Examining Improvement in Risk Prediction

Dave Glidden
24 November 2009

Risk Model

- Existing set of risk factors for disease
- Novel marker or test
  genetic, proteomic, imaging, biomarker
- New marker add predictive information
- How much better are predictions?
  do they aid in clinical decision making

Examples

- Breast density for breast cancer risk
  Tice et al. 2008
- C-reactive protein for CV risk
  Cook et al. 2006
- PSA for prostate cancer risk

The Easy Question

and least interesting

- Is there a statistically significant contribution of the new measure/marker?
- Fit a logistic/Cox regression model
- Examine p-value for new measure adjusting for other factors
- Gives p-value
- Implies better prediction

Why not interesting?

- Turns out to be an easy threshold
- Many factors can reach this
- Always need to know how much better we do?
- Relevant summary needed

Measure of Prediction

- TPF = true positive fraction
  proportion of case "positive"
  aka: sensitivity
- FPF = false positive fraction
  proportion of control "positive"
- Trade-off
- Two kinds of errors

Context is Important

- Prostate cancer screening:
  FPF = 10%, TPF = 50%,
  prevalence = 5% => PVP = 21%
- Metastatic disease in breast cancer
  TPF = 92%, FPF = 42%
- What are implications of false positive, false negative?

Odds Ratio

- Suppose FPF = 10%
  good specificity (90%)
- Suppose OR = 3.0
  pretty hefty OR
- Implies TPF = 25%
  very low specificity = 25%

Odds Ratio

- FPF = 0.10
  90% specificity
- TPF = 0.80
  80% sensitivity
- OR = (0.80/0.20) / (0.10/0.90) = 36.0
- Huge OR associated with good discrimination

Dist of Marker and OR

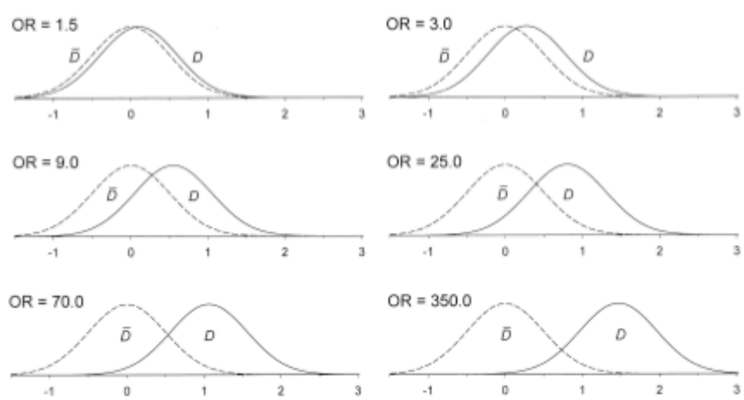


FIGURE 2. Probability distributions of a marker, *X*, in cases (solid curves) and controls (dashed curves) consistent with the logistic model logit$P(D = 1|X) = \alpha + \beta X$. It has been assumed that *X* has a mean of 0 and a standard deviation of 0.5 in controls so that a unit increase represents the difference between the 84th and 16th percentiles of *X* in controls. The marker is normally distributed, with the same variance in cases. The odds ratio (OR) per unit increase in *X* is shown.

Am J Epidemiol 2004;159:882–890

ROC Curves

- Can be problematic for risk
- Don't display absolute risk values
- Two tests graphed =>
  FPF v. TPF (varying threshold)
- C-statistic doesn't relate to clinical decisions
  measure ability to classify

Tice at al 2008

- Breast cancer risk
- SEER data: 1,095,484 women
- age, race/ethnicity, family hx, biopsy hx
- added value of breast density
  HR 4.2 (less than 65), 2.2 (65 or older)
  statistically significant
- 5 year risk model (Cox model)
- How much does density add to risk assessment?

Another Approach

- Based on results of a model, calculate risks
- Group risks in context dependent way
  Tice: <1%, 1-1.6%, 1.6-2.5%, > 2.5%
- Examine calibration
- Compare between models

Model w/o Density

Strata	<1%	1.0-1.66%	1.67-2.5%	> 2.5%
N	215,402	201,972	148,795	63,105
% event	0.7%	1.3%	1.8%	2.9%

Model with Density

Strata	<1%	1.0-1.66%	1.67-2.5%	> 2.5%
N	249,959	186,106	124,420	68,744
% event	0.7%	1.3%	2.0%	3.1%

Calibration

- % events by category
- Both models well calibrated
- % events observed falls within expected
- Calibration has subjective aspect
  if model predicts 2% but is really 3%
  close enough?

NR Cook

- Compare two *nested* prediction models
  one includes extra predictor
  like breast density
- Cross-tabulate risk predictions based on model
- Examine who gets reclassified
- What % and in which direction?

