

Whole-herd risk factors associated with wean-to-finish mortality under the conditions of a Midwestern USA swine production system

Edison S. Magalhães¹, Jeffrey J. Zimmerman¹, Pete Thomas³, Cesar A. A. Moura³, Giovanni Trevisan¹, Derald J. Holtkamp¹, Chong Wang^{1,2}, Christopher Rademacher¹, Gustavo S. Silva¹, Daniel C. L. Linhares^{1*}

¹ Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa;

² Department of Statistics, College of Liberal Arts and Sciences, Iowa State University, Ames, Iowa;

³ Iowa Select Farms, Iowa Falls, Iowa;

***Corresponding author**

Daniel C. L. Linhares

2221 Lloyd, 1809 S Riverside Drive

Department of Veterinary Diagnostic and Production Animal Medicine

Iowa State University, Ames, Iowa 50011

dcl.linhares@gmail.com; linhares@iastate.edu (DCLL)

Abstract

Swine wean-to-finish (W2F) mortality is a multifactorial, dynamic process and a key performance indicator of commercial swine production. Although swine producers typically capture the relevant data, analysis of W2F mortality risk factors is often hindered by the fact that, even if data is available, they are typically in different formats, non-uniform, and dispersed among multiple unconnected databases.

In this study, an automated framework was created to link multiple data streams to specific cohorts of market animals, including sow farm productivity parameters, sow farm and growing pig health factors, facilities, management factors, and closeout data from a Midwestern USA production system. The final dataset (master-table) contained breeding-to-market data for 1,316 cohorts of pigs marketed between July 2018 and June 2019. Following integration into a master-table, continuous explanatory variables were categorized into quartiles averages, and the W2F mortality was log-transformed, reporting geometric mean mortality of 8.69% for the study population. Further, univariate analyses were performed to identify individual variables associated with W2F mortality ($p < 0.10$) for further inclusion in a multivariable model, where model selection was applied. The final multivariable model consisted of 13 risk factors and accounted for 68.2% (R^2) of the variability of the W2F mortality, demonstrating that sow farm health and performance are closely linked to downstream W2F mortality.

Higher sow farm productivity was associated with lower subsequent W2F mortality and, conversely, lower sow farm productivity with higher W2F mortality e.g., groups weaned in the highest quartiles for pre-weaning mortality and abortion rate had 13.5%, and 12.5%, respectively, which was statistically lower than the lowest quartiles for the same variables (10.5%, and 10.6%). Moreover, better sow farm health status was also associated with lower

subsequent W2F mortality. A significant difference was detected in W2F mortality between epidemic versus negative groups for porcine reproductive and respiratory syndrome virus (15.4% vs 8.7%), and *Mycoplasma hyopneumoniae* epidemic versus negative groups (13.7% vs 9.9%).

Overall, this study demonstrated the application of a whole-herd analysis by aggregating information of the pre-weaning phase with the post-weaning phase (breeding-to-market) to identify and measure the major risk factors of W2F mortality.

Keywords: swine; wean-to-finish; mortality; whole-herd; risk factors.

Introduction

Wean-to-finish (W2F) mortality is commonly used as a key performance indicator of swine production systems (Agostini et al., 2014). As reviewed by Gebhardt et al. (2020), W2F mortality involves interactions among multiple factors related to animal physiology, management practices, and the pig's environment. This interaction is dynamic and fluid over time and geographical space. A further complexity, risk factors are often site-specific and vary markedly among production sites and within the same site over time. Thus, minimizing W2F mortality to optimize swine performance requires strategic management of multiple factors (Fablet et al., 2018).

