (like iphones).

Ongoing discovery, validation and FDA-approval of new biomarkers for clinical practice imply a need to incorporate them into existing calculators rather than collect a new cohort from scratch (like adding a room to a house rather than building a whole new house).
Completion of randomized trials and studies have brought about a change in the clinical landscape since 2006.

Different case-mixes of hospital settings + changes in clinical practice imply constant updates to calculators are necessary (like iPhones).

Ongoing discovery, validation and FDA-approval of new biomarkers for clinical practice imply a need to incorporate them into existing calculators rather than collect a new cohort from scratch (like adding a room to a house rather than building a whole new house).
Prostate Biopsy Collaborative Group (PBCG): in response to urological research community gone out of control

Andrew Vickers
Memorial Sloan-Kettering Cancer Center

Prediction Models: Revolutionary in Principle, But Do They Do More Good Than Harm?
Andrew J. Vickers, Memorial Sloan-Kettering Cancer Center, New York, NY
See accompanying article on page 2959; listen to the podcast by Dr Cooperberg at www.jco.org/podcasts

It can sometimes seem as though we are drowning in prediction models. Every month brings a multitude of newly published risk calculators and nomograms to add to the multitude already in the literature—there are more than 100 prediction models on prostate cancer alone1—and Web sites such as www.nomogram.org, www.nomogram.org, and www.cancer.nomograms.org continue to proliferate. As such, it is easy to become somewhat inviolate to prediction modeling and thus to forget that it constitutes an important shift in the way that medicine is practiced.

prostate cancer have a risk threshold of 20% those who would do no more than four have a risk threshold of 25%. Patients with risks higher than those thresholds from a prediction model would accordingly be advised to consider biopsy.

Similarly, predicted risk can be used to individualize decision making. Given a patient averse to biopsy, the results of a prediction model can be used as part of shared decision making. Informing a patient that his risk is 60% versus 26% is far more conducive to decision making than reporting a PSA of 11 ng/mL versus one of

7 European, 3 US biopsy cohorts
25,772 biopsies from 23,070 patients
8,503 prostate cancers

AIM: Validation is a property of BOTH the prediction tool and the cohort to which it is applied.

Vickers et al., Clinical Cancer Research, 2010

10/7/2013
### Table 1. Description of study cohorts

<table>
<thead>
<tr>
<th>Name of cohort</th>
<th>Location</th>
<th>Type of cohort</th>
<th>Biopsy algorithm</th>
<th>Indication for biopsy</th>
<th>Decision for biopsy a clinical decision?</th>
<th>Biopsy scheme</th>
<th>Prior screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERSPC Göteborg Round 1</td>
<td>Sweden</td>
<td>Screening</td>
<td>Biopsy algorithm</td>
<td>PSA ≥3 ng/mL</td>
<td>No</td>
<td>6-core*</td>
<td>No</td>
</tr>
<tr>
<td>ERSPC Göteborg Rounds 2-6</td>
<td>The Netherlands</td>
<td>Screening</td>
<td>PSA ≥3 ng/mL or ≥4 ng/mL, depending on year</td>
<td>No</td>
<td>6-core*</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Round 1</td>
<td>ERSPC Rotterdam</td>
<td>Screening</td>
<td>PSA ≥3 ng/mL or ≥4 ng/mL†</td>
<td>No</td>
<td>6-core*</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>ERSPC Tarn France</td>
<td>Screening</td>
<td>PSA ≥3 ng/mL</td>
<td>Yes</td>
<td>Primarily 10- to 12-core</td>
<td>Mixture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 1</td>
<td>SABOR San Antonio, TX</td>
<td>Screening</td>
<td>PSA ≥2.5 ng/mL, abnormal DRE, or family history</td>
<td>Yes</td>
<td>10- to 12-core</td>
<td>Mixture</td>
<td></td>
</tr>
<tr>
<td>Cleveland Clinic ProtecT United Kingdom</td>
<td>Screening</td>
<td>PSA ≥3 ng/mL</td>
<td>No</td>
<td>10-core</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyrol Austria</td>
<td>Screening†</td>
<td>PSA ≥1.25 ng/mL, percent free PSA, abnormal DRE</td>
<td>Most men with elevated PSA were biopsied</td>
<td>6-, 10-, or 10- to 15-core†</td>
<td>Mixture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durham VA Durham, NC</td>
<td>Clinical</td>
<td>Elevated PSA, abnormal DRE</td>
<td>Yes</td>
<td>6-, 10-, or 12-core†</td>
<td>Mixture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCPT U.S.</td>
<td>Screening</td>
<td>PSA ≥4ng/mL or abnormal DRE for “for cause” biopsies; end of study biopsy offered to all men</td>
<td>In the case of “for cause” biopsies</td>
<td>Primarily 6-core</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Externally validate the PCPTRC 2.0 by 3 criteria


