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

Individualized Risk Assessment of Prostate Cancer

Disclaimer

The Prostate Cancer Prevention Trial Prostate Cancer Risk Calculator (PCPTRC) was developed based upon 5519 men in the placebo group of the Prostate Cancer Prevention Trial. All of these 5519 men initially had a prostate-specific antigen (PSA) value less than or equal to 3.0 ng/ml and were followed for seven years with annual PSA and digital rectal examination (DRE). If PSA exceeded 4.0 ng/ml or if an abnormal DRE was noted, a biopsy was recommended. After seven years, all men were recommended to have a prostate biopsy, regardless of PSA or DRE findings. PSA, family history, DRE findings, and history of a prior negative prostate biopsy provided independent predictive value to the calculation of risk of a biopsy that showed presence of cancer.

The results of the PCPTRC may not apply to different groups of individuals. As about 80% of men had a prostate biopsy with six cores, if more than six cores are obtained at biopsy, a greater risk of cancer may be expected. Most men in this study were white and results may be different with other ethnicities or races.

The calculator is in principle only applicable to men under the following restrictions:

- Age 55 or older
- No previous diagnosis of prostate cancer
- DRE and PSA results less than 1 year old

The PCPTRC is applicable for men who are undergoing prostate cancer screening with PSA and DRE as it was derived from a group of men in the Prostate Cancer Prevention Trial who underwent annual PSA and DRE screening. The risk estimate from the calculator does not reflect an endorsement of either PSA or DRE for screening for prostate cancer.

This calculator is designed to provide a preliminary assessment of risk of prostate cancer if a prostate biopsy is performed. Additional clinical information may modify this risk. No specific level of risk is recommended for prostate biopsy and this decision should be an individual choice based upon a physician-patient relationship.

The original PCPTRC was developed and validated using six pieces of information: PSA, age, DRE, race/ethnicity, any history of a prior prostate biopsy, and family history of prostate cancer. It was subsequently also extended in the Prostate Cancer Prevention Trial to include whether the individual is taking finasteride.

Subsequent to this, additional tests have been found to modify levels of risk of prostate cancer in individual men. For example, body mass index (a measure of obesity), has been
d or urine tests including percent free PSA, proPSA, and PCA3. Although these tests were not performed in
se tests have been demonstrated to improve, in some men, the diagnosis of PSA and the other risk factors.

f the test results. A physician must request these tests and would be best informed as to which patients
provided to assist physicians and their patients with the interpretation of these results and the inclusion of
r the original PCPTRC.

kerst@uthscsa.edu.

Continue to Calculator

f, Lucia MS, Feng Z, Parnes HL, Coltman CA Jr. Assessing prostate cancer risk: Results from the Prostate Cancer 006.

Enter Your Information

Race	Caucasian
Age	60
PSA Level ?	1.0 ng/ml
Family History of Prostate Cancer ?	No
Digital Rectal Examination ?	Normal
Prior Prostate Biopsy ?	Never Had A Biopsy

Calculate Cancer Risk

WE MAKE LIVES BETTER™
UT HEALTH SCIENCE CENTER™
SAN ANTONIO

www.prostate-cancer-risk-calculator.com

PCPTRC 1.0

Web Counter

Results

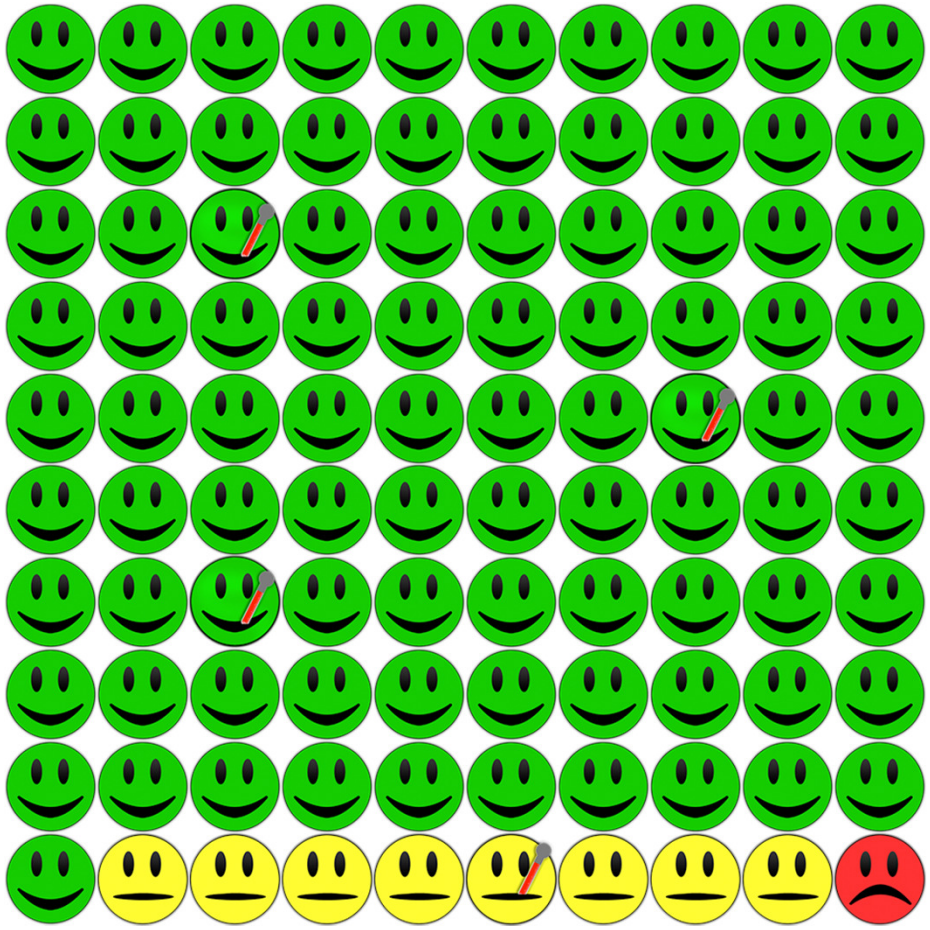
Based on the data provided, the person's estimated risk of biopsy-detectable prostate cancer is **14.2%**.