Much of our understanding of the causality of post-weaning mortality is based on studies performed under experimental conditions, an environment that cannot represent the complexity and dynamic interactions among the risk factors that drive mortality in the field (Bello et al., 2018). The idea of researching the risk factors of post-weaning mortality under field conditions is not new. For example, Schwabe (1982) described the concept of "epidemiological diagnostic medicine", a whole-herd approach based on collecting and analyzing herd-specific data. Today, most of the data required to implement this approach are already collected by producers, but the day-to-day realities of production often compromise the full use of these data. For example, a cohort, i.e., a group of marketed pigs or "closeout", may have originated from multiple sow farms, each of different health status, or commingled (mixed with other growing groups) at some point during production. Further, even if the data was collected correctly, it may be stored in different locations and formats.

Despite the practical and technical challenges, achieving the integration of diverse and disperse data streams into a "master-table" of cohort-specific management, health, and environmental data provide the foundation for achieving Schwabe's whole-herd approach.

Therefore, the objective of this paper was to describe the process of integrating multiple data streams from one production system into a master-table, thereby creating a longitudinal integrated database consisting of retrospective data from breeding-to-market for each cohort. Furthermore, the objective was also to analyze the master-table to identify the major risk factors associated with W2F mortality in this company, with particular emphasis on the impact of sow farm performance metrics on the downstream mortality of the cohorts. Here we used the W2F mortality as the outcome. However, technically any other performance parameter present in the master-table can be used as the outcome.

Materials and Methods

2.1 Overview

This study used production and health data from 1,316 groups of pigs (cohorts) marketed between July 2018 and June 2019 by one swine production system (Iowa Select Farms, Iowa, USA). Cohort-specific longitudinal data (breeding-to-market) were aggregated into a single dataset from 6 separate data streams concerning multiple events related to the production process. The single master-table was then analyzed to identify the risk factors associated with pig mortality during the W2F phase.

2.2 Study Design

The observational unit was defined as a group (cohort) of pigs marketed from one finishing site at one point in time ($n = 1,316$). In the simplest case, a cohort originated from a single sow farm, was weaned into a single-stock wean-to-finish site, and remained there until marketed. The outcome was defined as the proportion of pigs in each cohort that died in the growing phase, here defined as W2F mortality. Within the production system, the pattern of pig flow varied to accommodate the available facilities, sow herd productivity, and management's assessment of weaned pig quality (Figure 1). Thus, a cohort may have originated from one or multiple sow farms and may have been weaned to nurseries or to single or double-stocked wean-to-finish sites. Double-stock wean-to-finish sites are stocked with the double number of animals of the final capacity of market age pigs. In contrast, single-stock wean-to-finish sites are stocked at weaning with the maximum limited capacity of market-age pigs.

2.3 Observational unit and characteristics of data streams

A cohort of marketed pigs was the observational unit for all analyses. For cohorts originating from multiple sow farms and/or multiple nurseries, a weighted fixed average of explanatory variables was calculated after marketing, based on the number of animals stocked and moved. The study was based on historical data for each cohort of pigs and the herd(s) from which they originated, i.e., all longitudinal data collected from breeding to market identified by cohort. Data provided by the production system in six separate Microsoft® Excel data streams (Table 1) were integrated into a single dataset, here referred as master-table, for analysis using SAS® Version 9.4 (SAS Institute, Inc., Cary, NC).

2.3.1 Productivity Data by Production Phase (Stream A)

Productivity data for the breeding-to-wean phase and growing phase were collected (Table 2) via password-protected online reports (MetaFarms, Inc., Burnsville, MN) concerning

the 1,316 cohorts. Breed-to-wean information was provided in sow farm performance monitor reports individually downloaded for each sow farm (41 reports); growing phase data were provided in nursery, wean-to-finish, and finisher closeout reports (3 reports). Sow farm-specific reports (July 2017 to January 2019) contained time-based weekly productivity information for the insemination, farrowing, and weaning events occurring in each week. For the variables concerning the farrowing event, we utilized the values of the previous 21 days of weaning in order to match, for each weaned group, their respective values for farrowing and weaning events. On the other hand, for variables concerning the insemination event, the values for each weaned cohort represented the performance in the sow farm at the time of weaning. Growing phase reports, i.e., the cohort-specific aggregated nursery and finisher reports, tracked productivity in the three pig flows illustrated in Figure 1, and information concerning the stocking conditions of the weaned cohorts.