1.) **Calibration**: How close are predicted risks to observed risks?

2.) **Discrimination**: How well does a risk prediction discriminate between those with and without the disease?

3.) **Clinical net benefit**: Decision-curve analysis that compares the net benefit of using a risk prediction tool to refer patients to biopsy versus referring all or no patients to biopsy (not shown).

There are many more, some such as the Brier score, combine multiple metrics; these 3 are most seen in Urology.
Calibration of PCPTRC 2.0

- Fits the European screening cohorts that similarly use the 6-core biopsy technique on the left.
- Under-fitting for the clinical cohorts on the right—that is to be expected since these men are referred with symptoms and these use a 12-core biopsy.
## Discrimination of PCPTRC 2.0

Area underneath the receiver-operating-characteristic curve (AUC) gives the probability that for a randomly selected cancer case and control, the cancer case would have a higher PCPTRC risk. It varies from 50% (no better than random guessing) to 100% (perfect).

AUC varies from 52% to 68%, a bigger range than any new biomarker has ever pushed an AUC.

We should be worrying more about fixing the cohort effect problem than improving models or how to measure improvement of models...

<table>
<thead>
<tr>
<th>Cohort</th>
<th>AUC for no cancer versus low-grade cancer (%)</th>
<th>AUC for no cancer versus high-grade cancer (%)</th>
<th>AUC for low-grade versus high-grade cancer (%)</th>
<th>Generalized c-index (%) (prevalence weighted AUCs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERSPC Goet. R1</td>
<td>55.6</td>
<td>88.1</td>
<td>77.0</td>
<td>62.9</td>
</tr>
<tr>
<td>ERSPC Goet. R2-6</td>
<td>46.0</td>
<td>74.3</td>
<td>70.8</td>
<td>51.6</td>
</tr>
<tr>
<td>ERSPC Rott. R1</td>
<td>51.8</td>
<td>82.4</td>
<td>76.2</td>
<td>63.8</td>
</tr>
<tr>
<td>ERSPC Rott. R2-3</td>
<td>50.4</td>
<td>74.5</td>
<td>72.3</td>
<td>57.0</td>
</tr>
<tr>
<td>ERSPC Tarn</td>
<td>56.8</td>
<td>74.5</td>
<td>65.9</td>
<td>66.3</td>
</tr>
<tr>
<td>SABOR, US</td>
<td>67.6</td>
<td>71.3</td>
<td>60.8</td>
<td>67.9</td>
</tr>
<tr>
<td>Cleveland Clinic, US</td>
<td>56.8</td>
<td>62.1</td>
<td>61.7</td>
<td>59.6</td>
</tr>
<tr>
<td>ProtecT, UK</td>
<td>57.1</td>
<td>75.9</td>
<td>70.1</td>
<td>64.0</td>
</tr>
<tr>
<td>Tyrol, Austria</td>
<td>60.5</td>
<td>73.0</td>
<td>65.3</td>
<td>64.6</td>
</tr>
<tr>
<td>Durham VA, US</td>
<td>61.4</td>
<td>71.6</td>
<td>66.5</td>
<td>66.1</td>
</tr>
</tbody>
</table>
Can one risk calculator fit all? We don't think so.

Empirical risk curves according to PSA across 11 cohorts in the PBCG

Vickers et al., Clinical Cancer Research, 2010

After adjusting for known risk factors, age, DRE, race, family history, a cohort effect is still significant.

There remains a case-mix effect across different types of hospitals that cannot be explained away by covariates, yet are not the fault of the model (Vergouwe et al, Am J Epidem, 2010).
Another question: what happens when your cohort becomes outdated?

