I. **Project Title:** Developing surveillance systems for emerging and foreign animal diseases of swine

NPB project number: SHIC #17-141

Principal Investigator: Jeff Zimmerman, DVM PhD

Institution: Iowa State University

Date Report Submitted: May 9, 2019

II. Industry Summary:

Effective surveillance should efficiently collect data for production and/or business planning, document freedom from specific pathogens, and guide a rapid, effective response to emerging and/or FADs. Current on-farm or regional surveillance programs routinely fail to meet these targets. In part, this is because the industry has changed over time and no longer conforms to the assumptions under which our surveillance systems were originally designed. As a result, surveillance either is not done or is done ineffectively.

<u>On-farm surveillance</u> The statistical theory on which on-farm surveillance was originally based assumes: (1) subjects (pigs) are independent, (2) all pigs have an equal probability of being selected for sampling, and (3) the farm has a stable, homogenous pig population. Traditional farms fit these assumptions - hence the "30 sample" approach worked in the PRV eradication program - but current swine production systems do not.

Contemporary production systems differ from traditional farms in ways that are incompatible with traditional surveillance: (1) Today's production systems are much larger than in the past. Iowa farms averaged a total inventory of 250 animals in 1980 (Flora et al., 2007) versus 3,265 according to a study commissioned by the Iowa Pork Producers Association in 2016 (https://www.iowapork.org/study-iowa-pork-industry-remains-important-economic-driver/). (2) Pigs no longer run free in pastures or feedlots. Instead, management of large swine populations requires physical segregation by age and stage into buildings and pens. (3) Swine populations on modern farms experience rapid turnover of animals and frequent introduction of new animals - often of a different disease status. Thus, current production systems rely on extensive movement of pigs, people, trucks, and feedstuffs between sites. This connects distant places/populations and facilitates the rapid movement of pathogens between them.

<u>Surveillance at the farm level</u> In NPB #13-157 (Rotolo et al., 2017), we showed that disease on contemporary farms moved in a spatiotemporal fashion (non-random). This led us to develop new surveillance guidelines for on-farm surveillance based on spatial (non-random) sampling. This "fixed spatial sampling" approach is being used in the U.S. and elsewhere.

<u>Surveillance at a regional level</u> Efficient regional surveillance is fundamental to detecting the incursion of new pathogens and in monitoring regional disease control/elimination projects. Thus, the current project moved surveillance to the regional level with the objective of developing more efficient regional surveillance methods (fewer samples, but better detection). In this project, we tested the hypothesis that disease exhibited a spatiotemporal pattern of spread at the regional level (just as we saw on farms). The emergence of PEDV in April 2013 provided the opportunity to examine this question.

Using PEDV testing results from the Iowa State University Veterinary Diagnostic Laboratory (at the county level to protect client confidentiality), we found a spatiotemporal pattern of PEDV spread. This means that, just as for on-farm sampling, the assumptions upon which regional surveillance have been based do not hold in today's world. This is important because it means that new guidelines for regional surveillance should be developed using statistically-appropriate modelling to account for the spatial and temporal correlation in disease spread. As a first effort in developing new guidelines, we have shown that spatially balanced sampling through generalized random–tessellation stratified (GRTS) gives a higher power of detection than traditional simple random sampling (SRS) using simulation studies mimicking real PEDV data.

Thus, our research has provided a better understanding of the spatiotemporal nature of disease spread. Initial assessment showed that use of a spatially balanced sampling scheme improved the power of disease detection and the efficiency of the disease surveillance.

III. Keywords

Emerging and foreign disease, spatiotemporal, spatially balanced sampling, surveillance, statistics.

IV. Scientific Abstract

Effective surveillance should efficiently collect data for production and/or business planning, document freedom from specific pathogens, and guide a rapid, effective response to emerging and/or FADs. Efficient regional surveillance is fundamental to detecting the incursion of new pathogens and in monitoring regional disease control/elimination projects. However, the industry has changed over time and no longer conforms to the assumptions under which our surveillance systems were originally designed.

