

Use of a Bayesian risk-mapping technique to estimate spatial risks for mare reproductive loss syndrome in Kentucky

James A. Thompson, DVM, DVSc; Stuart E. Brown II, DVM; W. Thomas Riddle, DVM; Janyce C. Seahorn, DVM, MS; Noah D. Cohen, VMD, MPH, PhD

Objective—To estimate spatial risks associated with mare reproductive loss syndrome (MRLS) during 2001 among horses in a specific study population and partition the herd effects into those attributable to herd location and those that were spatially random and likely attributable to herd management.

Animals—Pregnant broodmares from 62 farms in 7 counties in central Kentucky.

Procedure—Veterinarians provided the 2001 abortion incidence proportions for each farm included in the study. Farms were georeferenced and data were analyzed by use of a fully Bayesian risk-mapping technique.

Results—Large farm-to-farm variation in MRLS incidence proportions was identified. The farm-to-farm variation was largely attributed to spatial location rather than to spatially random herd effects.

Conclusions and Clinical Relevance—Results indicate that there are considerable data to support an ecologic cause and potential ecologic risk factors for MRLS. Veterinary practitioners with more detailed knowledge of the ecology in the 7 counties in Kentucky that were investigated may provide additional data that would assist in the deduction of the causal factor of MRLS via informal geographic information systems analyses and suggest factors for inclusion in further investigations. (*Am J Vet Res* 2005;66:17–20)

In 2001, excessive fetal losses were reported among broodmares in central Kentucky as a result of a syndrome that is now well described and often called **mare reproductive loss syndrome (MRLS)**.^{1,2} A combination of unusual environmental factors during April and May of 2001 coupled with unrevealing bacteriologic, virologic, and pathologic findings has led investigators to propose an ecologic-based exposure as the cause of this syndrome.³ Experimental evidence has suggested that **Eastern tent caterpillars (ETCs; *Malacosoma americanum*)** could be an agent or carrier of such an

exposure.^{4,5} Eastern tent caterpillars had not previously been considered a toxic threat, and it remains uncertain whether the outbreak of fetal losses among mares can be attributed to either an unusual dose or unusual potency of ETC larvae. In experimental studies, the pathogenicity of ETC larvae was removed through autoclaving⁴ or irradiation.⁵ Ecologic factors that influence naturally occurring exposures to and pathogenicity of ETCs warrant further investigation.

When ecologic factors contribute to disease causation in farm animals, 2 important problems hamper epidemiologic investigation. First, the preventive effectiveness of within-herd management decisions (such as those regarding which mares to put on pasture) on a farm will depend on the level of the ecologic risk on that farm. Second, when comparing whole-herd management choices made on that farm with those made on others, similar decisions may be imposed by the common ecologic risk on those neighboring herds. Results of 1 study² of MRLS have indicated the potential for varying ecologic risk on a small scale; in that study, mares maintained on 1 broodmare farm had significantly fewer early fetal losses (5%) than mares on 3 other farms (26%, 32%, and 35%). Those investigators proposed an ecologic cause of the fetal losses that may have been absent from the farm with low incidence of MRLS.

Several risk (and preventive) factors that have influence at the pasture and farm levels have been proposed,^{3,6,7} including, among others, feeding practices on the pasture and type of pasture. These factors could be common management practices among farms that are close neighbors. Ecologic factors, including the presence of cherry trees and concentrations of ETCs, have also been proposed; these ecologic risk factors are very likely similar among neighboring farms. Estimation of the spatial correlation and the ability to produce maps that illustrate risk estimates for specific geographic coordinates are now possible.⁸ The geographic risks can be estimated even in the presence of spatially random herd effects.⁹ This potential has developed with recent advances in fully Bayesian methods of analysis.

The fundamental difference between the frequentist and Bayesian statistical approaches is that in a frequentist approach, the parameter of interest is considered to be fixed and the sampled data are considered to be random, whereas in a Bayesian approach, the data are considered to be fixed and the parameter is considered to be random or uncertain.

Received April 12, 2004.

Accepted May 20, 2004.

From the Department of Large Animal Medicine and Surgery, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843-4475 (Thompson, Cohen); Hagyard-Davidson-McGee Associates, 4250 Iron Works Pike, Lexington, KY 40511-8491 (Brown II); Rood & Riddle Equine Hospital, PO Box 12070, Lexington, KY 40580-2070 (Riddle); and Equine Veterinary Specialists, 629 Craig Ln, Georgetown, KY 40324 (Seahorn).

