Examples

Why online calculators for cancer risk?

Il is not easy for a clinician to translate a list of odds ratios into a salient model for a patient. The other risk factors.

Cancer is explicitly influenced by several risk factors. The simultaneous effect of the risk factors is analyzed by several risk factor models. A tool like this can give the effect of each risk factor adjusted for other risk factors on cancer risk and report odds ratios or other risk models. The model itself is a coherent model for a patient.

Why online calculators for cancer risk?

For this purpose, the National Institutes of Health (NIH) recently announced the development of a website that translates risk models into online tools.

The tool, called Cancer Risk Evaluation (CARE), allows users to input their risk factors and receives a personalized cancer risk assessment.

The tool is designed to help patients understand their risk of developing cancer and to make more informed decisions about their health.

Why online calculators for cancer risk?

Science Center at San Antonio
Department of Biostatistics, University of Michigan
Department of Epidemiology, University of Michigan
Institute for Medical Informatics, IRB, and

Donna Pavelien, Ph.D.

Bayesian MCMC, Cancer prediction, Markov chain Monte Carlo (MCMC)
### Results: Gleason 7 Disease

<table>
<thead>
<tr>
<th>Risk (95% CI)</th>
<th>Variable</th>
<th>Parameter</th>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>Prostate biopsy</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>1.0</td>
<td>African American</td>
<td>2.4</td>
<td>1.55</td>
<td>1.44</td>
</tr>
<tr>
<td>1.0</td>
<td>Age</td>
<td>0.01</td>
<td>1.0</td>
<td>1.01</td>
</tr>
<tr>
<td>1.0</td>
<td>DRE</td>
<td>0.0</td>
<td>2.16</td>
<td>1.96</td>
</tr>
<tr>
<td>1.0</td>
<td>Prostate specific antigen (PSA)</td>
<td>0.0</td>
<td>3.64</td>
<td>3.4</td>
</tr>
</tbody>
</table>

One year of the PSA and DRE, the Prostate biopsy with a biopsy performed within
the Prostate biopsy by age, race, and history of a prior biopsy in 55 men from
the Gleason grade of prostate cancer, and history of prostate
development between prostate-specific antigen (PSA).

Using multiple logistic regression models, we analyzed the

### Risk factors for biopsy-detected prostate cancer

- Prostate biopsy
- African American
- Age
- DRE
- Prostate specific antigen (PSA)
**External Validation on the SAQOR**

This graph illustrates the external validation of the SAQOR model compared to internal cross-validation. The SAQOR model achieves a 65.5% AUC for PctP at the population level, indicating its effectiveness in predicting prostate cancer risk.

- **Only 5% Reduction from SAQOR**: Valued at 70% PctP.
- **Compared to Internal Cross-Validation**: AUC for PctP Risk.
- **Risk no Better than PSA (also in PctP)**: External validation on the SAQOR model.

---

**The Burden to Validate and Compare**

There is also the requirement to compare new prediction tools with existing external populations. To ensure accuracy and appropriate characteristics, it is important to validate new tools against established benchmarks. This helps in assessing the performance and reliability of new methodologies in comparison to existing tools.

The validation process is crucial for ensuring that new tools are effective and can be trusted by the public. It involves testing the tools on diverse populations to confirm their generalizability and applicability.

---

**Online PCTP Risk Calculator**

This calculator is designed to help predict the likelihood of prostate cancer based on user input. It provides a personalized risk assessment, aiding in the decision-making process for patients and healthcare providers.

- **Risk Assessment**: Accurate and tailored to individual needs.
- **User-friendly Interface**: Easy to use and understand.
- **Professional Guidance**: Supports informed decision-making.

The calculator is accessible online, offering a convenient and efficient way to access cancer risk information.
Screening for ovarian cancer

- Can longitudinal CA 125 screening improve early detection?
- CA 125 > 35 U/ml refer to ultrasound.
- CA 125 is a blood test that increases for most ovarian cancers up to 2 years prior to clinical detection.
- CA 125 is a blood test that increases for most ovarian cancers up to 2 years prior to clinical detection.
- No ultrasonic screening in the United States due to low incidence and high cost.
- 90% of women do not know Stage I: 8% get symptoms, go to doctor.
- 20% Stage II, IV: 20% 5-year survival rate if detected in