In this project, we tested the hypothesis that disease exhibited a spatiotemporal pattern of spread at the regional level. Using PEDV testing results from the ISU VDL (at the county level to protect client confidentiality), we found a spatiotemporal pattern of PEDV spread. Subsequently, we found that spatially balanced sampling through generalized random–tessellation stratified (GRTS) gave a higher power of detection than traditional simple random sampling (SRS) using simulation studies mimicking real PEDV data. Thus, thus research provides a better understanding of the spatiotemporal feature in disease spread. Application of the spatially balanced sampling scheme is shown to improve the power of disease detection and the efficiency of the disease surveillance.

V. Introduction

"Representative sampling", i.e., testing a subset of the population, was first described in 1895 (Kruskal and Mosteller, 1980), but not widely applied to swine surveillance until the U.S. pseudorabies (PRV) eradication program in the 1980's. The convention of sampling 30 "randomly selected" animals for surveillance is a legacy of the PRV eradication programs. This number is loosely based on a sample size providing a 95% probability of detecting ≥1 positive animals in a population with ≥10% prevalence (Anderson et al., 2008).

The statistical theory on which on-farm surveillance was originally based assumes: (1) subjects (pigs) are independent, (2) all pigs have an equal probability of being selected, and

(3) the farm has a stable, homogenous pig population. Traditional farms fit these assumptions, but current swine production systems do not (see Industry Summary).

In NPB #13-157 (Rotolo et al., 2017), we showed that disease on contemporary farms moved in a spatiotemporal fashion (non-random) and developed new surveillance guidelines based on spatial (non-random) sampling. This "fixed spatial sampling" approach is being used in the U.S. and elsewhere.

<u>Surveillance at a regional level</u> Efficient regional surveillance is fundamental to detecting the incursion of new pathogens and in monitoring regional disease control/elimination projects. Thus, the current project moved surveillance to the regional level with the objective of developing more efficient regional surveillance methods (fewer samples, but better detection). In brief, the logic behind spatially balanced sampling is as follows:

Populations have a spatial structure e.g., neighboring farms are more likely to have the same infectious disease status vs farms distant from each other.

Because neighbors tend to be similar in disease status, efficient sampling design should spread out sampling so that neighboring farms are not sampled. How far apart sampling should occur is one of the calculations in the GRTS approach. However, we also need randomization at some level because it optimizes detection. GRTS achieves both by dividing the area into grids and then randomly selecting sites within grids.

VI. Objectives

The focus of this project was on technical aspects/experimental design related to the development of more efficient and cost-effective surveillance systems, with an emphasis on preparing the swine industry for detecting and eliminating emerging and/or foreign animal diseases (FAD). Herein we report results of the **assessment of "spatially balanced sampling" (generalized random tessellation stratified design - GRTS) using ISU VDL PEDV test results.**

Step 1: Developing the database.

Diagnostic data from the PEDV outbreak beginning in April 2013 was used to assess the efficacy of a spatially balanced sampling system for disease surveillance and detection. This dataset is unique because it represents the emergence of a disease into a completely naïve population. This is similar to the situation we will face in the event of the introduction of an FAD.

Client confidentiality was maintained throughout by removing client identifiers and analyzing the data at the county level. Records included in the PEDV dataset were based on case submissions to the Iowa State Diagnostic Laboratory (ISU VDL) from April 2013 to April 2017. The criteria for cases included any cases submitted for PEDV testing regardless of assay selection and/or sample type from swine facilities. After evaluating the data for quality, approximately 222,000 records were considered suitable for analysis.

Step 2: Analysis of the VDL data.

Based on our previous work, we knew that spatial distribution is an important consideration in designing an efficient survey or monitoring program for the detection of infectious disease. Often, spatially balanced sampling, that is collection of samples that are more or less evenly distributed over the area of interest, is more efficient than simple random sampling. In Step 2, we compared random sampling to spatially-based sampling using historic PEDV testing data. Which design would be most effective for disease surveillance in the face of a transboundary disease outbreak?

