

---

# Bayesian Estimation of Aggregate Test Accuracy Based on Different Sampling Schemes

Chun-lung Su and Ian A. Gardner  
Department of Medicine and Epidemiology  
University of California, Davis  
and  
Wesley Johnson  
Department of Statistics  
University of California, Irvine

# Goals

---

- Develop Bayesian models to estimate cluster/herd-level test characteristics
- Herd level sensitivity, specificity, prevalence and predictive values
- Sampling schemes involve a single test case and three sequential test cases
- Cluster-level characteristics are calculated and compared based on different sample size, sampling schemes, individual-level sensitivities, specificities, and cut-off values

- 
- Compare posterior estimates of individual-level and cluster-level characteristics for these sampling schemes with simulated data
  - Illustration with Johne's disease in cattle

# Definitions of Herd Level Characteristics

- Herd level prevalence ( $\pi^h$ ) is the prop of herds that are infected ( $\geq 1$  infected animal)
- Herd level sensitivity ( $\eta^h$ ) is the proportion of infected herds that test positive
- Herd level specificity ( $\theta^h$ ) is the proportion of disease-free herds that test negative
- Herd level PVP is the prop of pos herds that are actually infected
- Herd level PVN is the prop of neg herds that are actually disease/infection free

# Sampling Designs

---

- Herd level tests involve decision rules based on one or more individual animal tests applied to samples of animals
- O1: Herd is pos if the prop of single-test-pos animals in a sample exceeds a cutoff (eg. 0.05)
- S1: Take a second sample if the first proportion exceeds the cutoff and declare the herd pos if both propns exceed their cutoffs.
- S2: Declare the herd pos if either proportion exceeds the cutoff
- S3: Apply two tests to the same sample units and declare the herd positive if both sample propns exceed their cutoffs

- Different choices result in different test accuracy properties
- For example, we can show theoretically that

$$\eta_{O1}^h \geq \eta_{S2}^h \geq \eta_{S1}^h, \quad \theta_{O1}^h \leq \theta_{S2}^h \leq \theta_{S1}^h$$