2.3.2 Sow Farm Health Records Data (Stream B)

Sow farm health data consisted of retrospective weekly electronic records of clinical signs and laboratory testing performed in the herd for porcine reproductive and respiratory syndrome virus (PRRSV), porcine epidemic diarrhea virus (PEDV), Senecavirus A (SVA), and *Mycoplasma hyopneumoniae* (*Mhp*) (Table 3). Disease status classifications (negative, endemic, epidemic) in the datasets were utilized for PRRSV and *Mhp*, and were based on informed judgments (e.g., clinical signs, diagnostics, and changes in performance) by herd veterinarians.

Epidemic status (coded as 2) implied the first 16 weeks after an outbreak with the presence of clinical signs compatible with PRRSV and/or *Mhp*; endemic status (coded as 1) reflected a period from the week 17 after an outbreak until the herd is determined negative based on diagnostic testing; negative status (coded as 0) for PRRSV and *Mhp* indicated the absence of

nucleic acid detection by laboratory tests and the absence of clinical signs. For other pathogens, the disease classification criteria relied on classifying them as positive or negative according to clinical and diagnostic evidences of disease occurrence.

2.3.3 Sow Herd PRRSV Management Data (Stream C)

Sow herd PRRSV mass vaccination (whole-herd) dates were recorded for each sow farm and each weaned group identified in the context of week post-vaccination. Thus, groups weaned before sow herd mass vaccination were “0”, groups identified as “1” were weaned within 3 weeks of mass vaccination, groups “2”, “3”, and “4” were weaned ≥ 4 to 6, and ≥ 7 to 9, and 10 to 12 weeks post mass vaccination, respectively. The logic for comparing performance in different weeks before and after PRRS mass vaccinations was established based on the study conducted by Moura et al. (2019).

2.3.4 Pig Movement Data (Stream D)

Stream “D” recorded data on transportation events of weaned and growing animals from July 2017 to July 2019. Data included the respective number of animals moved, sites involved, date of transport, and the unique group alphanumeric identifier (letter to identify source plus numeric date/group identifier). The group alphanumeric identifier allows tracking cohorts from the pre-weaning to the post-weaning phase, even if multiple movements and days were required to assemble the group.

2.3.5 Miscellaneous data (Stream E)

The data stream “E” (Table 4) consisted of data on types of feeders, the season of the year when the weaned cohorts were stocked into the respective growing groups, number of sow farms originating each cohort, and air filtration (yes/no). Growing phase reports available before

and after the installation of filters on source sow farms made it possible to compare downstream productivity with and without air filtration for sow farms filtered during the study period.

2.3.6 Growing Pig Health Data (Stream F)

Growing pig health data included clinical assessments made by system personnel based on clinical observation and diagnostic test results. More specifically, groups were classified as positive or negative for specific infectious diseases (Table 5) based on the perceived presence or absence of clinical disease identified at any time throughout the growing phase. For example, cohorts with clinical signs for enteric coronaviruses would be tagged as groups with “break”. Cohorts that did not report the disease occurrence of a specific pathogen were considered as negative, albeit lack of clinical signs does not exclude the possibility of subclinical infections.

2.4 Data Aggregation

For each cohort, the 6 production data streams described in Section 2.2 were merged into a single dataset (master-table) in a 7-step process (Figure 2). Initially, SAS algorithms (PROC SET and PROC SQL) were used to create unique alphanumeric identifiers in each data stream, tagging every record with the location/date of occurrence and providing for complete traceability of cohort data from breeding to marketing. Thereafter, SAS algorithms were developed (PROC IMPORT, PROC DATA, PROC MERGE, PROC SET, PROC SQL, PROC SORT) to match and merge data streams into a single master-table based on the identifiers (Table 6). To assure data quality, responses inconsistent with specific variables were identified during the aggregation process and affected cohorts removed from the dataset (Table 7). After completing data aggregation, a data validation was performed by screening the dataset for irregularities related to specific productivity parameters (stream A) and values exceeding established ranges. Extreme values that represented actual events, e.g., high mortality resulting from infectious disease, were

retained as given. In the final step, the data originating from streams B - F were evaluated for completeness, where variables with $\geq 10\%$ of missing data were excluded from the final consolidated dataset.