The **95%** Confidence Interval for this prediction is **13%** to **15.4%**.
More information about the confidence interval

The person's estimated risk of biopsy-detectable high grade prostate cancer is **1.2%**.

The **95%** Confidence Interval for this prediction is **0.8%** to **1.6%**.
More information about the confidence interval

PCPTRC 2.0



Please consult your physician concerning these results. Click **here** to watch a video overview of these results.

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.

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 Prediction Models

The following risk prediction models are grouped by cancer site and whether their methodology and results have been peer-reviewed.

Bibliography of Peer-Reviewed Risk Prediction Models

- Bladder cancer
- Breast cancer
- Cervical cancer
- Colorectal cancer
- Lung cancer
- Melanoma
- Other cancers or multiple sites
- Ovarian cancer
- Pancreatic cancer
- Prostate cancer
- Testicular cancer

Online Risk Prediction Tools and Calculators

- Breast Cancer
  - breastcancerprevention.org Risk Prediction Tool
  - Halls Breast Cancer Risk Calculator
  - Susan Komen Foundation Risk Factors Table
  - womenshealth.org Risk Assessment Tool
- Colorectal Cancer
  - Cleveland Clinic Colorectal Cancer Risk Assessment Tool
  - Dana-Farber Genetic Risk Model
- Prostate Cancer
  - Prostate Cancer Canada Assessment Tools
  - Prostate Cancer Canada Risk Assessment Quiz
  - RealAge Health Assessment Prostate Cancer Risk Questionnaire
- Multiple Sites
  - Cancerfacts.com-NexProfiler Tool
  - Central PA Oncology Group Risk Assessment Questionnaire
  - Health 24 Cancer Risk Calculator
  - MyGenerations by Evanston Northwestern Healthcare Center for Medical Genetics
  - Siteman Cancer Center Tool
  - University of Texas Southwestern Medical Center Web site
  - Women's Cancer Network Cancer Risk Assessment Survey

Cleveland Clinic

http://www.lerner.ccf.org/qhs/risk_calculator/

Mike Kattan
Cleveland Clinic

Cleveland Clinic LERNER RESEARCH INSTITUTE Home Education Research Services

QUANTITATIVE HEALTH SCIENCES

QHS Home

Services

- Request Collaboration
- Skill Finder
- Grant Writing Services
- Statistical Software/Code
- Clinical Risk Calculators
- Courses We Teach

Contact

- Department Contact Information
- People Directory
- Request Collaboration
- Job Postings

Sections

- Biostatistics Section
- Clinical Trials Design and Analysis Section
- Health Outcomes Research and Clinical Epidemiology Section
- Statistical Genetics and Bioinformatics Section
- Research Informatics
- Regulatory Compliance Office

Research Activities

- Most Recent Publications
- 2011 Publications (PDF)

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.

Make Your Own!

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

VOLUME 29 · NUMBER 22 · AUGUST 1 2011
JOURNAL OF CLINICAL ONCOLOGY E D I T O R I A L S

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 alone[1]—and Web sites such as www.nomograms.org, www.nomogram.org, and www.cancernomograms.org continue to proliferate. As such, it is easy to become somewhat inured 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

Table 1. Description of study cohorts

Name of cohort	Location	Type of cohort	Biopsy algorithm		Biopsy scheme	Prior screening
			Indication for biopsy	Decision for biopsy a clinical decision?		
ERSPC Göteborg Round 1	Sweden	Screening	PSA ≥3 ng/mL	No	6-core*	No
ERSPC Göteborg Rounds 2-6	Sweden	Screening	PSA ≥3 ng/mL	No	6-core*	Yes
ERSPC Rotterdam	The Netherlands	Screening	PSA ≥3 ng/mL or ≥4 ng/mL, depending on year	No	6-core*	No
ERSPC Rotterdam Rounds 2-3	The Netherlands	Screening	PSA ≥3 ng/mL or ≥4 ng/mL†	No	6-core*	Yes
ERSPC Tarn Round 1	France	Screening	PSA ≥3 ng/mL	Yes	Primarily 10- to 12-core	Mixture
SABOR	San Antonio, TX	Screening	PSA ≥2.5 ng/mL, abnormal DRE, or family history	Yes	10- to 12-core	Mixture
Cleveland Clinic	Cleveland, OH	Clinical	Elevated PSA, abnormal DRE, rapid rise in PSA	Yes	Primarily ≥8-core	Mixture
ProtecT	United Kingdom	Screening‡	PSA ≥3 ng/mL	No	10-core	No
Tyrol	Austria	Screening‡	PSA ≥1.25 ng/mL, percent free PSA, abnormal DRE	Most men with elevated PSA were biopsied	6-, 10-, or 10- to 15-core†	Mixture
Durham VA	Durham, NC	Clinical	Elevated PSA, abnormal DRE	Yes	6-, 10-, or 12-core †	Mixture
PCPT	U.S.	Screening	PSA ≥4ng/mL or abnormal DRE for "for cause" biopsies; end of study biopsy offered to all men	In the case of "for cause" biopsies	Primarily 6-core	Yes

