

**Models, Methods and
Inferences for Prevalence
and Test-Accuracy Based on
Imperfect-Dependent-Test
Screening Data**

Wesley O. Johnson, UCD

Ian A. Gardner, UCD

Timothy E. Hanson, UNM

Marios Georgiadis

References

- Hanson, T.E., Johnson, W.O. and Gardner, I.A. (2003). Hierarchical models for the estimation of disease prevalence and the sensitivity and specificity of dependent tests in the absence of a gold-standard. *Journal of Agricultural, Biological and Environmental Statistics*.
- Georgiadis, M,P., Johnson, W.O., Singh, R. and Gardner, I.A. (2003). Correlation-Adjusted Estimation of sensitivity and specificity of two diagnostic tests. *Applied Statistics* 52.

The Problem

- Screening a population for disease **D**
- Screening accomplished with two binary diagnostic tests: T_1 and T_2 , each taking the values (+) or (-).
- In the absence of a gold-standard we wish to estimate:
 - The *Prevalence* of **D**, $\pi = P(\mathbf{D})$.
 - The *Sensitivity* of each test $\eta_i = P(T_i^+ | D)$,
 - The *Specificity* of each test $\theta_i = P(T_i^- | \bar{D})$,
 - The *Conditional Correlations* between the tests:
 $\rho_D = \text{corr}(T_1, T_2 | D)$, $\rho_{\bar{D}} = \text{corr}(T_1, T_2 | \bar{D})$.

The Specific Situation:

- Assume $K \geq 2$ populations are sampled
- Assume sensitivity/specificity are equal across populations.
- Data for i^{th} popn are cross-classified as:

		T_2		
		+	-	
T_1	+	X_{i11}	X_{i10}	Population i
	-	X_{i01}	X_{i00}	

- Let $X_i = \{X_{ijk}\}$
- Define the prevalence of **D** in population i

$$\pi_i = P(\mathbf{D}|\text{pop'n } i)$$

- Assume

$$\mathbf{X}_i \sim \text{Multinomial}(n_i, \mathbf{p}_i), \text{ for } i = 1, \dots, k.$$

- By the Law of TP:

$$\mathbf{p}_i = \pi_i \boldsymbol{\eta} + (1 - \pi_i) \boldsymbol{\theta}, \text{ for } i = 1, \dots, k.$$

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_{11} \\ \eta_{10} \\ \eta_{01} \\ \eta_{00} \end{pmatrix} = \begin{pmatrix} P(T_1+, T_2+ | D+) \\ P(T_1+, T_2- | D+) \\ P(T_1-, T_2+ | D+) \\ P(T_1-, T_2- | D+) \end{pmatrix};$$

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{11} \\ \theta_{10} \\ \theta_{01} \\ \theta_{00} \end{pmatrix} = \begin{pmatrix} P(T_1+, T_2+ | D-) \\ P(T_1+, T_2- | D-) \\ P(T_1-, T_2+ | D-) \\ P(T_1-, T_2- | D-) \end{pmatrix}.$$

Hierarchical Model

- Assume the π_i 's are exchangeable, eg.

$$\pi_1, \dots, \pi_k \stackrel{iid}{\sim} \text{Beta}(\alpha, \beta)$$

where α and β are not known and therefore must be modeled with prior probability

- This corresponds to a belief that the prevalences are distinct, but that there is no knowledge that a particular prevalence may be larger or smaller than another
- Moreover, if one of the prevalences were known to be large, for example, then one may tend to believe that other prevalences are also large

- Let

$$\alpha = \mu\gamma \quad \beta = \gamma$$

- Then the mean of π_i is μ and the variance is $\mu(1 - \mu)/(1 + \gamma)$
- So μ is the “average” prevalence among herds, and large γ implies that the prevalences from herd to herd are not very spread out; small γ implies much variability among herd prevalences
- The model induces positive correlation between prevalences for different populations

Modeling Test Accuracy and Correlation

- We consider situations where there may be *reasonably accurate* prior information for η_1 and θ_1 eg.

$$\eta_1 \sim \text{Beta}(a_{\eta_1}, b_{\eta_1}), \theta_1 \sim \text{Beta}(a_{\theta_1}, b_{\theta_1})$$

- This is ideal, and is often necessary
- Consider the re-parametrization where the π_i 's and the Se and Sp for test 1 remain as above, but where

$$\lambda_D = Pr(T_2+|T_1+, D), \gamma_D = Pr(T_2+|T_1-, D)$$

$$\lambda_{\bar{D}} = Pr(T_2-|T_1-, \bar{D}), \gamma_{\bar{D}} = Pr(T_2-|T_1+, \bar{D})$$