- We can also show (under a particular case) that

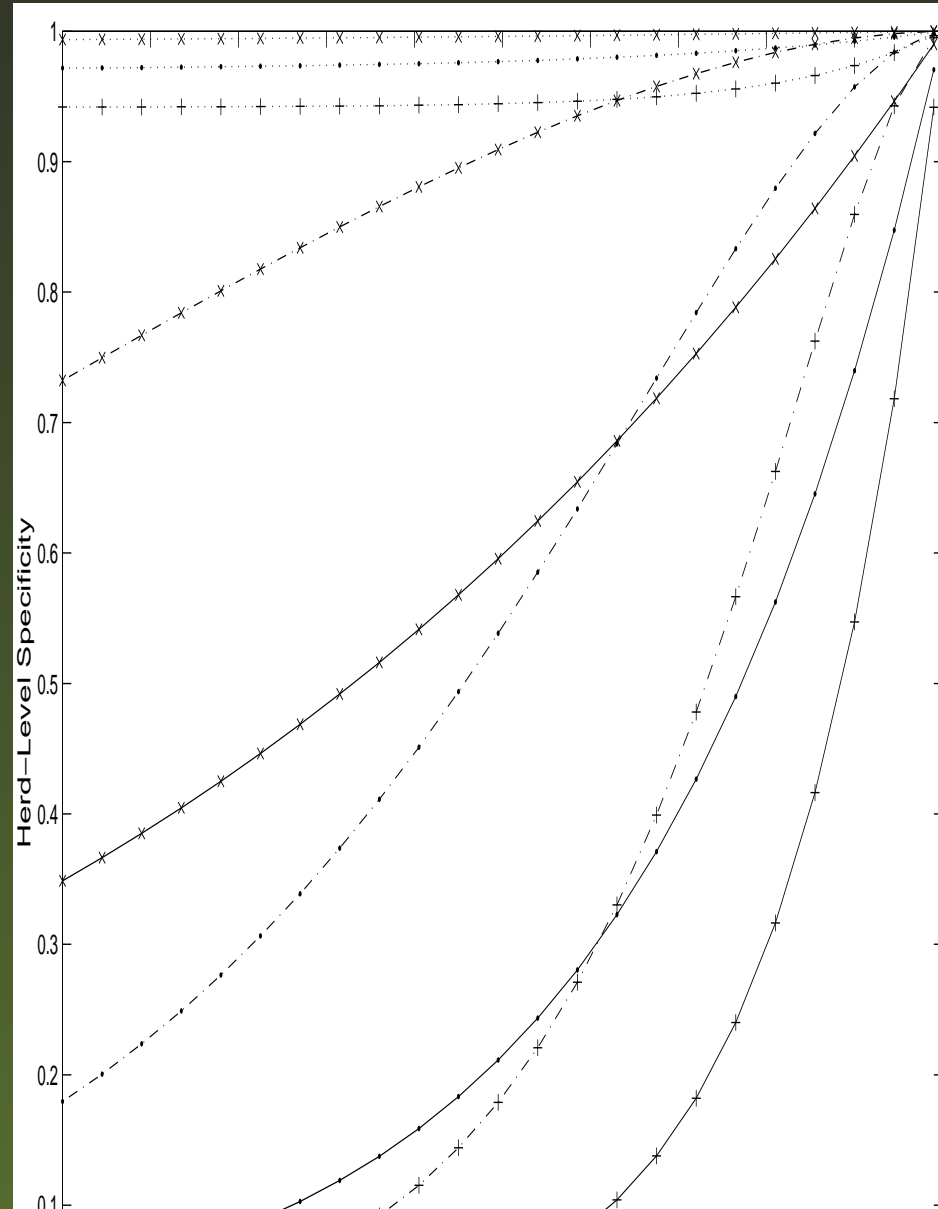
$$\eta_{S3}^h > \eta_{S1}^h, \quad \theta_{S3}^h \leq \theta_{S1}^h$$

- Can also show that if the correlation between tests one and two for non-diseased animals is zero, that these two specificities are equal
- So  $S3$  is preferable to  $S1$  in this instance