Externally validate the PCPTRC 2.0 by 3 criteria

Steyerberg E. Clinical Prediction Models, Springer, 2009

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

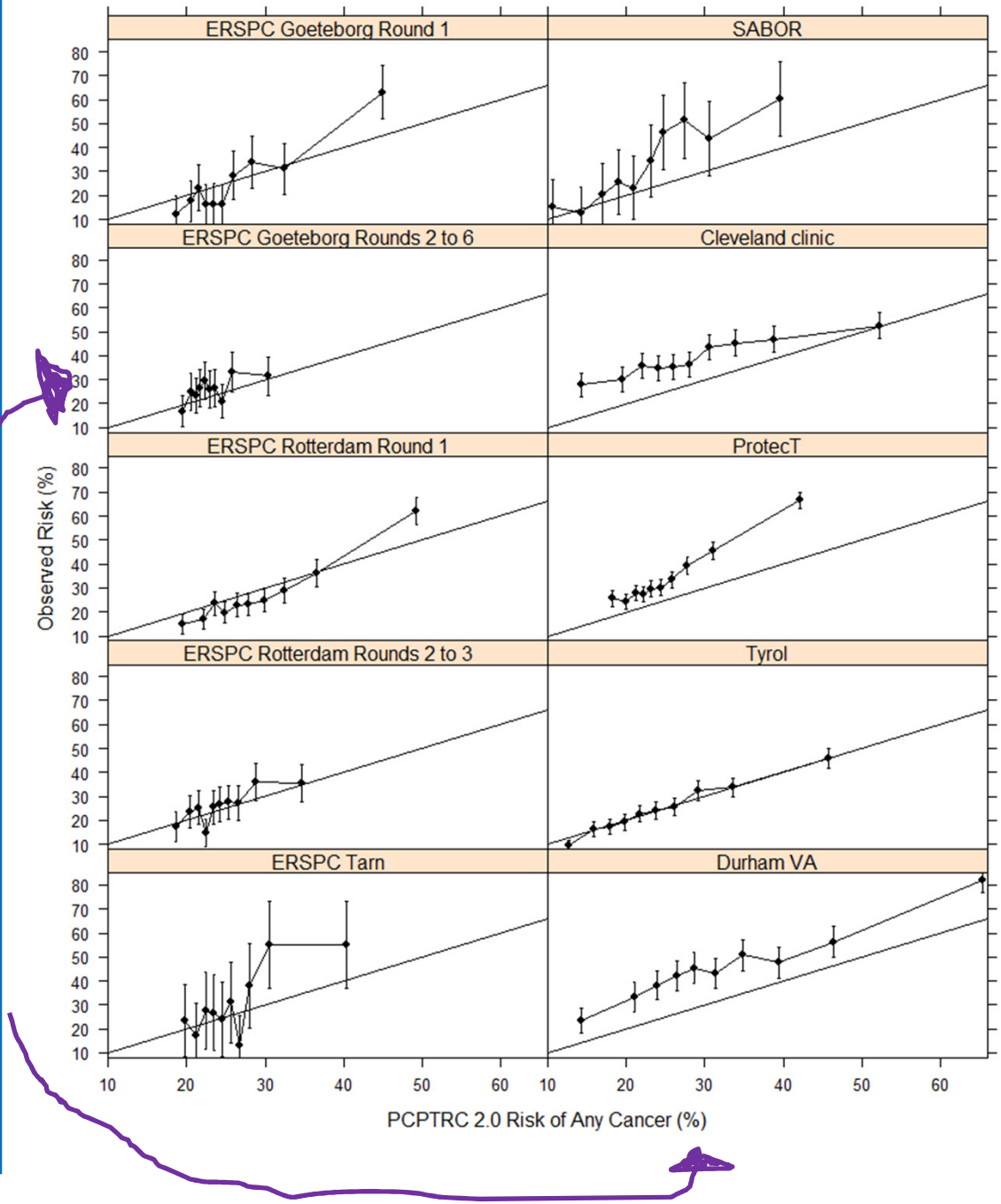
2.) **Discrimination**: How well does a risk prediction discriminate between those with and without the disease?

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

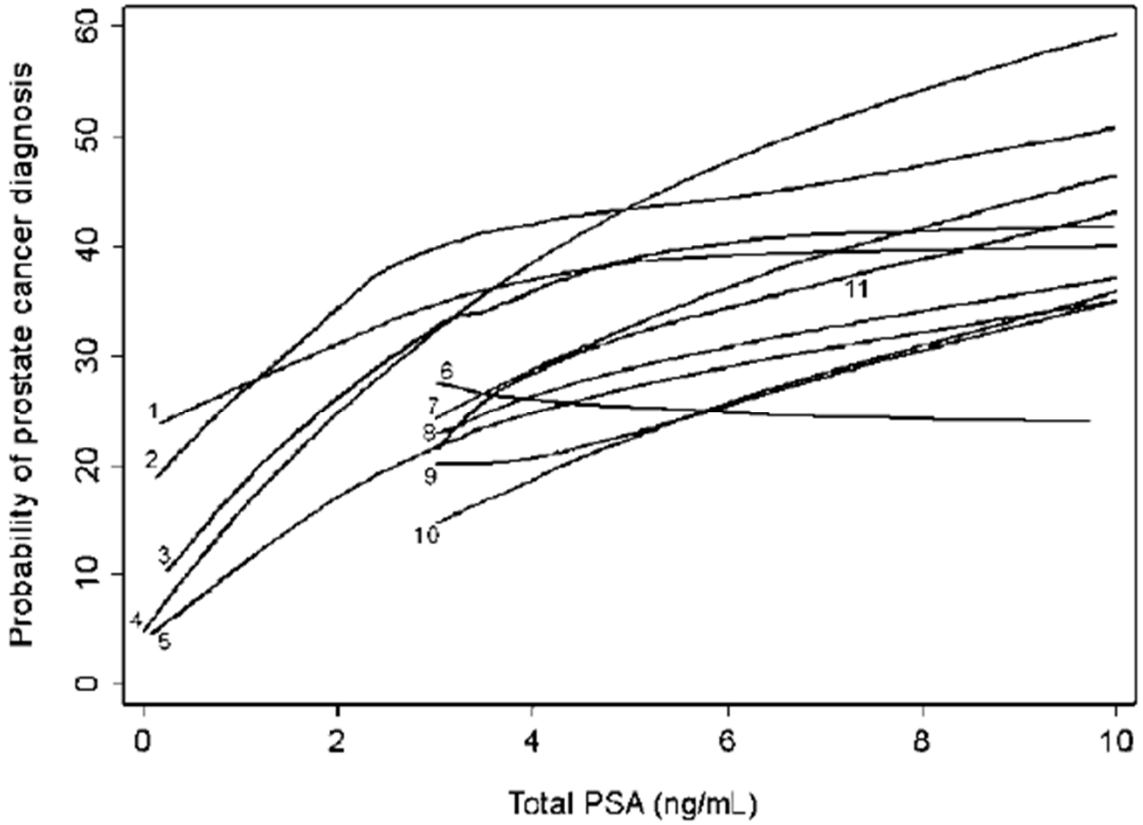
- Area underneath the receiver-operating-characteristic curve (AUC) gives the probability that for a randomly selected cancer case and control, the cancer case would have a higher PCPTRC risk. It varies from 50% (no better than random guessing) to 100% (perfect).
- AUC varies from 52% to 68%, a bigger 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…