2.5 Data analyses

Among the explanatory variables for which data were collected (streams “A” to “F”), only those with sufficient quality in the data validation and completeness steps were included in the final consolidated dataset (master-table). All continuous variables included in the master-table were categorized into quartiles, where the average value of the 25% of data inside each quartile was reported, while categorical variables were preserved in the original format.

For cohorts originating from multiple sow farms or weaned into multiple nurseries, a weighted fixed average of productivity parameters (stream A) was calculated for each cohort (PROC SQL) after the 6 data streams had been integrated. In other words, groups originated from multiple sow farms had the value of the stream A parameters calculated based on the average value of these variables in each sow farm, and on the number of incoming weaned pigs. For categorical variables from other data streams, the worse status was considered for the whole group, independently of the number of pigs, e.g., a cohort that receive 100 weaned piglets from an epidemic sow farm for PRRS and 900 pigs from a negative sow farm for PRRS would be considered as an epidemic group for PRRS.

Univariate analysis was conducted on the outcome (% of W2F mortality) to verify the assumption of normality (Shapiro-Wilk test), followed by its log-transformation to normalize the distribution. Thereafter, 27 univariate analyses (ANCOVA) were performed separately between each explanatory variable and the log-transformed W2F mortality (outcome), using a linear mixed model (PROC GLIMMIX in SAS) and a identity link function between the predictor and

the outcome with a normal distribution, and including the sow farm originating the cohorts as a random effect. The objective of this step was to identify the set of variables associated with W2F mortality., where only variables with p-values (p) < 0.10 in this step were eligible for inclusion in the initial multivariable model.

Before including all the significant variables from the univariate analyses in the multivariable model, Spearman's rank bivariate correlation coefficients were estimated among all the significant parameters (PROC CORR) to avoid multicollinearity in the multivariable model.

When the correlation between two variables was ≥ 0.50 , the variable occurring last in the pigs' life cycle was excluded. Thus, the multivariable regression analysis was conducted using a linear mixed model (PROC GLIMMIX), composed only by variables with significant association with the outcome in the previous univariate analysis ($p < 0.10$) and with low correlation (< 0.50) with other variables.

To build the multivariable model a manual stepwise selection was performed. Shortly, all variables with $p < 0.10$ in the univariate analysis were included in the multivariable model and the variables that had $p > 0.05$ were excluded one at a time until obtaining the final model with all variables having $p < 0.05$. Also, the Akaike's Information Criterion (AIC) value was utilized to compare the model goodness of fit when excluding the variables, by targeting the lowest value (Akaike, 1974). Furthermore, confounders and interactions were tested based on the biological plausibility of the variables based on the first author's (EM) perspective. Multicollinearity was tested in the in the final set of predictors included in the multivariable model by measuring the Variance Inflation Factor (VIF) for each parameter and excluding variables with $VIF > 5$. Also, all predictors not included in the initial multivariable model due to large p-value on the

univariate analyses were included again, one at a time, back into the final multivariable model, with the objective of testing if these variables remain not significant in the presence of potential confounders as mentioned by Dohoo et al. (2003), as seen that the effect of a predictor can become evident only when a confounder is controlled, which does not occur on the univariate analyses. On the multivariable model, the impact of each explanatory variable was assessed through pairwise comparisons (Tukey-Kramer test) by comparing the least square means estimates of the back-transformed log-W2F mortality between categories. Lastly, the Correlation Coefficient (R) and the coefficient of determination (R^2) were reported between the overall W2F mortality estimated by the final model and the actual value that occurred, describing the variance in W2F mortality explained by the model.