Supported by the Grayson-Jockey Club Research Foundation, the Commonwealth of Kentucky, and the Link Equine Research Endowment.

Address correspondence to Dr. Thompson.

A result of a frequentist analysis expresses uncertainty under further hypothetical data collections. For example, when a frequentist analyst states that a P value is < 0.05 , this means that if the parameter is true (ie, fixed), then the likelihood of gathering more extremely distributed data would be $< 5\%$. With this result, the frequentist analyst will usually reject the initial or fixed parameter value. Aside from being used to reject the original parameter value, the frequentist P value does nothing analytically to identify a more likely parameter value. Although a frequentist analyst may conjecture a more likely parameter value, the frequentist analysis does not support the conjecture; thus, the inability to learn and adapt to new information is a weakness of the frequentist approach. In contrast, a result of a Bayesian analysis will often identify all possible parameter values and a probability plot for these values. Bayesian estimates are influenced by both prior beliefs and information that is derived from a clinical test or clinical testing procedure. Prior beliefs are designated as informative if they are known precisely; thus, they are allowed to influence the posterior likelihood (the test prediction). When the prior beliefs are designated as vague or uninformative, the results are primarily influenced by the data. Thus, the traditional frequentist and the Bayesian approaches produce analogous results when there is no useful prior information. With an uninformative prior belief, the Bayesian approach is limited to inclusion of the same information used in the frequentist approach, namely the data likelihood. The fundamental differences between the 2 approaches and the technical terminology used to describe Bayesian methods have been reported in detail.¹⁰⁻¹² The usefulness of a Bayesian approach to clinical predictions was important among our considerations for statistical modeling in this study. We believe that clinicians are familiar with Bayesian updating applications in evidence-based medicine, especially for incorporating accumulation of evidence.¹⁰ Other reasons for selecting the Bayesian approach included the proven usefulness of Markov chain Monte Carlo (MCMC) sampling methods for complex models.^{11,12}

The objective of the study reported here was to estimate spatial risks associated with MRLS during 2001 among horses in a specific study population and partition the herd effects into those attributable to herd location and those that were spatially random and likely attributable to herd management. It was anticipated that the spatial risk estimates would enhance our understanding of the epidemiology of MRLS.

Materials and Methods

Study population—Veterinarians from 2 large equine clinics in Lexington, Ky, were asked to provide a list (including addresses) of farms to which they provided equine reproductive veterinary ser-

vices. These farms were located throughout 7 counties, including Bourbon, Fayette, Franklin, Harrison, Jessamine, Scott, and Woodford. For each farm, veterinarians from the 2 clinics provided proportions of pregnancy losses for the farm (defined as the number of mares that were verified as pregnant at 28 days after breeding and that subsequently lost the fetus divided by the number of mares verified as pregnant at 28 days after breeding) during 2001. The geographic coordinates (latitude and longitude) were obtained by accessing a United States Postal Service-based database that provided geographic coordinates from street addresses.

Covariogram—An empirical covariogram was plotted as described by Christensen et al.¹³ The empirical covariogram plots covariance for MRLS incidence proportions between farm pairs. The covariogram's function is to graphically represent the extent to which groups of farm pairs that are located close together are more similar than one would expect by chance alone.

Bayesian spatial modeling—The spatial modeling of binomial proportions reflecting the herd proportions of abortions used generalized linear kriging⁸ expanded to include extrabinomial variation at the cluster level.⁹ Briefly, the model used a Bayesian method of inference, with vague prior beliefs and an MCMC implementation. The MCMC implementation was performed by use of a readily available software package.^a The prior beliefs included a flat, improper prior belief for the intercept, for which all real numbers were equally possible, and vague gamma prior beliefs (gamma[0.01, 0.01]) for variance components, including the range and nugget (spatially random herd effect) and spatial effects (spatially dependent herd effect). The initial 500 iterations were discarded to allow for convergence. Convergence was determined by starting 2 chains with disparate initial values to observe for convergence. Adaptive-rejection sampling was performed, and every 100th iteration was retained until 1,000 iterations had been saved. Four models were compared by use of the deviance information criteria (DIC).¹⁴ Better