- The PCPT cohort was collected from the late 1990’s through 2004.
- The PCPT protocol for the biopsy procedure was a 6-core sample, but modern practice collects 12- or even more cores.
- It has been documented that a higher number of cores increases the likelihood of detecting cancer and high-grade cancer.
Prostate Biopsy Collaborative Group PBCG 2.0
Cheaper to build a new house if the foundation is too old?

Data elements: Same as before but now ask if ever had a prior PSA test and if it was elevated.
Steyerberg recalibration versus Bayesian methods

To yearly update a risk model:

- Build a new model from scratch
- Recalibration in the large: Use log PCPTRC 2.0 risk as offset and estimate new intercept in nominal logistic regression (NLR)
- Recalibration: NLR to estimate new intercepts and slopes for log PCPTRC 2.0 risk as single covariate
- Revision: Same as recalibration but allow individual risk factors to enter separately as covariates
- Bayesian: Use prior to posterior updating on parameters
- Bayesian likelihood ratio: Use PCPTRC 2.0 as prior odds and update through likelihood ratio on all covariates

Automate it from electronic medical records
Completion of randomized trials and studies have brought about a change in the clinical landscape since 2006

Different case-mixes of hospital settings + changes in clinical practice imply constant updates to calculators are necessary (like iPhones).

Ongoing discovery, validation and FDA-approval of new biomarkers for clinical practice imply a need to incorporate them into existing calculators rather than collect a new cohort from scratch (like adding a room to a house rather than building a whole new house).
Updating an existing risk tool

- Cancer biomarker research is dynamic.
- New markers are discovered/tested/validated.
- Cannot measure these markers retrospectively on the original participants of a cohort.
- For rare genetic markers, large multi-institutional consortiums are required.

Problem to be solved

How to update a risk calculator built on one cohort with a new risk factor measured on a different cohort?

Solution

Bayes theorem
From prior to posterior risk

\[ X = \text{PCPT Risk factors: PSA, DRE, family history, prior biopsy, race, age} \]

\[ Y = \text{New markers} \]

Posterior Odds Cancer(Y,X) = Likelihood Ratio(Y|X) \times \text{Prior Odds Cancer (X)}

within a given strata of \( X \), how much more likely is the new marker to be observed in cases rather than controls; estimated from a separate study to PCPT

\[
\frac{P(\text{Cancer} \mid X, Y)}{P(\text{No Cancer} \mid X, Y)} = \frac{P(Y \mid X, \text{Cancer})}{P(Y \mid X, \text{No Cancer})} \times \frac{P(\text{Cancer} \mid X)}{P(\text{No Cancer} \mid X)}
\]
Single continuous marker

X = PCPT Risk factors: PSA, DRE, family history, prior biopsy
Y = \log(\text{PCA3})

\[
\frac{P(\text{Cancer} \mid X, Y)}{P(\text{No Cancer} \mid X, Y)} = \frac{P(Y \mid X, \text{Cancer})}{P(Y \mid X, \text{No Cancer})} \times \frac{P(\text{Cancer} \mid X)}{P(\text{No Cancer} \mid X)}
\]

Linear regressions of Y on X in cancer cases and controls separately.

\[
\frac{1}{\sqrt{\sigma_{\text{cancer}}^2}} \exp \left\{ -\frac{1}{2\sigma_{\text{cancer}}^2} (Y - \mu_{\text{cancer}})^2 \right\}
\]

\[
\frac{1}{\sqrt{\sigma_{\text{no cancer}}^2}} \exp \left\{ -\frac{1}{2\sigma_{\text{no cancer}}^2} (Y - \mu_{\text{no cancer}})^2 \right\}
\]

\[
\mu_{\text{cancer}} = 1.1926 - 0.0836\log(\text{psa}) + 0.0376\text{age} + 0.1055\text{dre} + 0.0658\text{priorbiop}
\]

\[
\mu_{\text{no cancer}} = -0.6915 - 0.1137\log(\text{psa}) + 0.0577\text{age} - 0.3345\text{dre} + 0.1260\text{priorbiop}
\]

\[
\sigma_{\text{cancer}} = 1.0366
\]

\[
\sigma_{\text{no cancer}} = 1.0191
\]

\[
\exp(\beta'X) = -1.7968 + 0.8488\log(\text{psa}) + 0.2693\text{famhist} + 0.9054\text{dre} - 0.4483\text{priorbiop}
\]

Confidence, prediction intervals for posterior risk by delta rule.
The first validation online shortly after added to website; The impact of publishing algorithms.