Results

3.1 Data Aggregation and Quality Assurance

The final master-table contained breeding-to-market data on 1,316 W2F cohorts (~4,200,000 pigs), originated from six data streams and from 41 sow farms, consolidating disperse information into a cogent "whole-herd" master-table suitable for statistical analyses. The data validation step resulted in the exclusion of 186 cohorts due to typographical errors in 2.43% (n = 32) and missing values in 11.7% (n = 154) and resulted in a final dataset containing 1,130 cohorts. Twenty-seven of 41 variables met the quality standard to be included in the analyses, excluding a total of 14 variables from streams "A", "E", and "F".

3.2 Univariate Analyses

The Shapiro-wilk test demonstrated that the outcome W2F mortality distribution was not normal, thus requiring its log-transformation for the univariate and multivariable analyses. The back-transformed geometric mean log-W2F mortality of the study population was 8.6% (95% confidence interval (CI) 8.4% - 8.9%).

The univariate analyses were conducted on the 27 eligible variables with the objective of identifying the variables with significant association with the outcome to be included in the multivariable model (Table 8), using the outcome on the log-scale and the sow farm originating the closeouts as the random effect, evaluating this way the effect of each risk-factor across the sow farms.

3.3 Multivariable Regression Analysis

From 27 variables analyzed in the univariate step, 5 variables were not included in the initial multivariable model due to large p-value ($p > 0.10$). These variables included capacity utilization, pigs weaned/mated female/year, herd parity, number of sources, and stocking type. Thereafter, correlation analyses were performed among the 22 remaining parameters identified as important to W2F mortality. Born alive, lactation length, pigs weaned/sow, non-productive days, and sow parity removal were excluded in the multivariable model due to high correlation with other variables. Overall, among the 27 variables in the univariate analysis, 17 met the eligibility criteria for inclusion in the initial multivariable regression analysis. Initial analysis of the multivariable model revealed that sow death rate, sow PRRS mass vaccination, litters/female/year (LFY), and SVA status had no significant association with downstream W2F mortality ($p > 0.05$). After excluding non-significant variables, the final multivariable model was composed of 13 explanatory variables and 6 interaction terms. The 5 variables initially excluded from the multivariable model, due to large p-value in the univariate step, were included again in

the model but remained not significant in the presence of other potential confounders ($p > 0.10$), thus confirming the low association with the outcome demonstrated on the univariate analyses. Also, no multicollinearity was observed in the final model ($VIF < 5$). The pairwise comparisons estimates are demonstrated in Table 8, and the total variance explained by the model in figure 3. All variables used in this study to obtain the final multivariable model and their respective p-value at each step are described in the Supplement File (Table A), as well as correlation coefficient when assessing pairwise correlations between predictors eligible for inclusion in the initial model (Table C), the AIC values between the multivariable models, and the Variance Inflation Factor (VIF) of each variable included in the final multivariable model when evaluating multicollinearity (Table B).

Overall, higher sow farm productivity was associated with lower subsequent W2F mortality and, conversely, lower sow farm productivity was associated with higher W2F mortality. Sow farm productivity variables in the final model included pre-weaning mortality, weaning age, service repeat rate, abortion rate, farrowing rate, average total born, birth loss, gestation length, and average sow parity at farrowing. Marketed cohorts weaned from the highest average quartile for pre-weaning mortality (18.0%) and abortion rate (5.3%) were associated with the highest W2F mortality category (13.5% and 12.5% respectively). In other words, marketed cohorts originated from sow farms weaning groups within the highest quartile average for pre-weaning mortality or abortion rate had the highest downstream W2F mortality. Furthermore, groups marketed within the youngest average age quartile for weaning age (15.3 days) had the highest W2F mortality (13.1%) compared to the other age categories. For the variable parity at farrowing, cohorts that were weaned in weeks where the average parity at farrowing represented the quartiles of 2.45 and 3.45 cycles, were associated with a higher W2F

mortality when compared to the oldest average parity category of 4.56 (12.1% and 11.9% vs 10.3%).