- Observe that

$$\eta_{11} = \eta_1 \lambda_D, \eta_{12} = \eta_1 (1 - \lambda_D)$$

$$\theta_{22} = \theta_1 \lambda_{\bar{D}}, \theta_{12} = (1 - \theta_1) \gamma_{\bar{D}}$$

etc

- We place independent Beta priors on all parameters
- These new parameters may not be easy to think about
- So $U[0,1]$ priors are a natural first thought
- However, since we may expect

$$Pr(T_2 + |T_1 -, D) \leq \eta_2 \leq Pr(T_2 + |T_1 +, D)$$

it may be sensible to center γ_D and λ_D on a reasonable prior guess for η_2

- Then place a “fat” Beta distribution on each
- For example, we may believe that η_2 is around 0.8. We could then place Beta priors on γ_D and λ_D that have mode 0.8 and 5th %tile equal to 0.4, or less
- This prior reflects a slight belief that $\rho_D = 0$ but allows for ρ_D to be either positive or negative, leaving it to the data to decide
- Similar arguments lead to specifying priors for $\lambda_{\bar{D}}$ and $\gamma_{\bar{D}}$ that are focused on a prior guess for θ_2 , again with large dispersion

Informative Prior for the Prevalence Distribution

- In the hierarchical model, we require a prior distribution on μ and γ
- By considering the median and 95th percentile of $\mu = E(\pi_i)$ we model

$$\mu \sim \text{Beta}(a_\mu, b_\mu),$$

where $\mu_0 = (a_\mu - 1)/(a_\mu + b_\mu - 2)$ is the best guess for the average prevalence

- We can induce informative gamma prior for γ by considering, say, the 90th %tile of the prevalence distribution, conditional on μ_0 .

Models are Not Identifiable

- Previous approaches assumed *conditional independence* of the tests to achieve identifiability (Hui and Walter, 1981; Joseph et al (1995), Johnson et al (2001)).
- The hierarchical model is not identifiable. However, there exist consistent estimates of all parameters even without a prior specification as $K \rightarrow \infty$.
- C.L. Su Dissertation presents asymp post results for GLLM's: asymptotic posts are free of the prior when $K \rightarrow \infty$, while not true if the number of obs within popns tend to ∞ with fixed K .

- The model with independent and *known* Beta priors on the prevalences is not identifiable and so it is *absolutely essential* to have *good* prior information for at least two parameters
- Georgiadis et al (2003) argued that η_1 and θ_2 should be known to within ± 0.05 of their true values in this instance, under the *dependence* model

- Hierarchical model works *in the absence of very informative priors* for test accuracy estimation when population sample sizes ~ 300
- Model can behave poorly if all of the n_i are below 100 (without extra prior information)
- $n_i > 500$ do little to improve inference
- Use of the parametric model or the DP extension for estimating prevalence requires more popns, but sig fewer subjects sampled from each population to achieve accurate results.
- The model may be modified to allow more than two tests or a single test, or additional information

Mycoplasma hyopneumoniae in Swine

- In 1992, an enzyme-linked immunosorbent assay (ELISA) (Feld et al.,1992) replaced the indirect hemagglutination test (IHA) for monitoring swine herds in the Danish specific-pathogen-free program for *Mycoplasma hyopneumonia*.
- Typically, 20 samples are collected each month to determine the herd's status for the pathogen.
- For several months both tests were run on all samples from the 123 participating herds to compare their accuracy.
- True status of each herd known based on a combination of clinical observations.

- Eighteen herds confirmed infected and 105 were free of *Mycoplasma hyopneumonia* based on the available evidence (Sorensen et al., 1992).
- The results from the 105 certified disease-free herds were combined into one “super-herd”, herd zero. This information is simply incorporated into the prior on θ .
- The remaining 18 herds have unknown, non-zero prevalence, and are assumed to be *iid* draws from a $\text{Beta}(\alpha, \beta)$ distribution with a flat prior on (α, β) .
- The data are tabled here:

Herd	$E+, I+$	$E+, I-$	$E-, I+$	$E-, I-$	total
0	1	4	5	2118	2128
1	6	9	0	5	20
2	3	10	0	7	20
3	1	3	0	15	19
4	5	9	1	5	20
5	6	4	0	2	12
6	6	11	0	3	20
7	2	10	0	7	19
8	1	9	0	19	29
9	9	10	0	0	19
10	1	19	0	0	20
11	2	6	0	12	20
12	1	15	0	4	20
13	8	12	0	0	20
14	3	10	1	6	20
15	1	5	0	14	20
16	2	10	1	7	20
17	8	12	0	0	20
18	4	13	0	3	20

- Prior and posterior means and 95% equal-tailed credible intervals are

Par.	Prior	Posterior
η_1	0.93 (0.76,1.0)	0.97 (0.92,1.0)
η_2	0.50 (0.26,0.74)	0.30 (0.24,0.35)
θ_1	0.970 (0.860,1.0)	0.998 (0.995,0.999)
θ_2	0.970 (0.860,1.0)	0.997 (0.994,0.999)

- New (ELISA) test has higher sensitivity than the old (IHA) test.