The strategy followed is as follows:

Random sample selection were compared with sample selection based on a generalized random–tessellation stratified (GRTS) sampling design. Stevens and Olsen (2004) proposed GRTS sampling as a means to achieve spatially balanced sampling of natural resources. Based on creating a function that maps two-dimensional space into one-dimensional space, thereby defining an ordered spatial address, the method uses restricted randomization to randomly order the addresses. Thus, systematic sampling along the randomly ordered linear structure results in a spatially well-balanced random sample. The assumption is that samples are more or less evenly dispersed over the extent of the resource. Theoretically, this provides more efficient surveillance than simple random sampling.

In this project, we applied the GRTS sampling design to sampling a finite farm population in Iowa over the course of the PEDV outbreak that began in April 2013 based on PEDV testing data. The data for number of farms within each county and relevant farm sizes was acquired through several resources, including the National Agriculture Statistics Service (NASS) and the Department of Natural Resources (DNR).

An unstratified and unequal probability GRTS survey design was applied using the "sp" package in R. For a given sample size "n", the unequal selection probability for each county was set to be proportional to the total farm number within each county. Samples were selected and spatial balance evaluated using the definitions in the paper. For a sample of points S = ($s_1,\ldots,s_n$), let $v_i = \int_{\phi_i} \pi(s)d\phi(s)$ be the total inclusion probability of the voronoi polygon ϕ_i for the i-th sample point. Then $\zeta = \mathrm{Var}(v_i)$ can be used for evaluating the spatial balance. This GRTS sampling design was then compared with simple random sampling by calculating the corresponding spatial balance factor ζ.

The exact locations of the farms was denoted as "x" and the PEDV status of each farm analyzed using a spatial generalized liner mixed model (SGLMMs). Let Y_{ij} be the disease status of farm j in county i, x_{ij} be the location of farm j in county i. S(x) be the spatial surface which is a stationary Guassian process with:

$$\mathrm{E}\big(\mathrm{S}(\mathrm{x})\big) = 0, \mathrm{cov}\big(\mathrm{S}(\mathrm{x}), \mathrm{S}(x')\big) = \sigma^2\rho(x - x') + \tau^2 I_{\{x=x'\}}$$

use correlation function $\rho(\mathrm{h}) = \exp(-\mathrm{h}/\phi)$, where h is defined as the Euclidean distance between x and x', i.e. $\mathrm{h} = ||\mathrm{x} - x'||$. Then conditionally on S(x), assume that Y_{ij} are mutually independent Bernoulli variables, which becomes:

$$Y_{ij}|\mathrm{S}\big(x_{ij}\big) \sim \mathrm{Bernoulli}(\mathrm{p}(x_{ij}))$$

where $\mathrm{p}(x_{ij})$ is the probability that farm j in county i is diseased and

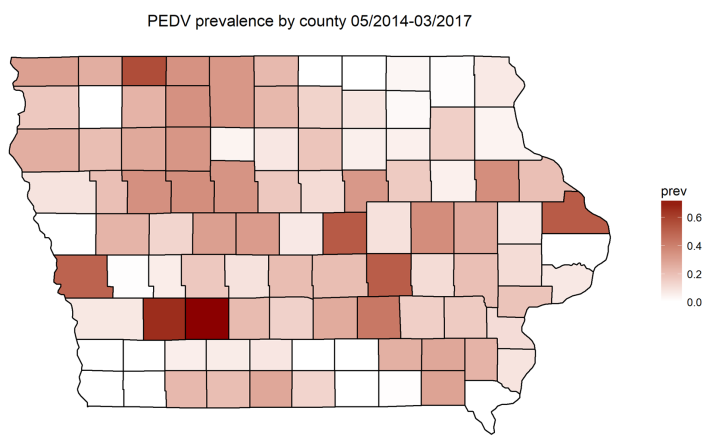
$$\mathrm{logit}\Big(\mathrm{p}\big(x_{ij}\big)\Big) = \mu + \mathrm{S}\big(x_{ij}\big).$$

