Consolidating diverse and disperse datasets to reveal production system-specific drivers of swine performance under field conditions

E. S. Magalhães[1], DVM; D. S. Rosero[2], PhD; T. Donovan[2], DVM; P. Thomas[3], DVM, MS; C. A. A. Moura[1], DVM, PhD; G. Trevisan[1], DVM, MBA, PhD; D. J Holtkamp[1], DVM, MS; J. J. Zimmerman[1], DVM, PhD, DACVPM; C. Rademacher[1], DVM; D. C. L. Linhares[1], DVM, MBA, PhD
[1]Veterinary Diagnostic and Production Animal Medicine Department, Iowa State University, Ames, Iowa; [2]The Hanor Company, Enid, Oklahoma; [3] Iowa Select Farms, Iowa Falls, Iowa.

Introduction

Understanding swine wean-to-finish (W2F) mortality, a key swine performance indicator, is not a simple task due to the influence of multiple factors that occur at different points in time from breeding of specific W2F cohorts until marketing, a process that takes about nine months. Therefore, a comprehensive understanding of whole-herd risk factors associated with W2F mortality is required by integrating diverse data, from multiple disperse sources. Nowadays, with the development of new technologies, an increasing volume of data is generated at an accelerating pace. This presents a complex challenge to livestock production systems to manage and utilize their big data in an informative manner. Therefore, this paper describes important concepts and the challenges when standardizing and integrating multiple data streams from swine production systems, with the objective of creating a robust integrated dataset that can be utilized for different analytical approaches (e.g., assessment of causal pathways, forecasting results, and measure impact of diseases or the effect of health interventions).

Process to build datasets

Our research team has been working with swine production systems to clean and integrate their existing data streams into a single consolidated dataset having information on cohorts of growing pigs carrying information from breeding to farrowing to weaning to finishing, constructing therefore flow of all the groups weaned and later marketed, from a specific period of time.

The process of constructing production system-specific master-tables utilized SAS 9.4, a data management and analytics software. More specifically, SAS scripts scanned dates and unique identifiers across data streams (premises id. and lots id.) and pig movement reports to build cohorts of W2F groups. Whole-herd health & production data incorporated into the W2F cohorts consisted of several datapoints including: sow farm ID, health status for selected diseases including PRRSv, PEDv and *Mycoplasma hyopneumoniae;* sow farm productivity including farrowing rate, prenatal and pre-weaning, abort rate, total born, pigs weaned per sow, and weaning age; type of grow-finish flow (e.g., nursery and finishing, single or double-stock wean-to-finish); feed mill source; and supervisors. When W2F cohorts were originated from multiple sources, the model used weighted averages from source farms. Additionally, to source farm information, the model also incorporated into corresponding W2F cohort diagnostic data from the ISU-VDL. All processes of data integration are fully automated, based on running SAS scripts.

The first time that we conducted this process we identified particularities of the production system related to their internal management strategies for breeding herds and growing groups (e.g., flows, diagnostic monitoring protocols, productivity targets, stocking strategies), as well as differences related to data collection and software. When working with another production system similar challenges were identified, however, with a new particularity of how data was available, in this case with all the information being available in a cloud database. On the other hand, similarities concerning the type of data stream were identified, such as productivity performance reports for closeouts and breeding herds, diagnostic datasets, economic performance reports, and packer plant information, as well as infrastructure, internal management, and health status internal records.

Applications for the aggregated dataset

One of the key benefits from having an integrated master-table is to enable measuring the effect of sow farm health & productivity data on the subsequent W2F performance. A great portion (70 to 80%) of W2F mortality variability was explained by sow farm data, supporting the 'field perception' that quality of weaned pig is highly predictive of subsequent W2F performance. For example, W2F mortality increased proportionally to decrease in farrowing rate, increased pre-weaning mortality, and positive-unstable status for PRRS or *M hyopneumoniae*.[1] Another important finding was that W2F mortality was 1.5% higher in groups that had tissue submitted for diagnostic testing for one system, compared to groups without tissue submission. The timing of tissue submission was also relevant: the earlier during grow-finish period that samples were submitted to diagnostic investigation, the greater was the W2F mortality. A similar approach of data integration has been applied in different studies,[2,3] however without integrating the data from breeding herds to the growing phase, and also without automated data integration and analysis process.

Another key benefit of having a consolidated master-table is the opportunity to be constantly updated with incoming information of weaned groups and marketed cohorts, in an automated fashion, enabling thus constant data integration and analysis. Furthermore, the integrated dataset can be used for internal research purposes (i.e., internal field trials to measure specific interventions or strategies), with the opportunity to control for other recorded risk factors, as well as to understand significant interactions (e.g., weaning weight and age, disease status at weaning, and stocking density). Lastly, but not least important, once the aggregated dataset is being utilized by the production system, the inclusion of new variables/predictors is simple (e.g., vaccination, medication or management procedure at processing), as well as other outcomes, such as feed conversion, full value pig, and the cost/revenue analysis by developing a system-specific economic model.

Conclusions

Multiscale data consolidation within a swine production system supports precision swine health & production management strategies, by providing the capability of collectively analyze field data that was previously disperse, identifying the major whole-herd drivers of swine performance. The interactions between health, productivity, and environmental factors change dramatically across systems and time. Therefore, measuring the true effect of different drivers of mortality under specific field conditions requires on-going, on-farm analysis by collecting, integrating, and analyzing data systematically and routinely.

References

1. Magalhaes ES, Thomas P, Moura CAA, Trevisan G, Rademacher CJ, Linhares DCL, editors. Whole-herd drivers of wean-to-finish mortality in a USA swine production system. Allen D Leman Swine Conference; 2019 Sep 14-17; Saint Paul, MN: University of Minnesota.

2. Agostini, P. S., A. G. Fahey, E. G. Manzanilla, J. V. O'Doherty, C. De Blas, and J. Gasa. 2014. Management factors affecting mortality, feed intake and feed conversion ratio of grow-finishing pigs. Animal. 8:1312–1318. **doi:10.1017/S1751731113001912**.

3. Oliveira J, Yus E, Guitián FJ (2009) Effects of management, environmental and temporal factors on mortality and feed consumption in integrated swine fattening farms. Livestock Science 123, 221–229.