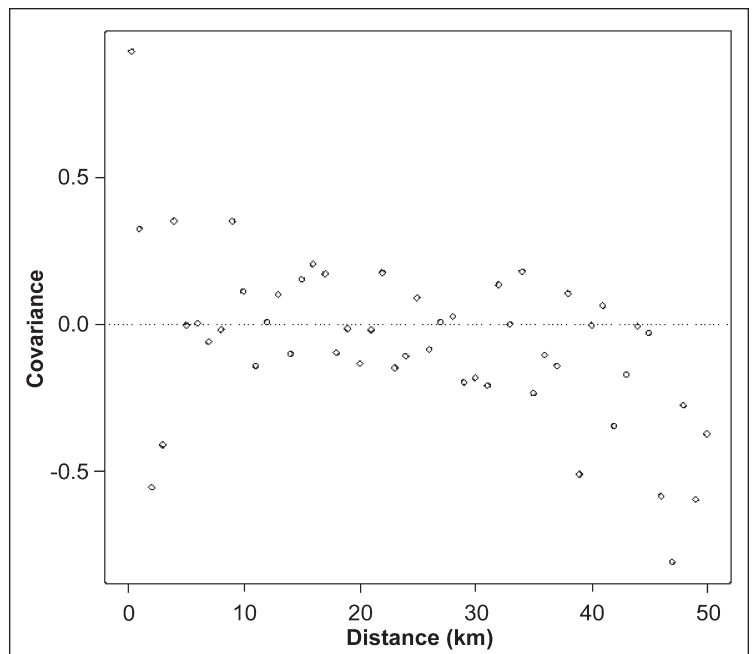


Figure 1—An empirical covariogram (covariance vs distance between groups of farm pairs) generated from data obtained for pregnant broodmares from 62 farms in 7 counties in central Kentucky regarding the outbreak of mare reproductive loss syndrome (MRLS) during 2001. Notice that covariance decreases with increasing distance between groups of farm pairs.

models are identified by low DIC. The 4 models included intercept only (model 1), random effects (model 2), spatial effects (model 3), and random and spatial effects (model 4). The fourth model is sometimes referred to as the convolution model in spatial epidemiology. For all models, the variance function chosen was exponential with the covariance between farm_i and farm_j modeled as a function of the distance between the 2 farms (d_{ij}) and the rate of decline of covariance (ϕ) as follows:

$$f(d_{ij}, \phi) = \exp(-[\phi d_{ij}]).$$

With an exponential rate of declining covariance, it is convenient to define the range as the distance between farm pairs at which the expected covariance is < 5%. This definition of range was used throughout the study, and with the exponential model, the range is equal to 3 times the range parameter (ϕ).

Mapping—Bayesian spatial prediction was performed for a grid of points with each point representing the centroid of a 0.5 × 0.5-km area encompassing the entire 7-county area. The predicted surface was plotted by use of geographic information systems (GIS) software.^b

Results

Data were obtained from 62 farms. The attack rate ranged from 2% to 63% (median attack rate, 19%). The empirical covariogram indicated that the covariance decreased with increasing distance between groups of farm pairs. Rather than approaching zero covariance asymptotically, the covariance decreased across the width of the entire study region (> 50 km; Figure 1).

Four models were compared by use of DIC. The lowest DIC was detected for the spatial model with no spatially random herd effect. The DIC of both the random effects and spatial effects models were considerably lower than that of the intercept only model (Table 1). The model with both spatial and random effects (the convolution model) had lower DIC than the random effects model but higher DIC than the spatial effects model.

Model parameters for the convolution model were estimated in the log-odds scale. The overall intercept around which the spatial and random effects were centered was -1.50, which corresponds to abortions in 18.2% of at-risk mares. The spatially random component is often called the nugget in standard geostatistical analyses, and the mean estimate of the nugget was 0.16. The mean of the spatial covariance, often referred to as the partial sill, was 0.58. The mean practical range was 29.0 km. The mean estimate for the percentage of the farms' SD that was attributable to spatial effects was 60% (Table 2). The 95% Bayesian confidence intervals for the parameters demonstrated considerable uncertainty.

Spatial estimation was performed by use of the model with the lowest DIC (ie, the spatial effects model). The model demonstrated small pockets of both increased and decreased risk (Figure 2). The odds ratio (OR) for each mapped point is considered to be relative to the overall odds of disease corresponding to an incidence proportion of 18.2%. The map of the location ORs indicated that ORs ranged from 0.46 to 3.06; the relative OR between these 2 specific locations was

Table 1—Deviance information criteria (DIC) comparison of binomial models of data obtained for pregnant broodmares from 62 farms in 7 counties in central Kentucky regarding the outbreak of mare reproductive loss syndrome (MRLS) during 2001.