Cohort	AUC for no cancer versus low-grade cancer (%)	AUC for no cancer versus high-grade cancer (%)	AUC for low-grade versus high-grade cancer (%)	Generalized c-index (%) (prevalence weighted AUCs)
ERSPC Goet. R1	55.6	88.1	77.0	62.9
ERSPC Goet. R2-6	46.0	74.3	70.8	51.6
ERSPC Rott. R1	51.8	82.4	76.2	63.8
ERSPC Rott. R2-3	50.4	74.5	72.3	57.0
ERSPC Tarn	56.8	74.5	65.9	66.3
SABOR, US	67.6	71.3	60.8	67.9
Cleveland Clinic, US	56.8	62.1	61.7	59.6
ProtecT , UK	57.1	75.9	70.1	64.0
Tyrol, Austria	60.5	73.0	65.3	64.6
Durham VA, US	61.4	71.6	66.5	66.1

Can one risk calculator fit all? We don't think so.

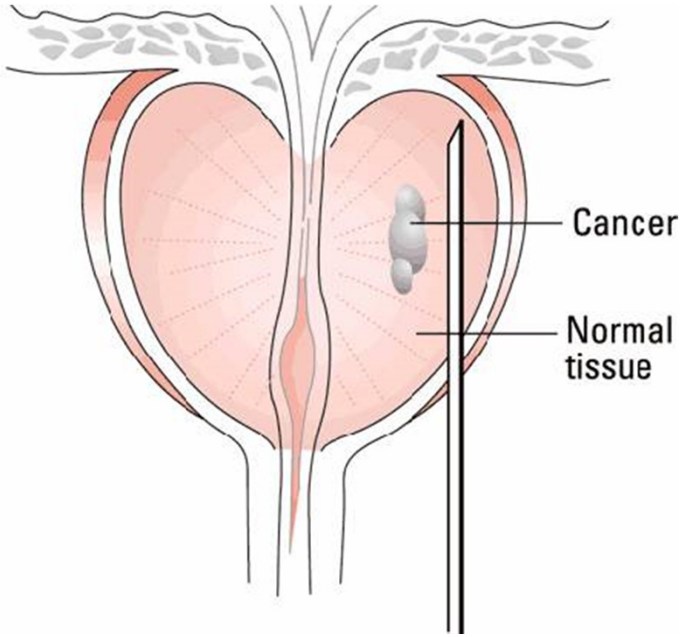
Empirical risk curves according to PSA across 11 cohorts in the PBCG

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).

