

Diagnostic Screening Without Cutoffs

Wesley O. Johnson Young-Ku Choi

Department of Statistics

and

Mark C. Thurmond

Department of Medicine and Epidemiology

University of California, Davis

● Introduction

- Standard diagnostic test procedures involve dichotomization
- Critical value or cutoff is determined to optimize trade off between sensitivity and specificity
- Values high above the cutoff are treated the same as units with values just barely above the cutoff, etc

- Inherent information loss in dichotomization.
- Predictive values positive and negative are exactly the same, regardless of actual (non-dichotomized) test result
- We thus develop a diagnostic screening method based on data that are not dichotomized within the Bayesian paradigm

- Method determines the *predictive probability of infection* based on specific serologic test result
- Provides inferences about the prevalence of infection in the population sampled
- Methods are illustrated on problems of association of *Neospora* infection to natural abortion in dairy cattle, without a gold standard test but with training data

- **Bayesian Probability**

- Diagnostic Assignment**

- We now describe the classic epidemiologic problem
- Let E and \bar{E} denote presence and absence of “exposure”
- Let D and \bar{D} denote presence and absence of “disease”

- Interested in prevalences

$$P(E|D) P(E|\bar{D})$$

- and corresponding RR,

$$P(E|D)/P(E|\bar{D})$$

- With case-control data, independent samples of units known to be D and \bar{D} are taken

- We make these classic epidemiologic inferences, only with data that is one step removed from the standard type
- Instead of observing E , we actually see S , which serves as a proxy for E
- We don't know if a cow is E or not, but we calculate

$$P(E|S)$$

- Given information about $P(D)$ in the overall population, we can make inferences about the RR

$$P(D|E)/P(D|\bar{E})$$

- In our illustration, E is the infectious agent, *Neospora caninum* infection in dairy cattle, and D is the event of natural abortion

- We require availability of std training data, eg. two independent sets of serologic values obtained from infected and non-infected cows

$$\{T_{jk}; k = 0, 1, \quad jk = 1, \dots, n_k\}$$

- We also assume a parametric model

$$T_{jk} \sim p_1(\cdot|\theta)^k p_0(\cdot|\gamma)^{1-k}$$

$$k = 0, 1, \quad j = 1, \dots, m_k$$

- Prior information for all parameters is incorporated in the form of a prior probability distribution, namely

$$p(\theta, \gamma) = p(\theta)p(\gamma)$$

- We consider data presented in Thurmond *et al.* (2002)

- Training data consist of ELISA S/P values for $m_1 = 196$ **cattle infected with *N. caninum*** and for $m_2 = 553$ **not infected with *N. caninum***
- Also have ELISA values for $n_1 = 58$ **aborted cows (cases)** and $n_0 = 57$ **non-aborted cows (controls)**, collected from a Calif herd with an abortion problem

- Infection status unknown.
Prevalences of infection among aborted and among non-aborted cows unknown
- The probability of natural abortion, $P(A) = q$, is known to be about 0.108 from the herd records
- Using the Training Data, Weib(α, λ) and IG(a, b) distributions were reasonable choices for ELISA S values for inf and non-inf animals

– Prior distributions used were

$$\alpha \sim \text{Gamma}(0.12, 0.1)$$

$$\lambda \sim \text{Gamma}(0.1, 0.1)$$

$$a \sim \text{Gamma}(0.4, 0.1)$$

$$b \sim \text{Gamma}(0.08, 0.1)$$

– These elicited from
Dr. Thurmond, independently
of the current data

- We elicited a ‘best estimate’ for the median serologic value for infected cows (0.76), and the corresponding 95th percentile (2.5) based on Dr. Thurmond’s experience with previous studies
- This resulted in best guesses $\alpha \simeq 1.22$ $\lambda \simeq 0.97$
- Similarly, $a \simeq 4$ and $b \simeq 0.8$ based corresp ‘best’ estimate (0.16), and 95th percentile (0.6) for un-inf animals

- Dr. Thurmond's best estimate for the prevalence of infection for aborted cows was 0.35, and the corresponding 95th percentile was 0.65
- Best estimate of prev for non-aborted animals is 0.2 with corresp 95th %tile 0.35
- We obtained

$$\pi_1 \sim \text{beta}(3.48, 5.61)$$

$$\pi_0 \sim \text{beta}(6.77, 24.06)$$

- Posterior results for (α, β) and (a, b) for our method were obtained using the training data in conjunction with case-control data
- The Bayesian method makes inferences about all parameters in a single framework incorporating all data in a single analysis
- Used 15,000 Markov chain Monte Carlo (MCMC) iterations in WinBUGS, after a burn-in of 500

Posterior Inferences

	$q = 0.108$	
	PM	95% PI
$P(E)$	0.19	(0.12, 0.27)
$P(E D)$	0.36	(0.24, 0.49)
$P(E \bar{D})$	0.17	(0.10, 0.26)
$P(D E)$	0.20	(0.13, 0.33)
$P(D \bar{E})$	0.09	(0.07, 0.10)
RR	2.42	(1.23, 4.62)

$P(E|S)$ in case-control
study ignoring abortion

S		status		
		Pr. BS	Prob Bayes	Inf ML
.19	-	.001	.004	0.004
.26	-	.004	.014	.013
.35	-	.025	.076	.076
.40	-	.062	.172	.176
.48	+	.218	.448	.467
.75	+	.927	.966	.972
.91	+	.987	.992	0.993

- Compare BPDA with Bayesian Approach Using Cutoff of 0.45
 - Dichotomized C-C Data (0.45)

	A	\bar{A}
$+$	22	9
$-$	36	48

- Dr. Thurmond's priors for Se and Sp are beta(731.4, 100.6) and beta(330, 11.2)
- The training data were also dichotomized.
- The number of positives among the known inf animals, and

the number of negatives among the non-inf animals were combined with the beta priors to obtain posteriors

- These then become priors for Se and Sp that can be combined with the case control data to make inferences
- The resulting priors are $Se \sim \text{beta}(753.4, 136.6)$ and $Sp \sim \text{beta}(378, 20.2)$
- We used the same priors for π_1 and π_0 and q that were used in the case-control study

W/ Cutoff

	PM	95% PI
$P(E)$	0.22	(0.12,0.35)
$P(E D)$	0.45	(0.22,0.72)
$P(E \bar{D})$	0.18	(0.08,0.33)
$P(D E)$	0.26	(0.11,0.48)
$P(D \bar{E})$	0.09	(0.04,0.16)
RR	3.02	(1.06,8.28)
W/O	Cutoff	
$P(E)$	0.19	(0.12,0.27)
$P(E D)$	0.36	(0.24,0.49)
$P(E \bar{D})$	0.17	(0.10,0.26)
$P(D E)$	0.20	(0.13,0.33)
$P(D \bar{E})$	0.09	(0.07,0.10)
RR	2.42	(1.23,4.62)