Model	Mean deviance	Deviance at posterior means	Effective parameters	DIC
1. Intercept only	486.0	485.0	1.0	487.0
2. Random effects	284.1	240.9	43.2	327.2
3. Spatial effects	282.9	239.0	43.9	326.8
4. Convolution model	283.0	239.0	44.0	327.0

Table 2—Model parameters and their distributions for the convolution model that includes both location-specific herd-effects and random herd-effects applied to data obtained for pregnant broodmares from 62 farms in 7 counties in central Kentucky regarding the outbreak of MRLS during 2001.

Node	Mean	2.5 percentile	Median	97.5 percentile
Intercept	-1.50	-1.72	-1.51	-1.24
Random herd variance (nugget)	0.16	0.00	0.10	0.52
Maximum of spatial variance (sill)	0.58	0.02	0.31	0.75
Range of spatial variance (km)	28.95	0.00	8.11	161.4
Percentage of herd effects attributable to location	0.60	0.16	0.64	0.89

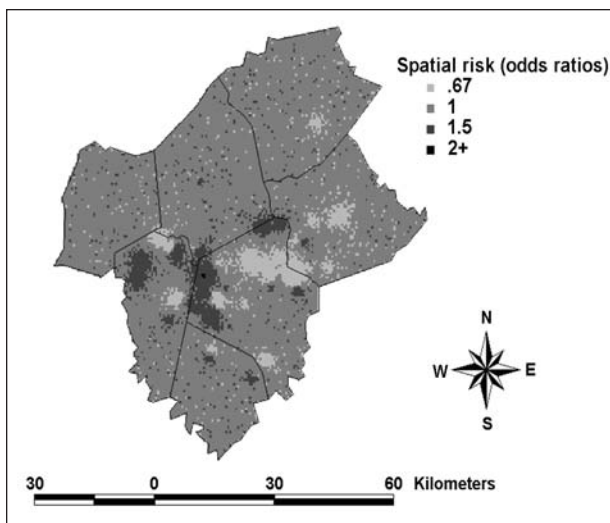


Figure 2—Map of location odds ratios generated from data obtained for pregnant broodmares from 62 farms in 7 counties in central Kentucky regarding the outbreak of MRLS during 2001. Notice that the odds ratios are relative to an overall intercept of an abortion incidence proportion of 18.2%.

6.65. The distance between the highest- and lowest-risk locations (located in Fayette County) was approximately 15 km.

Discussion

The complexity of our analysis was predicated on the need for hierarchical or mixed-level modeling. These models are required when the potential for cross-level confounding or interaction is present. The determination of effectiveness by use of within-herd management differences is conditional on the presence of the ecolog-

ic risk. For example, feeding practices that reduce pasture consumption of a toxin will be biased toward non-significance if areas without the ecologic risk are included in the study. Also, among-herd comparisons will be limited by a lack of independence in herd abortion proportions if the exposure to the ecologic risk is clustered geographically. Neighborhood farms will have related risks with the correlation being greatest for very close neighbors. The results of the present study have indicated the usefulness of fully Bayesian methods in the assessment of ecologic risks and their interaction with herd-level risk factors. These hierarchical models can also be extended to include individual mare risk factors.⁹

In a report² of 1 study of early fetal losses among mares on 4 farms in Kentucky during 2001, the authors proposed that the farm with the lowest incidence of fetal losses was not affected by MRLS. This investigation of a small sample of herds highlights the need for hierarchical modeling. For example, in an investigation of 3 farms that were exposed to an environmental toxicant and 1 farm that was not (ie, a small sample size of herds), the herd management differences detected could often, by chance, appear to explain the risk differences, even though herd location is the true risk. Furthermore, if data obtained from the 4 herds were pooled to study within-herd differences (such as identification of mares that were or were not exposed to pasture), the results would indicate that the risk or benefit of this management would depend on the extent of the environmental toxicant present. A mixed-level model provides 1 solution to analytical complexities of models that include ecologic risks.

Results of the present study have provided evidence of ecologic risk associated with MRLS among horses in Kentucky, as indicated by the correlated risk among neighboring farms. In addition to the local correlation, the OR map contains a large range of relative ORs. There was a relative OR of 6.65 between the lowest and highest risk locations. The size of this estimate suggests that there is considerable potential to define the ecologic differences that may explain these risk differences.

