# IDENTIFYING QUESTIONS IN THE AMERICAN ASSOCIATION OF SWINE VETERINARIAN'S PRRS RISK ASSESSMENT SURVEY THAT ARE IMPORTANT FOR RETROSPECTIVELY CLASSIFYING SWINE HERDS ACCORDING TO WHETHER THEY REPORTED CLINICAL PRRS OUTBREAKS IN THE PREVIOUS 3 YEARS

DJ Holtkamp[1], H Lin[2], C Wang[1,2], AM O'Connor[1]

[1]*Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA USA,* [2]*Department of Statistics, College of Liberal Arts and Sciences, Iowa State University, Ames, IA USA,* holtkamp@iastate.edu

## Introduction

The American Association of Swine Veterinarians (AASV) Production Animal Disease Risk Assessment Program (PADRAP) is a web-based program that offers a set of risk assessment surveys being used by veterinarians who are members of the AASV. AASV members use PADRAP to help producers systematically assess risk factors that may be associated with clinical outcomes. Completed surveys are added to the dataset of completed surveys. One of the surveys included in PADRAP is the porcine reproductive and respiratory syndrome (PRRS) Risk Assessment for the Breeding Herd. The aim of the study was to categorize questions in version 2 of the PRRS Risk Assessment for the Breeding Herd survey as important or unimportant for classifying herds according to whether they reported clinical PRRS outbreaks in the previous 3 years.

## Materials and Methods

Surveys from 896 sow herd sites in the United States and Canada completed between March 2005 and March 2009 were included in the analysis. The survey contained a large number of questions with a complex correlation structure among the questions where responses for several questions were dependent upon responses to others. To overcome these issues, an approach was developed using a series of statistical methods including random forest, principle component analysis, logistic regression and receiver operating characteristic (ROC) analysis to classify the herds using the questions in the survey as explanatory variables. Questions were ranked by importance and systematically excluded from least important to most important. The questions excluded, without significantly affecting the performance of the model for classifying herds were categorized as unimportant.

## Results

After first removing questions with the lowest Gini score in multiple of tens a significant decrease in area under the curve (AUC) (-0.02, $p < 0.01$) occurred when the 40 questions with the lowest Gini score were excluded but no significant decrease in AUC (0.00, $p = 1.00$) when 30 were excluded. Thus we further studied the questions with the 31st to the 39th lowest Gini score by removing them one at a time. Model summaries including the number of questions, number of principal components, point and 95% interval estimate of AUC for all models studied are reported in Table 1. Thirty-eight of the 127 questions analyzed could be removed from the model without significantly reducing the AUC and therefore categorized as unimportant. Sections of the survey where a large number of questions were categorized as unimportant included 1) entry of semen into the breeding herd and 2) transportation of live animals. Sections with a high percentage of questions identified as unimportant included 1) characteristics of the site, 2) disposal of dead animals and waste management and 3) employee and visitors.

**Table 1.** Summary of area under the curve (AUC) of all models and difference between AUC of Models 10 through 120 (reduced models) and Model 0 (full model).

| Model | Number of questions | Number of principal components | AUC | 95%CI | v.s. Model 0 P-value | v.s. Model 0 95%CI |
|---|---|---|---|---|---|---|
| Model 0 | 127 | 371 | 1.00 | (1.00, 1.00) | ----- | ----- |
| Model 10 | 117 | 355 | 1.00 | (1.00, 1.00) | 1.00 | (0.00, 0.00) |
| Model 20 | 107 | 337 | 1.00 | (1.00, 1.00) | 1.00 | (0.00, 0.00) |
| Model 30 | 97 | 316 | 1.00 | (1.00, 1.00) | 1.00 | (0.00, 0.00) |
| Model 31 | 96 | 314 | 1.00 | (1.00, 1.00) | 1.00 | (0.00, 0.00) |
| Model 32 | 95 | 313 | 1.00 | (1.00, 1.00) | 1.00 | (0.00, 0.00) |
| Model 33 | 94 | 311 | 1.00 | (1.00, 1.00) | 0.41 | (0.00, 0.00) |
| Model 34 | 93 | 308 | 1.00 | (1.00, 1.00) | 0.13 | (0.00, 0.00) |
| Model 35 | 92 | 305 | 1.00 | (1.00, 1.00) | 0.31 | (0.00, 0.00) |
| Model 36 | 91 | 302 | 1.00 | (1.00, 1.00) | 1.00 | (0.00, 0.00) |
| Model 37 | 90 | 299 | 1.00 | (1.00, 1.00) | 1.00 | (0.00, 0.00) |
| Model 38 | 89 | 295 | 1.00 | (1.00, 1.00) | 0.28 | (0.00, 0.00) |
| Model 39 | 88 | 292 | 0.98 | (0.97, 0.99) | <.01 | (-0.03, -0.01) |
| Model 40 | 87 | 290 | 0.98 | (0.97, 0.98) | <.01 | (-0.03, -0.02) |
| Model 50 | 77 | 259 | 0.94 | (0.93, 0.96) | <.01 | (-0.07, -0.04) |
| Model 60 | 67 | 234 | 0.94 | (0.92, 0.95) | <.01 | (-0.08, -0.05) |
| Model 70 | 57 | 207 | 0.92 | (0.91, 0.94) | <.01 | (-0.09, -0.06) |
| Model 80 | 47 | 173 | 0.91 | (0.89, 0.93) | <.01 | (-0.11, -0.07) |
| Model 90 | 37 | 143 | 0.89 | (0.87, 0.91) | <.01 | (-0.13, -0.09) |
| Model 100 | 27 | 105 | 0.89 | (0.86, 0.91) | <.01 | (-0.14, -0.09) |
| Model 110 | 17 | 67 | 0.85 | (0.83, 0.88) | <.01 | (-0.17, -0.12) |
| Model 120 | 7 | 26 | 0.81 | (0.78, 0.84) | <.01 | (-0.22, -0.16) |

## Conclusions and Discussion

The analysis presented demonstrates how a program like PADRAP, that is supported by a professional association and used by a community of veterinarians, can generate valuable data that contributes to our understanding of the relative importance of risk factors for disease. Specifically, the results elucidate the relative importance of risk factors and areas of risk factors for clinical outcomes and removing unimportant questions may reduce the time required to complete the survey without affecting the quality of information obtained. The results show how the data may be used to iteratively improve the quality of the survey by decreasing the reliance upon expert opinion to identify questions that should remain in the survey.