For service repeat rate and farrowing rate there was statistical difference among some of the categories within each variable, however not following the linear pattern observed in the previous variables that indicated a lower W2F mortality associated with improved performance. Furthermore, there was no statistical difference between the least squared means estimates of birth loss, gestation length, and total born categories.

As mentioned above for sow farm productivity parameters, better sow farm health status at weaning was also associated with lower W2F mortality. That is, groups weaned from sow farms with "epidemic PRRS" or "epidemic *Mhp*" recorded W2F mortalities of 15.4% and 13.7%, respectively. In contrast, no significant difference was detected in W2F mortality between PRRSV endemic vs negative sow farms (8.7% vs 10.7%) or *Mhp* endemic versus negative sow farms (9.9% vs 10.6%).

Weaned cohorts stocked into growing sites during winter months had higher downstream W2F downstream mortality when compared to groups stocked during the other seasons of the year (12.5% vs 11.3%, 11.0%, and 10.6%). Lastly, cohorts that at any point of time throughout the growing phase reported clinical evidence of enteric coronaviruses infection had higher W2F mortality when compared to groups without report of clinical disease (13.4% vs 9.6%).

The final multivariable model provided a correlation coefficient (R) of 0.826 between the predicted and observed W2F mortality ($p < 0.0001$), with an overall predicted W2F mortality of 9.06% versus the predicted W2F mortality mean of 9.47% (Figure 3). The final model explained 68.2% of the variability (R^2) in the W2F mortality for the 1,130 cohorts. The majority of variables included in the model represented the pre-weaning production phase, with the

exception of enteric coronaviruses, revealing that events occurring on the sow farm were greatly associated with the downstream W2F mortality.

Discussion

4.1 Background on data collection and analysis in the swine industry

Beginning in the mid-1960's, livestock production evolved from extensive production on small farms to intensive production on larger farms. With this change, livestock diseases changed from single etiologies to multifactorial "production diseases" which, according to Schwabe (1982), could only be understood through "epidemiological diagnostic medicine". Consistent with this reasoning, Fourichon (1991), ascribed a whole-herd approach in investigating disease dynamics, noting that the solution would require developing new methods for collecting qualitative and quantitative data on populations, infrastructure, feeding, disease occurrence, and personnel. Thus, to understand the relationship between livestock performance and the factors impacting it, e.g., mortality, a dynamic, interactive, and multifactorial investigation of risk factors/events occurring in the field was required (Fourichon, 1991; Schwabe, 1977). The implementation of this philosophical shift required better data collection methods and the emergence of animal health specialists with the mindset and analytical skills to perceive disease holistically (Schwabe, 1993).

Initial data recording systems in the swine industry focused exclusively on the breeding herd (Stein et al., 1987). They relied on hand-written records of sow and litter performance from which to produce simple descriptive reports. Practical computer technology and data recording systems appeared and proliferated over the last three decades, but in swine production, remained focused on breeding herds. This led to improvements in data collection, analysis, and benchmarking of key performance indicators related to mating, farrowing, and weaning events

(Piñeiro et al., 2019), e.g., pigs weaned per sow per year or litters per female per year, but left gaps in other segments of the production cycle. In the present study, for example, the largest portion of data is related to the pre-weaning phase (breeding herds). Regardless, the analysis demonstrated a strong link between sow herd health and productivity and downstream W2F mortality. At this point, and particularly as increasing data-capture technologies are used in commercial livestock production, producers have the opportunity to apply a similar approach to data-driven decision-making across the production cycle (Piñeiro et al., 2019).