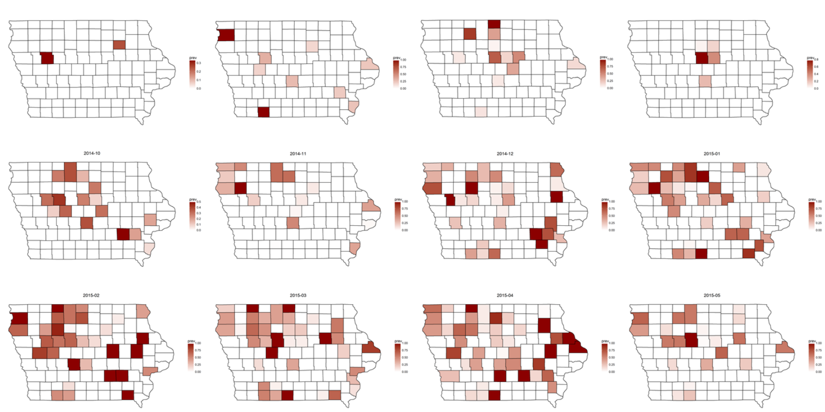
References. Stevens DL Jr, Olsen AR. 2004. Spatially-balanced sampling of natural resources. J Am Stat Assoc 99:262-278.

VII. Results

Assessment of "spatially balanced sampling" (generalized random tessellation stratified design - GRTS) using ISU VDL PEDV test results.

Binary (positive/non-positive) PEDV diagnostic test results from samples submitted to ISU Veterinary Diagnostic Laboratory were used in the analyses. We initially tested the hypothesis that disease exhibited a spatiotemporal pattern of spread at the regional level (just as we saw on farms). The emergence of PEDV in April 2013 provided the opportunity to examine this question. Using PEDV testing results from the ISU VDL (at the county level to protect client confidentiality), we found a spatiotemporal pattern of PEDV spread (Figure above, 06/2014 – 05/2015). This means that, just as for on-farm sampling, the assumptions upon which regional surveillance is currently based do not hold in today's world.





Spatio-temporal PEDV prevalence distribution. ISU VDL data 06/2014 – 05/2015

Thereafter, we evaluated the performance of spatially balanced sampling through simulation studies. Parameters in the simulation, included the strength of spatial correlation and disease transmission rates, as estimated from PEDV diagnostic data. Data simulation and sampling were programed using statistical software R. The GRTS sampling design was compared with simple random sampling (SRS) in terms of the power of detection statewide. Simulations were performed at various settings by controlling the average prevalence p_0, number of sites sampled n_1, and number of samples per site n_2.

Simulation at p_0=0.1%, n_1=10 showed that GRTS consistently performed better than SRS in terms of power of detection. The improvement in power is due to the spatially balanced distributions of samples over the whole state, whereas in SRS there is a possibility of spatially clustered (unbalanced) sampling

n_2	Month	GRTS	SRS
5	1	0.063	0.049
5	2	0.091	0.072
5	3	0.126	0.091
10	1	0.114	0.095
10	2	0.151	0.139
10	3	0.209	0.181

VIII. Discussion

Our results show that new guidelines for regional surveillance should be developed using statistically-appropriate modelling to account for the spatial and temporal correlation in disease spread. This conclusion is supported by the fact that that spatially balanced sampling through generalized random–tessellation stratified (GRTS) provided a higher power of detection than traditional simple random sampling (SRS) using simulation studies mimicking real PEDV data. Thus, application of the spatially balanced sampling scheme(s) can improve the power of disease detection and the efficiency of the disease surveillance. Future research should focus on 1. evaluating and adjusting for the effects of covariate variables, such as farm type, seasonality, distance to highway; 2. evaluation of other spatially balanced sampling methods, including Local Pivotal Method, spatially correlated Poisson sampling, cube method (for balanced sampling) and the local cube method. 3. developing methods for spatially balanced sampling over time.