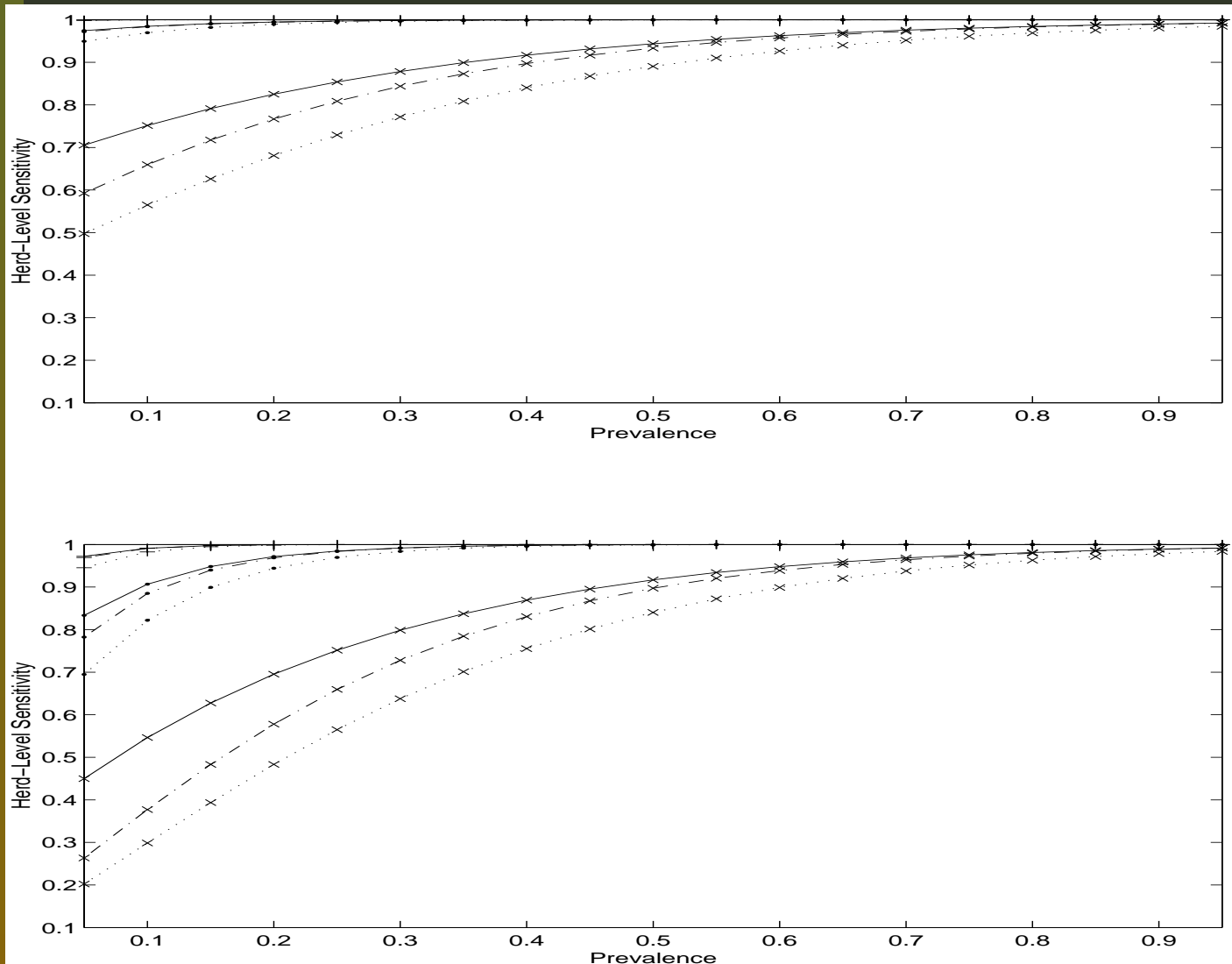
# Behavior of Herd Level Characteristics

- Moreover, in general, the herd level sensitivities are increasing in the animal level sensitivity, and decreasing in the animal level specificity
- The herd level specificities are increasing in the animal level specificities and of course don't depend on the animal level sensitivity
- With a fixed cutoff for the sample proportion of positives, the herd level sensitivity is increasing with the sample size and the herd level specificity is decreasing
- Increasing the cutoff decreases the sensitivity and increases the specificity