Prostate Cancer Detection in the "Grey Area" of Prostate-Specific Antigen Below 10 ng/ml: Head-to-Head Comparison of the Updated PCPT Calculator and Chun's Nomogram, Two Risk Estimators Incorporating Prostate Cancer Antigen 3.


Istituto Nazionale Tumori, Fondazione "G. Pascale," Napoli, Italy.

Abstract

BACKGROUND: Prostate cancer antigen 3 (PCA3) holds promise in diagnosing prostate cancer (PCa), but no consensus has been reached on its clinical use. Multivariable predictive models have shown increased accuracy over individual risk factors.

OBJECTIVE: To compare the performance of the two available risk estimators incorporating PCA3 in the detection of PCa in the "grey area" of prostate-specific antigen (PSA) <10 ng/ml: the updated Prostate Cancer Prevention Trial (PCPT) calculator and Chun's nomogram.

DESIGN, SETTING, AND PARTICIPANTS: Two hundred eighteen patients presenting with an abnormal PSA (excluding those with PSA >10 ng/ml) and/or abnormal digital rectal examination were prospectively enrolled in a multicentre Italian study between October 2008 and October 2009. All patients underwent ≥12-core prostate biopsy.

MEASUREMENTS: PCA3 scores were assessed using the Progensa assay (Gen-Probe, San Diego, CA, USA). Comparisons between the two models were performed using tests of accuracy (area under the receiver operating characteristic curve (AUC-ROC)), calibration plots, and decision curve analysis. Biopsy predictors were identified by univariable and multivariable logistic regression. In addition, performance of PCA3 was analysed through AUC-ROC and predictive values.

RESULTS AND LIMITATIONS: PCa was detected in 73 patients (33.5%). Among predictors included in the models, only PCA3, PSA, and prostate volume retained significant predictive value. AUC-ROC was higher for the updated PCPT calculator compared to Chun's nomogram (79.6% vs 71.5%; p=0.043); however, Chun's nomogram displayed better overall calibration and a higher net benefit on decision curve analysis. Using a probability threshold of 25%, no high-grade cancers would be missed; the PCPT calculator would save 11% of biopsies, missing no cancer, whereas Chun's nomogram would save 22% of avoidable biopsies, although missing 4.1% non-high-grade cancers. The small number of patients may account for the lack of statistical significance in the predictive value of individual variables or model comparison.

CONCLUSIONS: Both Chun's nomogram and the PCPT calculator, by incorporating PCA3, can assist in the decision to biopsy by assignment of an individual risk of PCa, specifically in the PSA levels <10ng/ml.
Incorporating multiple markers  
Ankerst et al Biom J 2012

\[ X = \text{PCPT Risk factors: PSA, DRE, family history, prior biopsy} \]

\[ Y = (\log \%\text{freePSA}, \log [-2]\text{proPSA})' \]

\[
\frac{P(\text{Cancer} \mid X, Y)}{P(\text{No Cancer} \mid X, Y)} = \frac{P(Y \mid X, \text{Cancer})}{P(Y \mid X, \text{No Cancer})} \times \frac{P(\text{Cancer} \mid X)}{P(\text{No Cancer} \mid X)}
\]

\[
\begin{align*}
|\Sigma_{\text{cancer}}|^{-1/2} \exp\left\{-\frac{1}{2} (Y - \mu_{\text{cancer}})' \Sigma_{\text{cancer}}^{-1} (Y - \mu_{\text{cancer}}) \right\} \\
|\Sigma_{\text{no cancer}}|^{-1/2} \exp\left\{-\frac{1}{2} (Y - \mu_{\text{no cancer}})' \Sigma_{\text{no cancer}}^{-1} (Y - \mu_{\text{no cancer}}) \right\}
\end{align*}
\]