The unusual ecologic conditions in central Kentucky during 2001 have been described, but briefly, unusual climate conditions were accompanied by an unusual number of ETCs living primarily in black cherry trees and other fruit trees (eg, apple, crabapple, and pear).³ It has been postulated that the ETCs contained a toxic agent that contaminated pastures and, when consumed, caused the syndrome either directly or via synergism with some other agent.³ Results of other studies^{3,6,7} have indicated that factors such as whether mares were managed on pasture, the nature of the supplemental feeding practices while on pasture, and the type of pasture were related to development of MRLS. In general, the investigators suggest that these relationships could represent a causal pathway or the risk factors could be markers of other ecologic factors.

The data analyses performed in our study have revealed evidence that MRLS among horses in Kentucky during 2001 was caused by an ecologic risk. The specific source of ecologic risk remains uncertain, as do the benefits of specific pasture management options. Further GIS analyses could include digital maps of various vegetative indices;

soil composition; water resources (eg, streams, ponds, watersheds, and water drainage patterns); ecologic regions (eg, wildlife populations); and specific climatic events. It is even possible that the locations of cherry trees in Kentucky during the MRLS outbreak have been recorded on high-resolution satellite images.

In presenting this brief report regarding the spatial patterns associated with the outbreak of MRLS in Kentucky during 2001, it was the authors' intent to help stimulate further study. With additional MRLS incidence data, perhaps provided by spatial modelers, epidemiologists, and other investigators as well as by farm owners and farm managers, it may be possible to derive more detailed and broader coverage for the map of location ORs. Also, veterinary practitioners with more intimate knowledge of the ecology in the 7 counties in Kentucky included in our study may be able to deduce the causal factor via informal GIS analyses and propose factors for inclusion in further investigations.

-
- a. Win BUGS version 1.4, MRC Biostatistics Unit, Cambridge, UK.
 - b. ArcView GIS 3.2, Environmental Systems Research Institute Inc, Redlands, Calif.
-

References

1. LeBlanc MM, Williams NM, Long W, et al. Mare reproductive loss syndrome. *J Equine Vet Sci* 2003;23:246–248.
2. Morehead JP, Blanchard TL, Thompson JA, et al. Evaluation of early fetal losses on four equine farms in central Kentucky: 73 cases (2001). *J Am Vet Med Assoc* 2002;220:1828–1830.
3. Dwyer RM, Garber LP, Traub-Dargatz JL, et al. Case-control study of factors associated with excessive proportions of early fetal losses associated with mare reproductive loss syndrome in central Kentucky during 2001. *J Am Vet Med Assoc* 2003;222:613–619.
4. Webb BA, Barney WE, Dahlman DL, et al. Eastern tent caterpillars (*Malacosoma americanum*) cause mare reproductive loss syndrome. *J Insect Physiol* 2004;50:185–193.
5. Sebastian M, Gantz MG, Tobin T, et al. The mare reproductive loss syndrome and the Eastern Tent Caterpillar: a toxicokinetic/statistical analysis with clinical, epidemiologic, and mechanistic implications. *Vet Ther* 2003;4:324–339.
6. Cohen ND, Carey VJ, Donahue JG, et al. Case-control study of late-term abortions associated with mare reproductive loss syndrome in central Kentucky. *J Am Vet Med Assoc* 2003;222:199–209.
7. Cohen ND, Donahue JG, Carey VJ, et al. Case-control study of early-term abortions (early fetal losses) associated with mare reproductive loss syndrome in central Kentucky. *J Am Vet Med Assoc* 2003;222:210–217.
8. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. *J R Statist Soc B* 1998;47:299–350.
9. Diggle P, Moyeed R, Rowlingson B, et al. Childhood malaria in the Gambia: a case-study in model-based geostatistics. *J R Stat Soc C* 2002;51:493–506.
10. Ashby D, Smith AF. Evidence-based medicine as Bayesian decision-making. *Stat Med* 2000;19:3291–3305.
11. Congdon P. *Bayesian statistical modeling*. West Sussex, UK: John Wiley & Sons, 2001;18–29.
12. Carlin BP, Louis TA. *Bayes and empirical Bayes methods for data analysis*. New York: Chapman & Hall/CRC, 2000;1–34.
13. Christensen OF, Moller J, Waagepetersen R. Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo. Research report R-00-2009. Aalborg, Denmark: Aalborg University, 2000.
14. Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc B* 2002;64:1–34.