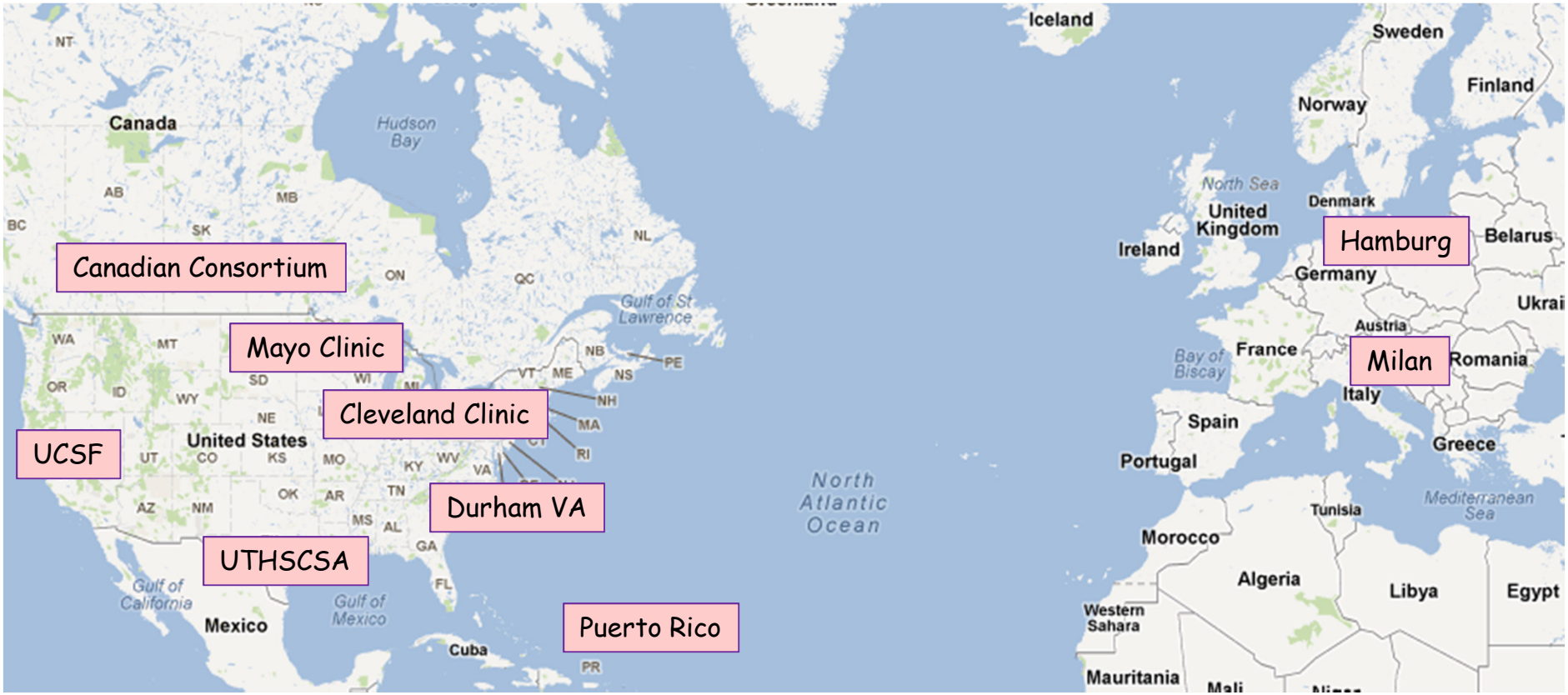
Another question: what happens when your cohort becomes outdated?



- 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.

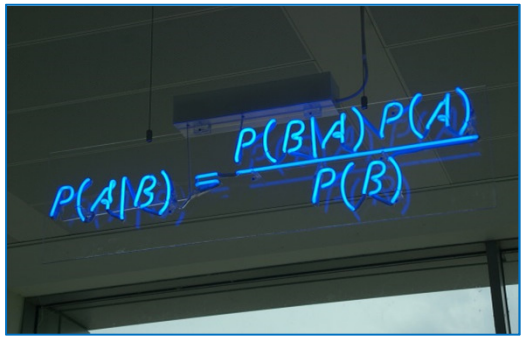
Problem to be solved

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

Solution

Bayes theorem

From prior to posterior risk



X = PCPT Risk factors: PSA, DRE, family history, prior biopsy, race, age
Y = New markers

Posterior Odds Cancer(Y,X) = Likelihood Ratio(Y|X) × 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

from PCPT risk calculator

$$\frac{P(Cancer \mid X, Y)}{P(No\ Cancer \mid X, Y)} = \frac{P(Y \mid X, Cancer)}{P(Y \mid X, No\ Cancer)} \times \frac{P(Cancer \mid X)}{P(No\ Cancer \mid X)}$$

Single continuous marker

Ankerst et al J Urol 2009

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

$$\frac{\mathrm{P(Cancer \mid X, Y)}}{\mathrm{P(No\ Cancer \mid X, Y)}} = \frac{\mathrm{P(Y \mid X, Cancer)}}{\mathrm{P(Y \mid X, No\ Cancer)}} \times \frac{\mathrm{P(Cancer \mid X)}}{\mathrm{P(No\ Cancer \mid X)}}$$

$$\frac{\frac{1}{\sqrt{\sigma^2_{cancer}}}\exp\left\{-\frac{1}{2\sigma^2_{cancer}}(Y-\mu_{cancer})^2\right\}}{\frac{1}{\sqrt{\sigma^2_{no\ cancer}}}\exp\left\{-\frac{1}{2\sigma^2_{no\ cancer}}(Y-\mu_{no\ cancer})^2\right\}}$$

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

PCPTRC 1.0 logistic regression

$$\exp(\beta'X)$$

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

$$\mu_{cancer} = 1.1926 - .0836\log(\mathrm{psa}) + .0376\mathrm{age} + .1055\mathrm{dre} + .0658\mathrm{priorbiop}$$

$$\mu_{no\ cancer} - .6915 - .1137\log(\mathrm{psa}) + .0577\mathrm{age} - .3345\mathrm{dre} + .1260\mathrm{priorbiop}$$

$$\sigma_{cancer} = 1.0366$$

$$\sigma_{no\ cancer} = 1.0191$$

Confidence,prediction intervals for posterior risk by delta rule.

The first validation online shortly after added to website; The impact of publishing algorithms.

Eur Urol. 2010 Oct 12. [Epub ahead of print]

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.

Perdonà S, Cavadas V, Di Lorenzo G, Damiano R, Chiappetta G, Del Prete P, Franco R, Azzarito G, Scala S, Arra C, De Sio M, Autorino R.