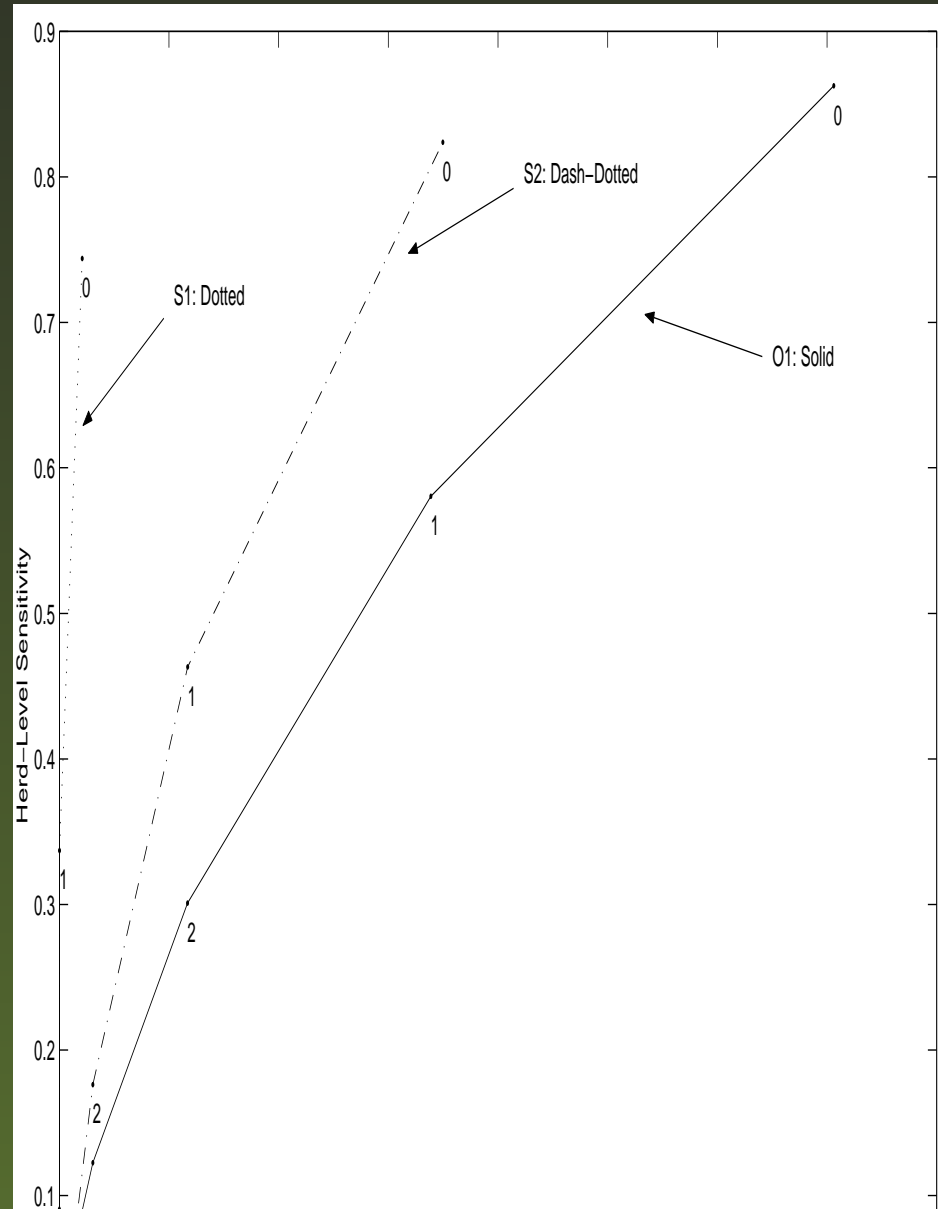
# Sp (x-axis) versus Herd-Level Sp



# HSE (x-axis) versus Prevalence



# 1- HSE (x-axis) versus HSP



# Statistical Model for Sampling $k$ Herds

---

- Let  $T_l$  denote the  $l$ th test, regardless of whether the same units are tested with a different test or different units are tested with the same test
- Let  $n_{lk}$  be the number of animals sampled within each herd for the  $l$ -th test,  $l = 1, 2, k = 1, \dots, n_h$ , where  $n_h$  is the number of herds sampled
- Let  $x_{lk}$  denote the number of test-positives out of  $n_{lk}$  animals for the  $l$ -th test in the  $k$ -th herd
- Suppose the herd size is ten times larger than the sample size such that we can treat the sampling scheme as binomial rather than hypergeometric (Su, et al., 2004)

- If this assumption is violated we propose a hypergeometric-based model (not further discussed here)
- The sensitivity and specificity for the  $l$ -th test are defined as

$$\eta_l = P(T_l^+ | D^+), \theta_l = P(T_l^- | D^-)$$

- Let  $\pi_k$  denote the prevalence for the  $k$ -th herd

# Assumptions

---

- If a herd is “truly noninfected”, then all animals in this herd are noninfected
- Test accuracies remain constant across the herds
- The distribution of prevalences across herds is assumed to be a mixture eg.