\[
\begin{align*}
\mu_{\text{cancer}} &= \begin{bmatrix} 2.667 - 0.365 \log \text{PSA} + 0.0110 \text{Age} \\ 1.385 + 0.627 \log \text{PSA} + 0.006 \text{Age} \end{bmatrix} \\
\Sigma_{\text{cancer}} &= \begin{bmatrix} 0.179 & 0.121 \\ 0.121 & 0.231 \end{bmatrix} \\
\mu_{\text{no cancer}} &= \begin{bmatrix} 3.276 - 0.235 \log \text{PSA} + 0.002 \text{Age} \\ 2.438 + 0.571 \log \text{PSA} - 0.008 \text{Age} \end{bmatrix} \\
\Sigma_{\text{no cancer}} &= \begin{bmatrix} 0.128 & 0.097 \\ 0.097 & 0.188 \end{bmatrix}
\end{align*}
\]

\[ \exp(\beta'X) \]

\[
\begin{align*}
\beta'X &= -1.7968 + 0.8488 \log(\text{psa}) \\
&+ 0.2693 \text{famhist} + 0.9054 \text{dre} \\
&- 0.4483 \text{priorbiop}
\end{align*}
\]

- For more flexibility use multivariate t, skew t, mixtures of skew t distributions
- Extend to more than 2 outcome groups.
**Integrated Discriminative Index:** Proposed in Pencina et al. (2008) for comparing risk prediction tools

\[
IDI = \left( \frac{1}{n_{\text{cancer}}} \sum_{i=1}^{n_{\text{cancer}}} P_{\text{new},i} - \frac{1}{n_{\text{control}}} \sum_{i=1}^{n_{\text{control}}} P_{\text{new},i} \right) - \left( \frac{1}{n_{\text{cancer}}} \sum_{i=1}^{n_{\text{cancer}}} P_{\text{old},i} - \frac{1}{n_{\text{control}}} \sum_{i=1}^{n_{\text{control}}} P_{\text{old},i} \right)
\]

Discrimination slope for risks from the updated calculator \((p_{\text{new}})\)

Discrimination slope for risks from the old (PCPT) calculator \((p_{\text{old}})\)

Evaluated on an external Early Detection Research Network cohort of 575 men yielded an improvement:

\[
IDI = 6.3\% \text{ (95\% confidence interval 3.0\% to 9.6\%).}
\]
Genomewide Association Study SNPS for prostate cancer

Some papers report genotype counts/some allele frequencies; latter can be transformed to genotypes assuming Hardy-Weinberg-Equilibrium.

Risk alleles (RA): higher odds for cancer than non-risk alleles.

Minor alleles (MA): lowest frequency.
### Example: SNP rs100860908, Al Olama et al. 2009

<table>
<thead>
<tr>
<th>freq (freq/n)</th>
<th>TT</th>
<th>TC</th>
<th>CC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 3646)</td>
<td>1913 (0.52)</td>
<td>1457 (0.40)</td>
<td>276 (0.08)</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=3939)</td>
<td>1933 (0.49)</td>
<td>1636 (0.40)</td>
<td>370 (0.09)</td>
</tr>
<tr>
<td><strong>LR (freq case/freq control)</strong></td>
<td>.52/.49 = 1.06</td>
<td>.40/.40 = 1.0</td>
<td>.08/.09 = 0.88</td>
</tr>
</tbody>
</table>

For a new individual with TT on this SNP, his PCPTRC prior odds gets inflated by 1.06 for computing his posterior odds/risk of cancer.
Single nucleotide polymorphisms

X = PCPT Risk factors: PSA, DRE, family history, prior biopsy; we believe that mutations are inherited or occur before X and so do not need to condition on X.

Y = SNP with published genotype or allele frequencies (example T,C).

\[
\frac{P(\text{Cancer} \mid X, Y)}{P(\text{No Cancer} \mid X, Y)} = \frac{P(Y \mid X, \text{Cancer})}{P(Y \mid X, \text{No Cancer})} \times \frac{P(\text{Cancer} \mid X)}{P(\text{No Cancer} \mid X)}
\]