4.2 Solving the problem of data collection/aggregation

To achieve a whole-herd approach to population health optimization, the data available for analysis must integrate all aspects of production. Banhazi and Black (2009) demonstrated that swine producers gathered data on multiple aspects of production, but data streams are typically collected and stored independently - and are thus beyond the easy reach of comprehensive analysis. The issue commonly confronted in the data integration process is incompatibility among data streams and software (Lenzerini, 2002). A number of studies have shown the feasibility of integrating multiple data streams for the purpose of identifying risk factors associated with performance in the growing phase (Agostini et al., 2015, 2014; Goumon and Faucitano, 2017; Larriestra et al., 2005a; Losinger et al., 1998; Oliveira et al., 2009; Passafaro et al., 2019) or the breeding phase (Liu et al., 2018; Pierozan et al., 2020). A common characteristic of these studies is the use of point-in-time, on-farm data collected and integrated using manual techniques and then evaluated using multivariable analyses.

In contrast, the current study achieved automated data integration in a commercial production system for ~4 million pigs in 1,130 cohorts across the entire production cycle (breeding-to-market). This approach allows automated on-going monitoring of the risk factors of

performance in swine populations. When pursuing a whole herd approach, the automated data integration process must be tailored to each production system because of the variety of methods used to manage data.

The use of SAS 9.4 in this study, a commercial data manipulation software, reduced data entry errors by providing criteria for the type and range of inputs but did not eliminate the need for a data quality step. Standards for data quality included identification of missing values and typographical errors. In this case, the immediate value of quality assurance procedure was to provide for reliable statistical analyses. Still, the long-term benefit is to assist the production system in achieving more consistent record collection over time with the goal of including more variables in future analyses and on-going development of the model.

In this study, whole-herd data integration was successful because it was possible to develop SAS algorithms to manage diverse and disperse records. This underscores the limitation that the data import process remains manual, and, as new data streams are introduced into the production environment, the algorithms used to manage data must likewise be updated.

4.3 Data Analyses

Larriestra et al. (2005a), Maes et al. (2004), and Oliveira et al. (2009), exploring the dynamics of variables associated with W2F mortality and their individual and collective impact, found large variation within and between herds in the effect of specific variables and showed the utility of multifactorial analyses in assessing these relationships. In the current study, identifying a diverse set of risk factors variables on W2F mortality was evaluated individually using univariate analysis, and their collective impact was investigated using multivariable regression analysis.

In the univariate and multivariable analyses, continuous variables were re-expressed in quartiles and the quartile means compared (ANCOVA). Compared to classical regression analysis, the conversion of continuous variables into categories resulted in some loss of information. Still, the loss was considered minimal due to the large sample size and the use of multiple (4) categories per variable. Overall, categorization facilitated the analysis and interpretation of results, as noted by Altman (2005). On the other hand, as stated by Bennette and Vickers (2012), the downside of categorization involves multiple hypothesis testing between the quartiles (pairwise comparisons), and the fact that assumes homogeneity within the groups, which does not always occur.

In the univariate analyses, ANCOVA was used to identify variables statistically associated with the W2F mortality, revealing that various sow farm health and reproductive performance parameters were associated with W2F mortality. Results of the univariate analysis (explanatory variable and outcome) should be interpreted with caution because variables identified as significant did not operate in isolation. For example, variables that had a statistically significant association with the outcome in the univariate analysis became non-significant in the initial multivariable analysis (e.g., sow herd PRRS vaccination, sow death rate, and LFY), when the model accounted for all the variables. Thus, the primary function of the univariate step was to identify variables associated with W2F mortality and decide for inclusion in the multivariable regression model.

Among the 27 variables analyzed in the bivariate step, 17 were included in the initial multivariable analysis ($p < 0.10$). Previous authors have included variables with p -values < 0.10 (Pieroza et al., 2020), < 0.20 (Maes et al., 2004), and < 0.25 (Agostini et al., 2015, 2014). As discussed by Dohoo et al. (1997), the rationale for including variables with p -values of 0.10 to

0.20 is that only variables with a significant unconditional association with the outcome are included in the final model. This could lead to the exclusion of some valid predictor variables from the final analysis if their effect were masked by another variable, i.e., if the effect of a variable in the univariate analysis becomes significant only when a confounder is controlled.