WinBUGS Code:

```
model; { for(i in 2:K){
  x[i, 1:4] ~ dmulti(p[i, 1:4], n[i])
  p[i,1] <- pi[i]*eta11 + (1-pi[i])*theta11
  p[i,2] <- pi[i]*eta12 + (1-pi[i])*theta12
  p[i,3] <- pi[i]*eta21 + (1-pi[i])*theta21
  p[i,4] <- pi[i]*eta22 + (1-pi[i])*theta22
  pi[i] ~ dbeta(alpha,beta)
}
pi[1] ~ dnorm(0,1) x[1,1:4]~dmulti(p[1,1:4], n[1])

  p[1,1] <- theta11
  p[1,2] <- theta12
  p[1,3] <- theta21
  p[1,4] <- theta22

pi[20] ~ dbeta(alpha,beta)

alpha<- mu*gamma
```

```

beta<- (1-mu)*gamma
eta11 <- lambdaD*eta1
eta12 <- eta1 - eta11
eta21 <- gammaD*(1-eta1)
eta22 <- 1 - eta11 - eta12 - eta21

theta11 <- 1 - theta12 - theta21 - theta22
  theta12 <- gammaDc*(1-theta1)
  theta21 <- theta1 - theta22
  theta22 <- lambdaDc* theta1

  eta2 <- eta11 + eta21
  theta2 <- theta22 + theta12

  rhoD<-(eta11 -eta1*eta2)/
sqrt(eta1*(1-eta1)*eta2*(1-eta2))

  rhoDc<-(theta22 -theta1*theta2)/
sqrt(theta1*(1-theta1)*theta2*(1-theta2))

  eta1 ~ dbeta(23.866, 2.721)

```

theta1 ~ dbeta(38.04, 2.146)

lambdaD ~dunif(0,1)

lambdaDc ~ dunif(0,1)

gammaD ~ dunif(0,1) gammaDc ~ dunif(0,1)

mu ~dunif(0,1) gamma ~dgamma(.1,.1) }

```
list(lambdaD=0.5,lambdaDc=0.5,  
gammaD=0.5,gammaDc=0.5,  
eta1=0.97,theta1=0.99,  
pi=c(0.1,0.10,0.10,0.10,0.10,0.10,0.10,0.10,0.10,  
0.10,0.10,0.10,0.10,0.10,  
0.10,0.10,0.10,0.10,0.10,0.1))
```

```
list(K=19, n=c(2128,  
20,20,19,20,12,20,19,29,19,20,20,20,  
20,20,20,20,20,20),  
x=structure(.Data=c(1,4,5,2118, 6, 9, 0,  
5, 3, 10, 0, 7, 1, 3, 0,  
15, 5, 9, 1, 5, 6, 4, 0, 2, 6, 11, 0, 3, 2,
```

```

10, 0, 7, 1, 9, 0, 19,
9, 10, 0, 0, 1, 19, 0, 0, 2, 6, 0, 12, 1,
15, 0, 4, 8, 12, 0, 0,
3, 10, 1, 6, 1, 5, 0, 14, 2, 10, 1, 7, 8,
12, 0 ,0, 4, 13, 0,
3),.Dim=c(19,4)))

```

node	mean	sd	MC	error	2.5%	med	97.5%
eta1	0.9632	0.02077	7.618E-4	0.9097	0.9678	0.990	
eta2	0.2866	0.02851	4.577E-4	0.2313	0.286	0.344	
mu	0.7141	0.06019	0.001447	0.5882	0.7174	0.824	
pi[1]	0.01175	0.9947	0.01449	-1.953	0.02135	1.981	
pi[2]	0.7671	0.09443	0.001617	0.5635	0.7737	0.929	
pi[3]	0.672	0.1037	0.001756	0.4586	0.677	0.861	
pi[4]	0.2762	0.1006	0.001367	0.1014	0.268	0.486	
pi[5]	0.7612	0.09542	0.001807	0.5553	0.7686	0.931	
pi[6]	0.8356	0.1057	0.002509	0.5941	0.8507	0.994	
pi[7]	0.8571	0.08135	0.001714	0.6759	0.8671	0.990	
pi[8]	0.6557	0.1069	0.001794	0.4322	0.6622	0.850	
pi[9]	0.3812	0.08978	0.001303	0.2181	0.3783	0.564	
pi[10]	0.9677	0.03757	0.001201	0.8653	0.981	1.0	
pi[11]	0.9685	0.03841	0.001261	0.8589	0.9829	0.999	
pi[12]	0.4431	0.105	0.001773	0.2452	0.4396	0.654	
pi[13]	0.8099	0.08973	0.001702	0.617	0.8151	0.967	
pi[14]	0.9682	0.03798	0.001095	0.8715	0.9817	0.999	
pi[15]	0.7119	0.1025	0.002037	0.4937	0.7171	0.896	
pi[16]	0.3499	0.1014	0.001176	0.1654	0.3449	0.559	