$$\pi_k \sim (1 - \tau)\delta_{\{0\}} + \tau * \text{Beta}(\mu * \gamma, (1 - \mu)\gamma)$$

where  $\tau$  is the proportion of the infected herds  
(Hanson et. al., 2003)

- $\mu$  is the average prevalence in the population of herds
- $\gamma$  is a measure of dispersion of herd prevalences since

$$\text{var}(\pi_k) = \frac{\mu(1 - \mu)}{\gamma}$$

# Likelihood and Prior: One Test Case

- The conditional distribution of  $x_{lk}$  is

$$x_{lk} | \pi_k, \eta_l, \theta_l \sim \text{Bin}(n_{lk}, AP_{lk})$$

$$AP_{lk} = \eta_l \pi_k + (1 - \theta_l)(1 - \pi_k) \quad k = 1, \dots, n_h$$

- Prior information is modeled independently using the following distributions:

$$\eta_l \sim \text{Beta}(a_{\eta_l}, b_{\eta_l}) \quad \theta_l \sim \text{Beta}(a_{\theta_l}, b_{\theta_l})$$

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

$$\gamma \sim \text{Gamma}(a_{\gamma}, b_{\gamma})$$

# Prevalence Distribution

---

- Prevalences for infected herds follow a Beta( $\mu\gamma$ ,  $(1 - \mu)\gamma$ ) distribution
- The (posterior) *prevalence distribution* has density

$$p(\pi^* | \text{data}) = \int p(\pi^* | \mu, \gamma) p(\mu, \gamma | \text{data}) d\mu d\gamma$$

# Herd Level Se and Sp

- The herd-level sensitivity and specificity for the  $k$ -th herd based on the *single* test  $T_l$  are

$$\eta_k^h = P(HT_{lk}^+ | H_k^+) = 1 - \text{bincdf}(n_{lk}c_l; n_{lk}, AP_{lk}^*)$$

$$\theta_k^h = P(HT_{lk}^- | H_k^-) = \text{bincdf}(n_{lk}c_l; n_{lk}, 1 - \theta_l)$$

- Define the “average” herd-level sensitivity

$$\eta_{\pi^*}^h(n, c, \eta, \theta, \mu, \gamma) = \int \eta^h(n, c, \pi^*, \eta, \theta) p(\pi^* | \mu)$$

- Define the “typical” herd-level sensitivity as

$$\eta_{\mu}^h \equiv \eta^h(n, c, \mu, \eta, \theta)$$

# Herd Level PVP and PVN

- Herd-level predictive values are:

$$\text{HPV}^+ \equiv P(H^+ | HT^+) = \frac{\tau \eta^h}{\tau \eta^h + (1 - \tau)(1 - \theta^h)}$$

$$\text{HPV}^- \equiv P(H^- | HT^-) = \frac{(1 - \tau)\theta^h}{(1 - \tau)\theta^h + \tau(1 - \eta^h)}$$

- We also consider models With Animal-Level Specificities Varying From Herd to Herd

# Illustration: Simulation

- Compare single test with sequential tests
- Assume the disease of interest is Johne's disease
- Tests one ( $T_1$ ) and two ( $T_2$ ) represent ELISA and fecal culture tests, respectively
- We assumed conditional independence of tests when performed on the same units
- Based on the previous data and the fact that  $\theta_2 \approx 1$
- Thresholds,  $c_1$ ,  $c_2$ , and  $c_{11}$ , are set to be zeros and  $n_{1k}c_{12}$  is set to be one, regardless of the value for  $n_{1k}$