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 = PCPT Risk factors: PSA, DRE, family history, prior biopsy

Y = (log %freePSA, log [-2]proPSA)′

$$\frac{\mathrm{P}(\mathrm{Cancer} \mid \mathrm{X}, \mathrm{Y})}{\mathrm{P}(\mathrm{No\ Cancer} \mid \mathrm{X}, \mathrm{Y})} = \frac{\mathrm{P}(\mathrm{Y} \mid \mathrm{X}, \mathrm{Cancer})}{\mathrm{P}(\mathrm{Y} \mid \mathrm{X}, \mathrm{No\ Cancer})} \times \frac{\mathrm{P}(\mathrm{Cancer} \mid \mathrm{X})}{\mathrm{P}(\mathrm{No\ Cancer} \mid \mathrm{X})}$$

SABOR

$$\frac{|\Sigma_{cancer}|^{-1/2} \exp\left\{-\frac{1}{2}(Y-\mu_{cancer})'\Sigma_{cancer}^{-1}(Y-\mu_{cancer})\right\}}{|\Sigma_{no\ cancer}|^{-1/2} \exp\left\{-\frac{1}{2}(Y-\mu_{no\ cancer})'\Sigma_{no\ cancer}^{-1}(Y-\mu_{no\ cancer})\right\}}$$

$$\mu_{cancer} = \begin{bmatrix} 2.667 - 0.365 \log PSA + 0.0110 Age \\ 1.385 + 0.627 \log PSA + 0.006 Age \end{bmatrix}$$

$$\Sigma_{cancer} = \begin{bmatrix} 0.179 & 0.121 \\ 0.121 & 0.231 \end{bmatrix}$$

$$\mu_{no\ cancer} = \begin{bmatrix} 3.276 - 0.235 \log PSA + 0.002 Age \\ 2.438 + 0.571 \log PSA - 0.008 Age \end{bmatrix}$$

$$\Sigma_{no\ cancer} = \begin{bmatrix} 0.128 & 0.097 \\ 0.097 & 0.188 \end{bmatrix}.$$

PCPT

$$\exp(\beta' X)$$

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

- 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

$$IDI = \left(\frac{1}{n_{cancer}} \sum_{i=1}^{n_{cancer}} p_{new,i} - \frac{1}{n_{control}} \sum_{i=1}^{n_{control}} p_{new,i} \right) - \left(\frac{1}{n_{cancer}} \sum_{i=1}^{n_{cancer}} p_{old,i} - \frac{1}{n_{control}} \sum_{i=1}^{n_{control}} p_{old,i} \right)$$

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

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

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

IDI = 6.3% (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.

SNP	Chr.[1]	Risk allele	Non-Risk allele	RA=MA[2]	Study
rs10187424	2	A	G		Kote-Jarai et al., 2011
rs12621278	2	A	G		Eeles et al., 2009
rs1465618	2	A	G	yes	Eeles et al., 2009
rs721048	2	A	G[3]	yes	Gudmundsson et al., 2008
rs10934853	3	A	C[3]	yes	Gudmundsson et al., 2009
rs10936632	3	A	C		Kote-Jarai et al., 2011
rs2660753	3	T	C	yes	Eeles et al., 2008
rs6763931	3	T	C	yes	Kote-Jarai et al., 2011
rs12500426	4	A	C	yes	Eeles et al., 2009
rs17021918	4	C	T		Eeles et al., 2009
rs7679673	4	C	A		Eeles et al., 2009
rs2121875	5	G	T	yes	Kote-Jarai et al., 2011
rs130067	6	G	T	yes	Kote-Jarai et al., 2011
rs9364554	6	T	C	yes	Eeles et al., 2008
rs10486567	7	G	A		Thomas et al., 2008
rs6465657	7	C	T	yes	Eeles et al., 2008
rs10086908	8	T	C		Al Olama et al., 2009
rs10090154	8	T	C	yes	Al Olama et al., 2009
rs1016343	8	T	C	yes	Al Olama et al., 2009
rs12543663	8	C	A	yes	Al Olama et al., 2009
rs13252298	8	A	G		Al Olama et al., 2009
rs1447295	8	A	C[3]	yes	Amundadottir et al., 2006
rs1512268	8	A	G	yes	Eeles et al., 2009
rs16901979	8	A	C[3]	yes	Gudmundsson et al., 2007a
rs16902094	8	G	A[3]	yes	Gudmundsson et al., 2009
rs2928679	8	T	C	yes	Eeles et al., 2009
rs445114	8	T	C[3]	yes	Gudmundsson et al., 2009
rs620861	8	C	T		Al Olama et al., 2009
rs6983267	8	G	T		Yeager et al., 2007
rs6983561	8	C	A	yes	Al Olama et al., 2009
rs10993994	10	T	C	yes	Eeles et al., 2008
rs4962416	10	C	T	yes	Thomas et al., 2008
rs10896449	11	G	A		Thomas et al., 2008
rs12418451	11	A	G		Zheng et al., 2009
rs7127900	11	A	G	yes	Eeles et al., 2009
rs7931342	11	G	T		Eeles et al., 2008
rs10875943	12	C	T	yes	Kote-Jarai et al., 2011
rs11649743	17	G	A		Sun et al., 2008
rs1859962	17	G	T	yes	Gudmundsson et al., 2007b
rs4430796	17	A	G	yes	Gudmundsson et al., 2007b
rs2735839	19	G	A		Eeles et al., 2008
rs8102476	19	C	T[3]	yes	Gudmundsson et al., 2009
rs9623117	22	C	T	yes	Sun et al., 2009
rs5759167	X	G	T		Eeles et al., 2009
rs5919432	X	A	G		Kote-Jarai et al., 2011
rs5945572	X	A	G[3]	yes	Gudmundsson et al., 2008
rs5945619	X	C	T	yes	Eeles et al., 2008

[1] chromosome on which SNP is located

Example: SNP rs100860908, Al Olama et al. 2009

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

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.