pi[17]	0.6673	0.1024	0.001833	0.4532	0.6745	0.852
pi[18]	0.9691	0.03826	0.001106	0.8609	0.9832	1.0
pi[19]	0.856	0.08436	0.002141	0.6699	0.8652	0.995
pi[20]	0.7223	0.255	0.003783	0.1245	0.7995	0.999
rhoD	-0.06293	0.09151	0.002778	-0.2335	-0.06526	0.107
rhoDc	0.2562	0.1371	0.002206	0.03997	0.2408	0.545
theta1	0.9967	0.00124	2.09E-5	0.9938	0.9968	0.998
theta2	0.9963	0.00133	2.427E-5	0.9933	0.9964	0.998

Bovine Brucellosis

- Twenty cow herds were randomly sampled in an endemically-infected region of Mexico,
- Within each herd several cows were sampled and tested by two imperfect serologic tests, BAPA (T_1) and Rivanol (T_2),
- Goal is to obtain an estimate of the prevalence distribution of the pathogen in the region.
- Added complication since a cow testing negative using the BAPA test was not tested by the Rivanol test due to high sensitivity for the BAPA test
- The classified counts are tabled below.

i	++	+-	T_{1-}	Tot	i	++	+-	T_{1-}	Tot
1	4	2	61	67	11	1	0	11	12
2	1	1	10	12	12	1	1	60	62
3	2	2	18	22	13	3	0	8	11
4	2	0	69	71	14	3	7	10	20
5	2	0	16	18	15	38	0	18	56
6	2	0	78	80	16	1	0	59	60
7	80	5	62	147	17	3	0	23	26
8	4	5	36	45	18	13	3	38	54
9	10	0	27	37	19	7	0	10	17
10	10	12	96	118	20	1	1	15	17

- Ian Gardner provided the following information for $\mu = E(\pi_i)$, $P(\mu \leq 0.250) = 0.5$ and $P(\mu \leq 0.35) = 0.95$,
- These yield the prior $\mu \sim \text{Beta}(14.85, 43.89)$.
- Prior for γ is $\gamma \sim \Gamma(7.23, 1.28)$ (best guess for 90th %tile is 0.5 and 95% sure that 90th %tile is less than 0.6).

- Accuracy point estimates for the BAPA test are provided by Stemshorn et al (1985): $\hat{\eta}_1 = 0.75$ and $\hat{\theta}_1 = 0.99$.
- Estimates for Rivanol are provided in Hall et al (1987): $\hat{\eta}_2 = 0.91$ and $\hat{\theta}_2 = 1.00$.
- $\eta \sim \text{Dirichlet}(14.4, 0.6, 3.6, 1.4)$ and $\theta \sim \text{Dirichlet}(0.564, 0.937, 0.437, 48.064)$.
- Approximately 90% of the herds are perhaps non to mildly infected ($\pi_i < 0.5$), while less than 10% herds are moderately to severely infected ($\pi_i > 0.5$).

Georgiadis et al Approach: Fixed small number of popula- tions sampled

- The Gibbs sampler consists of sampling independent beta's for all the parameters and the same binomial distributions as before. In short, it's very easy and efficient in this parameterization
- With moderately correlated tests, and except in high accuracy situations (where one of the tests has sensitivity and specificity near one), and with prior information for η_1 and θ_1 that specifies the true mode within ± 0.05 , results are very accurate and tend to be far superior to the conditional independence model.
- WinBugs code is very simple and fast, and need not specify latent (disease status) data.

Issues

- Must assume that test accuracies are the same from one population to the next.
- Diverse populations with different age structures may exhibit the “disease” differently eg. more clinical symptoms which would lead to easier detection eg. higher sensitivity.
- More diversity among populations may lead to more cross-reactivity and consequently varying specificities.
- Can check this by fitting one population at a time with the same priors. If the one-population results give different sensitivities/specificities, then the two-population analysis is in doubt

- One possible solution would be to obtain a higher level of information on staging of disease and to model sensitivities and specificities as functions of stage.
- Leads to more “latent” data eg. stage of individual will generally be unknown (at least in animals)
- Leads to necessity of more subjective input eg. prevalence and test accuracy at different stages
- Alternatively, could attempt to develop models that tie test-accuracy to the prevalence in the population.