# The Prior

- We adapted opinions of an expert (Dr. Robert Whitlock, University of Pennsylvania) for the proportion of infected herds, animal-level test accuracies for ELISA and fecal culture tests
- We use a modification of the method described in Hanson et. al. (2003) for eliciting prior information on  $(\mu, \gamma)$
- Assume that the expert's best guess for the unknown prevalence distribution is a beta distribution with mode (95th percentile) at 0.15(0.4) or equivalently  $\pi^* | (\mu, \gamma) \sim B(0.2024 \times 13.36, (1 - 0.2024) \times 13.36)$

# More on the Prior

- Set  $(\mu_0, \gamma_0) = (0.2024, 13.36)$   
The prior on  $\mu$  is set to be the beta distribution with mode  $\mu_0$  and 95th percentile 0.35, namely,  $\text{Beta}(7.02, 10.66)$
- Let  $\lambda(\mu_0, \gamma)$  be the 90th percentile of  $\pi^* | \mu_0, \gamma$
- Suppose the expert provided us with  $P(\lambda(\mu_0, \gamma) \leq 0.5) = 0.95$
- This implies  $P(\gamma \geq 3.175) = 0.95$  and then setting the mode equal to  $\gamma_0$ , we obtain the induced prior  $\gamma \sim \text{Gamma}(3.592, 0.194)$

# The Actual Prior

Table 3.1 Prior information for Johnes disease illustrations

	Mode(5th or 95th Percentile)	Beta(Gamma)( $a, b$ )
$\eta_1$	0.20 (0.40)	(4.5, 14.8)
$\theta_1$	0.96 (0.99)	(384.1, 17.0)
$\eta_2$	0.30 (0.45)	(10.2, 22.5)
$\theta_2$	0.999 (0.99)	(370.6, 1.3)
$\tau$	0.5 (0.30)	(3.3, 3.3)
$\pi^*$	0.15 (0.40)	(2.70, 10.66)
$\mu$	0.20 (0.35)	(7.02, 24.72)
$\gamma$	13.36 (36.95)	(3.592, 0.194)

# The Simulation

---

- The “true” values of the parameters were set equal to the prior guesses
- We consider herd sizes of 10, 30, 60, and 100
- We assume  $n_{1k} = n_{2k} = 60$  for all  $k$  and  $n_h = 800$
- We found that there were 765(112) herds or 45,900(6,720) animals for sampling design **S1(S2)** that were followed up with  $T_2$  and 2736 animals (approximately 46 herds, 60 animals per herd) for **S3**.

# Results

---

- All four methods yield posterior results that were close to their true values
- The one exception was  $\tau$ , the proportion of infected herds, for S3, which was underestimated due to lower sensitivities and lower within-herd prevalence.
- Increasing either the sensitivities of  $T_1$  or  $T_2$  or within-herd prevalence resulted in much better posterior estimates of  $\tau$
- For example, with  $\eta_1 = 0.4$  and  $\eta_2 = 0.5$ , with corresponding prior information Beta(7.6, 10.9) and Beta(8.0, 8.0) and with other settings unchanged, the posterior median (95% PI) for  $\tau$  was 0.48(0.43, 0.58)

# Johne's Disease

---

- In 1996, NAHMS sampled approximately 1000 dairy herds with at least 30 milk cows in 20 states
- 30 - 40 cows were sampled from each herd depending on the number of adult cows, and were tested with a commercial ELISA kit to detect serum antibodies to *M. avium* subsp. *paratuberculosis*
- The estimate of herd-level prevalence on U.S. dairy operations was about 22% although this was believed to be an underestimate of prevalence because the ELISA diagnostic test only has a sensitivity of about 20 – 25% in preclinically infected animals

- In 2002, NAHMS repeated the prevalence study in a smaller number of herds (n=105) but using more extensive testing by serum ELISA and culture of feces from the same cow Johne's Disease data
- The number of fecal and ELISA samples was herd-size dependent but ranged up to 360 and 580, respectively
- In order to compare a single test with sequential test **S3**, we selected animals having both ELISA and fecal culture results
- Forty eight herds (5921 animals) satisfy this criterion.