SNP	Chr.[1]	Risk allele	Non-Risk allele	RA=MA[2]	Study
rs10187424	2	A	G		Kote-Jarai et al., 2011
rs12621278	2	A	G		Eeles et al., 2009
rs1465618	2	A	G	yes	Eeles et al., 2009
rs721048	2	A	G[3]	yes	Gudmundsson et al., 2008
rs10934853	3	A	C[3]	yes	Gudmundsson et al., 2009
rs10936632	3	A	C		Kote-Jarai et al., 2011
rs2660753	3	T	C	yes	Eeles et al., 2008
rs6763931	3	T	C	yes	Kote-Jarai et al., 2011
rs12500426	4	A	C	yes	Eeles et al., 2009
rs17021918	4	C	T		Eeles et al., 2009
rs7679673	4	C	A		Eeles et al., 2009
rs2121875	5	G	T	yes	Kote-Jarai et al., 2011
rs130067	6	G	T	yes	Kote-Jarai et al., 2011
rs9364554	6	T	C	yes	Eeles et al., 2008
rs10486567	7	G	A		Thomas et al., 2008
rs6465657	7	C	T	yes	Eeles et al., 2008
rs10086908	8	T	C		Al Olama et al., 2009
[illegible]	8	T	C	yes	Al Olama et al., 2009
rs1016343	8	T	C	yes	Al Olama et al., 2009
rs12543663	8	C	A	yes	Al Olama et al., 2009
rs13252298	8	A	G		Al Olama et al., 2009
rs1447295	8	A	C[3]	yes	Amundadottir et al., 2006
rs1512268	8	A	G	yes	Eeles et al., 2009
rs16901979	8	A	C[3]	yes	Gudmundsson et al., 2007a
rs16902094	8	G	A[3]	yes	Gudmundsson et al., 2009
rs2928679	8	T	C	yes	Eeles et al., 2009
rs445114	8	T	C[3]	yes	Gudmundsson et al., 2009
rs620861	8	C	T		Al Olama et al., 2009
rs6983267	8	G	T		Yeager et al., 2007
rs6983561	8	C	A	yes	Al Olama et al., 2009
rs10993994	10	T	C	yes	Eeles et al., 2008
rs4962416	10	C	T	yes	Thomas et al., 2008
rs10896449	11	G	A		Thomas et al., 2008
rs12418451	11	A	G		Zheng et al., 2009
rs7127900	11	A	G	yes	Eeles et al., 2009
rs7931342	11	G	T		Eeles et al., 2008
rs10875943	12	C	T	yes	Kote-Jarai et al., 2011
rs11649743	17	G	A		Sun et al., 2008
rs1859962	17	G	T	yes	Gudmundsson et al., 2007b
rs4430796	17	A	G	yes	Gudmundsson et al., 2007b
rs2735839	19	G	A		Eeles et al., 2008
rs8102476	19	C	T[3]	yes	Gudmundsson et al., 2009
rs9623117	22	C	T	yes	Sun et al., 2009
rs5759167	X	G	T		Eeles et al., 2009
rs5919432	X	A	G		Kote-Jarai et al., 2011
rs5945572	X	A	G[3]	yes	Gudmundsson et al., 2008
rs5945619	X	C	T	yes	Eeles et al., 2008

[1] chromosome on which SNP is located

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

PCPT

$$\frac{\left(\pi^{TT}_{cancer}\right)^{I(Y=TT)}\left(\pi^{TC,CT}_{cancer}\right)^{I(Y=TC,CT)}\left(\pi^{CC}_{cancer}\right)^{I(Y=CC)}}{\left(\pi^{TT}_{no\ cancer}\right)^{I(Y=TT)}\left(\pi^{TC,CT}_{no\ cancer}\right)^{I(Y=TC,CT)}\left(\pi^{CC}_{no\ cancer}\right)^{I(Y=CC)}}$$

$$\approx \left(\frac{\hat{\pi}^{2}_{cancer}}{\hat{\pi}^{2}_{no\ cancer}}\right)^{I(Z=2)}\left(\frac{\hat{\pi}^{1}_{cancer}}{\hat{\pi}^{1}_{no\ cancer}}\right)^{I(Z=1)}\left(\frac{\hat{\pi}^{0}_{cancer}}{\hat{\pi}^{0}_{no\ cancer}}\right)^{I(Z=0)},$$