Published GWAS study

\[
\approx \left( \frac{\hat{\pi}_{\text{cancer}}^{2(X)} I(Z=2)}{\hat{\pi}_{\text{no cancer}}^{2}} \right) \times \left( \frac{\hat{\pi}_{\text{cancer}}^{1(X)} I(Z=1)}{\hat{\pi}_{\text{no cancer}}^{1}} \right) \times \left( \frac{\hat{\pi}_{\text{cancer}}^{0(X)} I(Z=0)}{\hat{\pi}_{\text{no cancer}}^{0}} \right),
\]

\[
\exp(\beta'X)
\]

\[
\beta'X = -1.7968 + 0.8488 \log(\text{psa}) + 0.2693 \text{famhist} + 0.9054 \text{dre} - 0.4483 \text{priorbiop}
\]

Z = no.of risk alleles (T).
Multiple SNPs in LD

\[ Y = Y_1, Y_2, ..., Y_r : \text{Multiple SNPs from different studies that are known from the Hapmap not to be in linkage disequilibrium (LD) } \Rightarrow \]

\[
\frac{P(\text{Cancer} \mid X, Y)}{P(\text{No Cancer} \mid X, Y)} = \frac{\prod_{i=1}^{r} P(Y_i \mid \text{Cancer})}{\prod_{i=1}^{r} P(Y_i \mid \text{No Cancer})} \times \frac{P(\text{Cancer} \mid X)}{P(\text{No Cancer} \mid X)}
\]

\[
= \prod_{i=1}^{r} LR_i \times \frac{P(\text{Cancer} \mid X)}{P(\text{No Cancer} \mid X)}
\]

- Multiple SNPs in LD $\rightarrow$ multiply LR’s
- Multiple SNPs not in LD $\rightarrow$ import LD/correlation from the HapMap.
Meta-analysis of SNPs from multiple GWAS studies


SNP rs8102476 (C/T, risk allele C)

Prostate, Lung, Colon and Ovarian Study (US)
Health Professionals Follow-up Study (US)
French Prostate Case-Control Study
Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (FI)
American Cancer Society Cancer Prevention Study II (US)
Spain (Zaragoza Hospital)
Vanderbilt University Medical Center and the VA Tennessee Valley Healthcare System (US)
Urology Outpatient Clinic of the Radboud University and Comprehensive Cancer Center (N)
Finland (Tampere University Hospital)
Pathology Core of Northwestern University's Prostate Cancer Specialized Program of Research Excellence (US)
Icelandic Cancer Registry

Genotype
- CC
- CT
- TT

Meta Analysis

LR = Probability of Genotype among Cases / Probability of Genotype among Controls
### Comparison of SNPs with self-report family history

<table>
<thead>
<tr>
<th>Marker</th>
<th>No. controls (%)</th>
<th>No. cases (%)</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs16901979 (No. allele A)</td>
<td>37848</td>
<td>2936</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>34799 (91.9)</td>
<td>2572 (87.6)</td>
<td>0.96</td>
</tr>
<tr>
<td>1</td>
<td>2985 (7.9)</td>
<td>351 (12.0)</td>
<td>1.53</td>
</tr>
<tr>
<td>2</td>
<td>64 (0.2)</td>
<td>13 (0.4)</td>
<td>2.54</td>
</tr>
<tr>
<td>No. FDR prostate cancer &lt; 60 years</td>
<td>303990</td>
<td>23630</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>302839 (99.6)</td>
<td>23407 (99.1)</td>
<td>0.99</td>
</tr>
<tr>
<td>1</td>
<td>1141 (0.4)</td>
<td>221 (0.9)</td>
<td>2.49</td>
</tr>
<tr>
<td>≥ 2</td>
<td>10 (0.01)</td>
<td>2 (0.01)</td>
<td>2.57</td>
</tr>
</tbody>
</table>

- **FDR**: first-degree relative; from Swedish Family-Cancer Database
- **SNP LR from meta-analysis of 3 GWAS studies**
The likelihood method allows addition of new SNPs or replacement of new LRs as more GWAS studies finish. However, projections show future SNP effects will be smaller and never compete with existing risk factors.
We are not the first to compartmentalize models for easy updating from multiple sources. **Gail et al, JNCI 1989** did this for the first online risk tool and has implemented a frequentist approach to incorporate SNPs. His frequentist approach has been replicated for colorectal and lung cancer.
Acknowledgements

TUM: Josef Hoefler, Sonja Grill
UTHSCSA: Russell MacShane, Robin Leach, Ian M. Thompson

Bordeaux