Table 1. **Five-Year Risks for Breast Cancer as Predicted by Models That Do and Do Not Include Breast Density***

5-Year Risk from Model without Breast Density	5-Year Risk from Model with Breast Density				Total
	0% to <1%	1% to 1.66%	1.67% to 2.5%	>2.5%	
0% to <1%					
Women, *n*	176 831	38 500	71	0	215 402
Events, *n*	1161	415	0	0	1576
Nonevents, *n*	175 670	38 085	71	0	213 826
Proportion of women with events	0.7	1.1	0.0	–	0.7
1% to 1.66%					
Women, *n*	64 297	99 456	37 149	1025	201 927
Events, *n*	526	1328	754	32	2640
Nonevents, *n*	63 771	98 128	36 395	993	199 287
Proportion of women with events	0.8	1.3	2.0	3.1	1.3
1.67% to 2.5%					
Women, *n*	8741	45 478	71 309	23 267	148 795
Events, *n*	74	609	1419	621	2723
Nonevents, *n*	8667	44 869	69 890	22 646	146 072
Proportion of women with events	0.9	1.3	2.0	2.7	1.8
>2.5%					
Women, *n*	90	2672	15 891	44 452	63 105
Events, *n*	0	38	340	1467	1845
Nonevents, *n*	90	2634	15 551	42 985	61 260
Proportion of women with events	0.0	1.4	2.1	3.3	2.9
Total					
Women, *n*	249 959	186 106	124 420	68 744	629 229
Events, *n*	1761	2390	2513	2120	8784
Nonevents, *n*	248 198	183 716	121 907	66 624	620 445
Proportion of women with events	0.7	1.3	2.0	3.1	1.4

Reclassification Index

with breast cancer

	<1%	1.0-1.66%	1.67-2.5%	> 2.5%
<1%	1161	415	0	0
1.0-1.66%	526	1328	754	32
1.67-2.5%	74	609	1419	612
> 2.5%	0	38	340	1467

Reclassification

- Risk strata same: 5375 (61%)
- Risk strata lower w/ dens 1587 (18%)
  wrong direction
- Risk strata higher w/ dens 1813 (21%)
  correct direction
- Net reclassification = 21%-18% = 3%
  in the right direction

Reclassification Index

no breast cancer

	<1%	1.0-1.66%	1.67-2.5%	> 2.5%
<1%	175,670	38,085	71	0
1.0-1.66%	63,771	98,128	36,395	993
1.67-2.5%	8,667	44,869	69,890	22,646
> 2.5%	90	2,634	1,551	42,985

Reclassification

- Risk strata same: 386,673 (64%)
- Risk strata lower w/ dens 121,582 (20%)
  right direction
- Risk strata higher w/ dens 98,190 (16%)
  wrong direction
- Net reclassification = 20%-16% = 3.8%
  in the right direction

Plots

- Can avoid cutpoints
- x axis: pred. prob with breast density
- y axis: pred. prob w/o breast density
- plot separately for cases and non-cases

Cautions

- Method doesn't work well in non-nested models
- Nested: one model has all predictors in other
- Can't be applied to case-control data
  not possible to estimate "risk"

Risk Reclassification

- Requires risk strata
  4 strata is typical
  arrived in a context-dependent way
- Shows calibration
- Direct measure of change
  due to adding marker
- See proportion reclassified

Diagnostic Likelihood Ratios

- DLR: familiar from clinical epi
- Describes how risk of disease is modified
  by new information
- pr(disease | x) / pr(disease)
  x is some diagnostic test information
  DLR is function of x
- DLR > 1, risk increased
- DLR < 1, risk decreased

Our scenario

- D: disease variable (D=1, has disease)
- X: an established risk factor for disease or even a risk score
- Y: a novel factor how much does it add?
- Extend the concept of a DLR

DLR

- DLR(Y) = P(D=1| X,Y) / P(D=1, X)
- the ratio of risk given X + Y to X alone
- a function of Y
- How does this differ from an odd ratio for Y?

Gu and Pepe

- $\log(p|X) = \alpha_0 + \alpha_1 * X$
- $\log(p|X,Y) = \beta_0 + \beta_1 * X + \beta_2 * Y$
- Then
  $\log\{DLR(Y)\} = (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) * X + \beta_2 * Y$
- CI for this is complicated
- But easily graphed

Physician's Health Study

- Randomized clinical trial in 1980s
- 430 men developed prostate cancer
- Nested case control study on stored specimens
- Total PSA and % free PSA

Total PSA alone

```
. logistic d tp, coef

Logistic regression                               Number of obs   =        683
                                                  LR chi2(1)      =     187.10
                                                  Prob > chi2     =     0.0000
Log likelihood = -342.10876                       Pseudo R2       =     0.2147

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        tpsa |   .3533952   .0399274     8.85   0.000     .2751389    .4316514
       _cons |  -1.886684   .1481591   -12.73   0.000    -2.177071   -1.596298
------------------------------------------------------------------------------
```

Total PSA alone

```
. logistic d fpsa tp, coef

Logistic regression                               Number of obs   =        683
                                                  LR chi2(2)      =     189.43
                                                  Prob > chi2     =     0.0000
Log likelihood = -340.94405                       Pseudo R2       =     0.2174

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fpsa |  -.3493467   .0829424    -4.21   0.000    -.5119108   -.1867826
        tpsa |   .4037308   .0471778     8.56   0.000      .311264    .4961976
       _cons |  -1.843289   .1438261   -12.82   0.000    -2.125183   -1.561395
------------------------------------------------------------------------------
```

log DLR

$\log\{DLR(Y)\} = (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)*X + \beta_2*Y =$