$$Z = \text{no.of risk alleles } (T).$$

$$\exp(\beta' X)$$

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

Multiple SNPs in LD

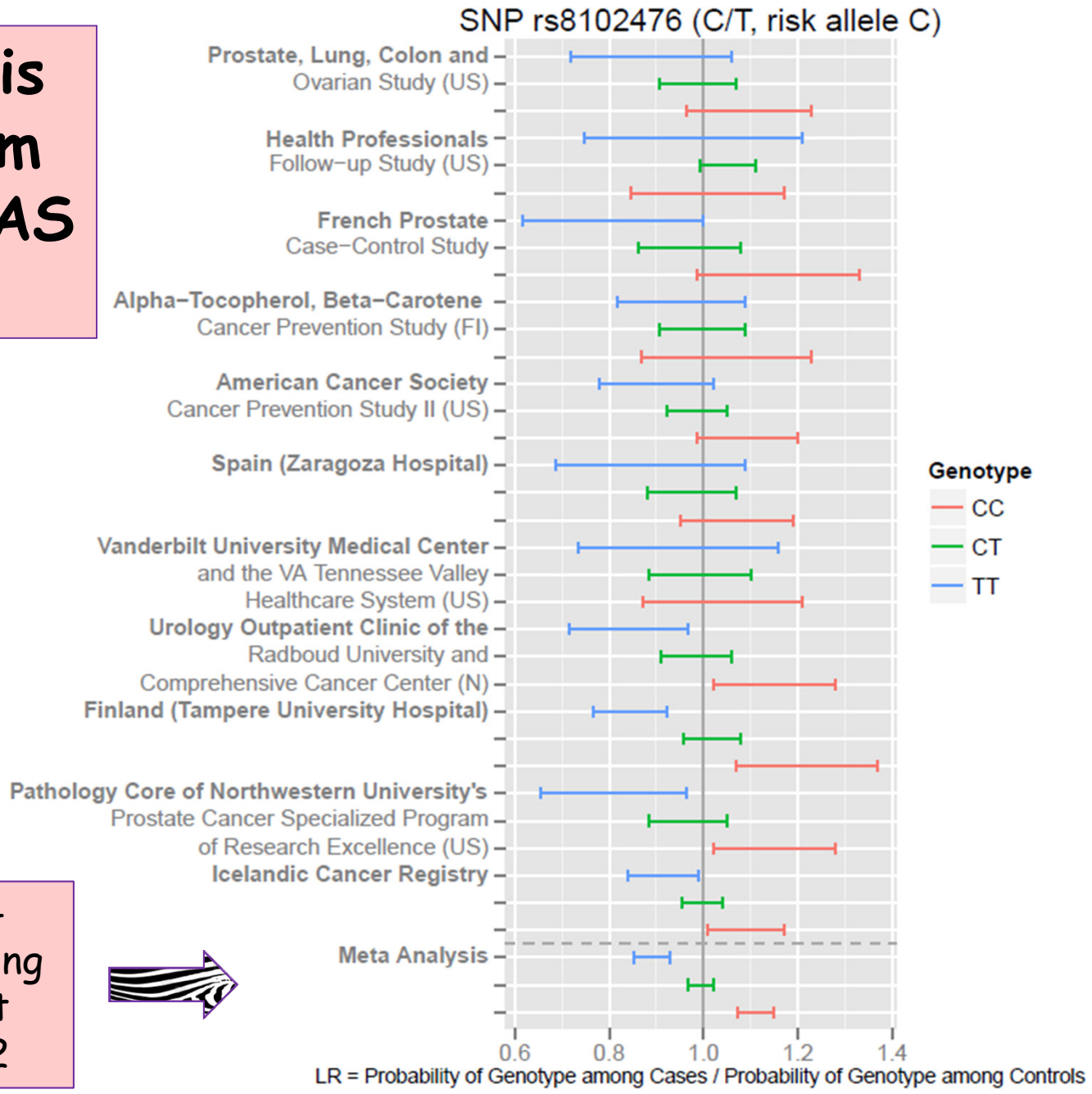
$Y = Y_1, Y_2, \ldots, Y_r$: Multiple SNPs from different studies that are known from the Hapmap not to be in linkage disequilibrium (LD) $\Rightarrow$

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

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

- Multiple SNPs in LD → multiply LR's
- Multiple SNPs not in LD → import LD/correlation from the HapMap.

Meta-analysis of SNPs from multiple GWAS studies



Multivariate meta-analysis of LR's using van Houwelingen et al, Stat Med, 2002

Comparison of SNPs with self-report family history

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

- FDR: first-degree relative; from Swedish Family-Cancer Database
- SNP LR from meta-analysis of 3 GWAS studies

Potential Usefulness of Single Nucleotide Polymorphisms to Identify Persons at High Cancer Risk: An Evaluation of Seven Common Cancers

Ju-Hyun Park, Mitchell H. Gail, Mark H. Greene, and Nilanjan Chatterjee

A B S T R A C T

Purpose
To estimate the likely number and predictive strength of cancer-associated single nucleotide polymorphisms (SNPs) that are yet to be discovered for seven common cancers.

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.

Results
Age-specific discriminatory accuracy (AUC) for models including FH and foreseeable SNPs ranged from 0.575 for ovarian cancer to 0.694 for prostate cancer. The proportions of patients in the highest decile of population risk ranged from 16.2% for ovarian cancer to 29.4% for prostate cancer. The corresponding false-positive ratios were 241 for colorectal cancer, 610 for ovarian cancer, and 138 or 280 for breast cancer in women age 50 to 54 or 40 to 44 years, respectively.

Conclusion
Foreseeable common SNP discoveries may not permit identification of small subsets of patients that contain most cancers. Usefulness of screening could be diminished by many false positives. Additional strong risk factors are needed to improve risk discrimination.

J Clin Oncol 30:2157-2162.

Closing remarks

Mitch Gail
NCI

Last modified date: 05/16/2011

- Risk Calculator
- About the Tool
- Breast Cancer Risk
- Mobile Access
- Download Source Code

Page Options
Print Page
Email Page

Quick Links
Breast Cancer Home Page
Breast Cancer: Prevention, Genetics, Causes
Current Clinical Trials: Breast Cancer *In Situ*: Treatment
Current Clinical Trials: Breast Cancer Prevention
Current Clinical Trials: Breast

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. The tool has been updated for African American women based on the Contraceptive and Reproductive Experiences (CARE) Study, and for Asian and Pacific Islander women in the United States based on the Asian American Breast Cancer Study (AABCS). See About the Tool for more information.

Before using the tool, please note the following:

- The Breast Cancer Risk Assessment Tool was designed for use by health professionals. If you are not a health professional, you are encouraged to discuss the results and your personal risk of breast cancer with your doctor.
- Although the tool may accurately estimate a woman's risk of developing breast cancer, these risk estimates do not allow one to say precisely which woman will develop breast cancer. In fact, the distribution of risk estimates for women who develop breast cancer overlaps the estimates of risk for women who do not.
- The tool should not be used to calculate breast cancer risk for women who have already had a diagnosis of breast cancer, lobular carcinoma in situ (LCIS), or ductal carcinoma in situ (DCIS).
- The BCRA risk calculator may be updated periodically as new data or research becomes available.
- Although the tool has been used with success in clinics for women with strong family histories of breast cancer, more specific methods of estimating risk are appropriate for women known to have breast cancer-producing mutations in the BRCA1 or BRCA2 genes.

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