Monitoring for prostate cancer recurrence

- US-NIH UK-SCREEN

2 Applications

- Bayesian Markov Chain Monte Carlo (MCMC) methods
  - to condition on the new person's longitudinal history.
  - with patients' variance, a prediction for a new person needs
    - to condition on the new person's longitudinal history.
    - the mean longitudinal trajectory is often
      - a Markov model for longitudinal model measurements and other
        - deterministic and time-to-event outcomes.
      - to condition on the new person's longitudinal history.
      - the mean longitudinal trajectory is often
        - a Markov model for longitudinal model measurements and other
          - deterministic and time-to-event outcomes.
  - Computational models are difficult to condition on joint statistical
    - potential more powerful since more information is contained
Longitudinal mixture models

\[ Y_{it} = \log \text{CA125 for patient } i \text{ at time } t \]

**Controls:** constant individual-specific mean over time
\[ Y_t \sim N(\mu, \sigma^2) \]

**Cases:** 85% have significant elevation in CA125 with cancer
- 15% of cancer cases do not express extra CA125

Longitudinal CA125 follows a mixture distribution

- Need reversible jump
- Must borrow information

\[ Y_{it} | \{ \mathbf{i}_t = 0 \} \sim N(\mu, \sigma^2) \]
\[ Y_{it} | \{ \mathbf{i}_t = 1 \} \sim N(\mu + \beta_1 \tau_t^*, \sigma^2) \]

\( \mathbf{i}_t \sim \text{Ber}(\pi) \) is estimated for each cancer case
\[ d_t - \tau_t \sim N(2, .75^2)I(0,5) \], where \( d_t \) is time of clinical detection,

\( I(\cdot) \) is indicator for the interval 0 to 5.

Fit to UK screening data

- Population distributions for all random effects.
- Prior distributions for all population parameters.
- Fit to 19766 controls and 61 cases using Markov chain Monte Carlo combination of Gibbs, Metropolis-Hastings and reversible jump steps.
- C program; 5,000 burn-in+5,000 iterations; 4 hours.
- Convergence diagnostics.
- Goodness of fit statistics.
- Sensitivity to prior.

Goodness of fit of CA125 model for 4 cancer cases

- Posterior probability of a changepoint, \( E(\mathbf{i}, \text{data}) \).
- Graphs begin 5 years prior to clinical detection and end at clinical detection.
Valuation on Independent Swedish Trial

Results: Simulated 7-Year Study

Valuation

Calculating risk of cancer for a new woman

\[
\begin{align*}
\text{Risk} &= \frac{\text{Missed 20\% cancers}}{\text{Standard} = 65\%} \\
\text{Risk} &= \frac{\text{Missed 35\% cancers}}{\text{But the sensitivity}} \\
\text{Both obtain specificity 99\%} \\
\text{Bayesian risk} \\
\text{False positive rate 1\%} \\
\text{Risk} < 1\% \\
\text{cALR} < 30 \mu/ml \\
\text{Detection rate 256 cases} \\
\text{Small number} \\
\text{Test set} \\
\text{4276 controls} \\
\text{10 cases} \\
\text{1000 population} \\
\text{age-adjusted population prevalence} \\
\text{cALR} = CA125 for a new woman} \\
\end{align*}
\]
Monitoring for prostate cancer recurrence
Screening for ovarian cancer

US-NIH UK-SCREEN

The importance of early stage detection

was accepted as part of a UK screening trial initiated in 2001. On the basis of these results, the Bayesian risk method only 99% by the standard approach.

would have captured 61% of early stage cancers compared to

Then in the simulations the Bayesian risk screening program

duration.

Suppose the early stage was the first 60% of the pre-clinical
detection = 0.1. - 1.

line between the change-point in CA125 and clinical

in the simulations. define the pre-clinical duration as the

Ongoing until 2012. Risk of cancer calculator

evaluated on one arm of UK Trial (n=200,000)
versus non-advanced patients. Salience hormone, chemotherapy and surgery. 
Radiation only (varying doses), radiation plus hormones.

RTTOG 9406
RTTOG 9413
RTTOG 9292
Massachusetts General Hospital: $=0.125
Hologic, Inc.: $=0.104

Followed up to 12 years. Large cohort.

Building a prediction model for diagnosis and monitoring models for longitudinal PSA.

**Proposed approach**

**Joint statistical models for longitudinal PSA.**

**Several shortcomings**

- PSA shows 2 or 3 consecutive rises over any time frame.
- PSA doubles in size from its lowest value (nadir) or
- Current PSA >2 c or
- Frequent to biopsy. Uncertainty in multiple PSA readings.
- Subsequent time frames can be ill-defined
- "Bouncing PSA" does not account for measurement error.

**Current practice**

Monitoring patients after radiotherapy for first presentation of prostate cancer.