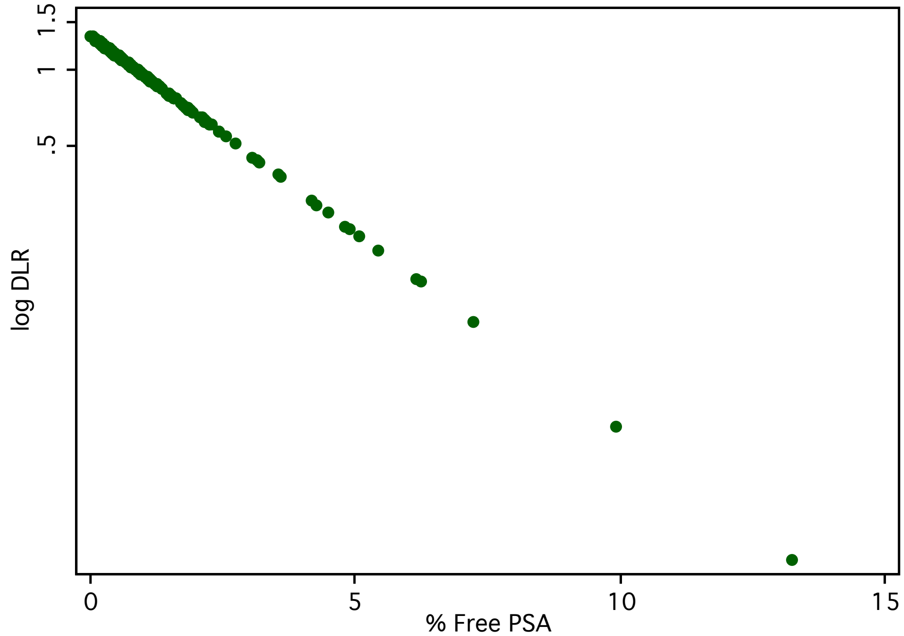
$0.05 + 0.051*TPSA - 0.35*FPSA$

Graph this at average TPSA =4.663

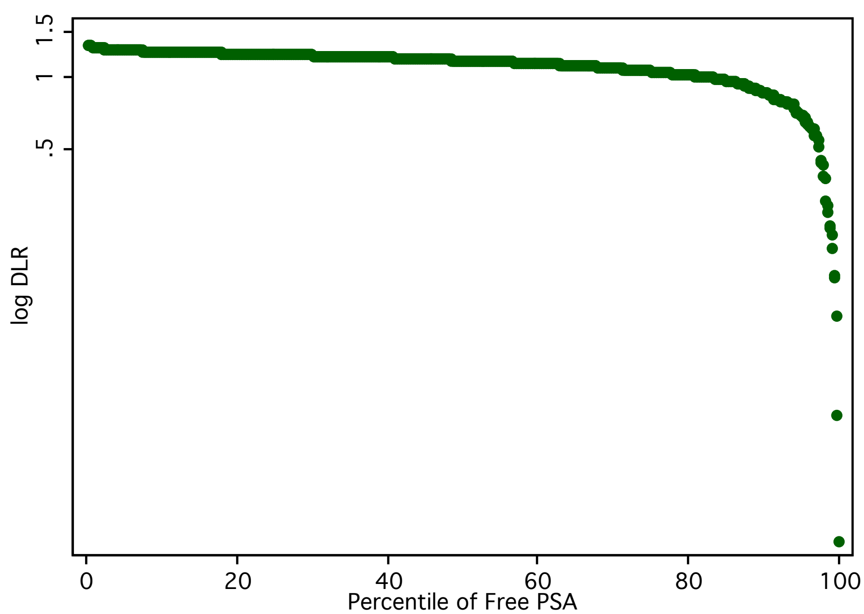
Stata

- gen log_dlr=0.05+0.051*4.663-.35*fpsa
- gen dlr = exp(dlr)
- twoway (scatter dlr fpsa), ytitle(log DLR) yscale(log) xtitle(% Free PSA)

DLR by % FPSA



DLR by % FPSA



Valued added

- High values indicate decreased risk
- 81% have DLR > 1
  lowest 81% of FPSA indicate increased risk
- High FPSA can indicated much lower likelihood of cancer (at a given PSA)
- Maybe helpful for ruling out disease in borderline PSA

Interactions

- Value of FPSA (or any marker)
- May depend on configuration of other predictors
- Can examine this using interactions in logistic models

OR

- Risk factor may have elevated OR
- Individual risk is modified by risk factor
- Population risk may not be greatly modified
- Especially true if risk factors cluster
- Both embody some level of truth

C-statistic

- Very difficult to budge
- Measure pure ability to classify
- However, strict classification is rarely of interest
- More interested in risk stratification

Summary

- These kinds of assessments very difficult
- Must weight very difficult trade offs
- Difficult to balance false positives and false negatives
- Utilities may be very different for different people