- Focus on prediction for a new patient—develop an
- Radiation Therapy and Oncology Group (RTOG)
- Radiation Oncology Group (RTOG)
- University of Michigan Cancer Center (UMCC) and
- Massachusetts General Hospital (MGH), University of
- Massachusetts General Hospital: $=0.125
- Hologic, Inc.: $=0.104

Followed up to 12 years. Large cohort.
A joint longitudinal failure time model.
Covariates Considered

$h(t) = a + \beta t - \log(t+1)$(t+1)

Longitudinal

Survival Component:

\(S(t) = \exp[-\int_0^t f(t) \; dt]

Logit link

Risk predictors:

\(\text{logit}(\pi) = \beta_0 + \sum \beta_i X_i

\)
Posterior Estimates of Survival Parameters

Model 1 (with frailty)

Logit link

Hazard ratios:

Model 2 (without frailty)

Logit link

Hazard ratios:

280.7 points above the optimal model

using only 4J (i) has an SBC value of 286.2.

Similar models with similar SBC: Populer model

Logits

Survival component:

Proportional hazards of logit:

Optimal Model (SBC = 247.36)

Markov Chain Monte Carlo (MCMC)

4 hours on a Dell computer.

5,000 burn-in + 5,000 samples.

Heating stops:

A combination of Gibbs and Metropolis.

Coded and run in C.

Check of Normal assumptions, influence in progress.

Longitudinal component

Goodness of fit checks show adequate individual fits for

Censored may be more measurement random effects and

L's of L for both

Currently finding this.

parameters have no problems.

(1) d = \log(t) < i)

\int_{i}^{\infty} \frac{1}{(1) d} \times \frac{t}{(1) d} dt

\int_{i}^{\infty} \frac{1}{(1) d} \times \frac{t}{(1) d} dt

Median length

only 4J (i) is in top German convergence criteria.

Most predictors no longer significant later on. Revised
Problem is that it is not feasible to check convergence in the non-monotone likelihood so converge in G0 iterations.

Results in a chain of a very informative prior and a proposal density in a Markov-Hastings update for G_0.

Let the posterior distribution as a prior distribution for G_0.

Method 2: mini MCMC chain for R,G_0

\[
\theta \sim p(\theta) \propto p(\text{data} | \theta) p(\theta | \theta_{\text{init}})
\]

Prediction for a new patient

Method 1: data augmentation

\[
\theta \sim p(\theta) \propto p(\text{data} | \theta) p(\theta | \theta_{\text{init}})
\]

Prediction for a new patient

Conditional Independence

\[
\theta \sim p(\theta) \propto p(\text{data} | \theta) p(\theta | \theta_{\text{init}})
\]

Note: includes all data from a training set, e.g. MGH

\[
\theta \sim p(\theta) \propto p(\text{data} | \theta) p(\theta | \theta_{\text{init}})
\]

new patient up until time t

\[
\theta \sim p(\theta) \propto p(\text{data} | \theta) p(\theta | \theta_{\text{init}})
\]

lowed up until time t

\[
\theta \sim p(\theta) \propto p(\text{data} | \theta) p(\theta | \theta_{\text{init}})
\]

Prediction for a new patient
Work in progress to combine data from MGH, UMC, RTGC.

Different institutions/populations.

- Positive predictive value: PPV (cancer detected/cancer).
- Sensitivity: true positive/true positive + false negative.
- Specificity: true negative/true negative + false positive.
- Accuracy: (true positive + true negative)/(true positive + true negative + false positive + false negative).

PSA = prostate-specific antigen. Risk of recurrence should be monitored.

Doctors are confused and patients are alarmed by "bouncing"

A probability of recurrence is easier for a patient and doctor.

Nevertheless.

**Evaluation criteria**

Some reasons for bad performance

Model

Some reasons for bad performance

PSA and Data

PSA not elevated in all patients that recur and

for patients that develop prostate cancer screening in the PCG.

Predict: clean up outliers.

- PSA elevation and require longer follow-up for patients not on androgen deprivation therapy (steroid).
- PSA may still be elevated with a low PSA level.
- Problems in the database: Many of the patients with a cancerous prostate are treated with ADT.

Zheng Y et al. 2009

Regression instead of Cox proportional hazards model

Survival model: Cox regression (Yu et al. 2004), covariates:

- Recurrence, Gleason score, Gleason grade, Gleason age, et al.

Regression instead of Cox proportional hazards model

Survival model: Cox regression (Yu et al. 2004), covariates

- Recurrence, Gleason score, Gleason grade, Gleason age, et al.

Regression instead of Cox proportional hazards model

Survival model: Cox regression (Yu et al. 2004), covariates

- Recurrence, Gleason score, Gleason grade, Gleason age, et al.

Regression instead of Cox proportional hazards model

Survival model: Cox regression (Yu et al. 2004), covariates

- Recurrence, Gleason score, Gleason grade, Gleason age, et al.