Given the absence of a definitive consensus in the literature concerning the inclusion of variables in multivariable models, we utilized a rigorous cut-off p-value of 0.10 due to the fact that the model accounted for a higher number of predictor variables and the dataset was extensive, thus increasing the probability of type I errors as the number of observations increase (Santman-Berends et al., 2021). Furthermore, multicollinearity among variables was investigated (correlation > 0.50) by using an approach proposed by Cohen (1988), similarly to that described by Maes et al. (2004), and using VIF as proposed by (Pak et al., 2002)

The majority of the variables included in the final multivariable model were related to the pre-weaning phase. It is likely that the small number of growing phase variables identified for inclusion reflects less availability of data from this phase of production when compared to breeding herds, not its potential impact on W2F mortality, as seen that most of the growing phase variables collected were not included in the whole analysis due to data quality matters. Greater effort to consistently collect growing phase variables for inclusion in the analyses could help to further explain variation in W2F mortality. This is an example of future improvements to the whole-herd approach in this system.

The final multivariable model provided a correlation coefficient (R) of 0.826, explaining 68.2% of the variability (R^2) in the W2F mortality of the 1,130 cohorts. In contrast, previous models based on growing pig stocking conditions variables explained 20.4% (Agostini et al., 2014), 12.06% (Larriestra et al., 2005a), and 20.0% (Maes et al., 2004) of W2F mortality. This

fact demonstrates the great portion of mortality explained by sow farm risk factors and the importance of the quality of the weaned group on their downstream performance.

Based on these results, sow farm reproductive performance variables may be viewed as a proxy of population health. That is, better breeding herd performance suggests that the population is managed under conditions that allow animals to achieve their full genetic potential. This hypothesis is supported by the fact that all breeding herds participating in this study had the same genetic breeding stock make-up. Thus, differences in productivity between sow herds were likely not due to their genetic potential.

Notwithstanding, the results were consistent with previous field studies reporting associations between sow farm productivity indicator variables and W2F mortality, e.g., weaning age (Davis et al., 2006; Leliveld et al., 2013; López-Vergé et al., 2019; Main et al., 2004), weaning weight (Collins et al., 2017; de Grau et al., 2005; Declerck et al., 2016; Dewey et al., 2006; Larriestra et al., 2005b), and birth weight (Declerck et al., 2016; Pardo et al., 2013; Škorput et al., 2018; Smith et al., 2007; Zotti et al., 2017)

For seasonality, the association between groups stocked in colder months and higher W2F mortality was also in agreement with previous reports (Agostini et al., 2014; Oliveira et al., 2009).

Likewise, the link between W2F performance and the sow farm status for PRRSV, *Mhp*, and PEDV is straightforward, i.e., these pathogens are known to affect growing pig health and productivity (Madson et al., 2014; Pieters and Maes, 2019; Zimmerman et al., 2019). Further, the results from the present study are consistent with previous reports of downstream mortality associated with PRRSV (Alvarez et al., 2015a; Fablet et al., 2018; Holtkamp et al., 2013), *Mhp* (Silva et al., 2019), and PEDV (Alvarez et al., 2015b; Yamane et al., 2016). It is also important

to note that the possibility of subclinical infections in cohorts that did not report epidemic status should not be excluded; thus, the non-significant difference between endemic and negative groups should be interpreted only for W2F mortality, and not for other productivity parameters such as average weight daily gain.

There is variation across disease classification systems utilized in the swine industry, and therefore the results related to the disease status should be interpreted carefully. Furthermore, causal inference models were not conducted. Instead, the analytical models focused on screening risk factors associated with swine wean-to-finish mortality. Therefore, extrapolating the effects of the differences between the categories for prospective groups is an inaccurate approach. On the other hand, the results were sufficient to document a strong association between sow farm health and productivity and subsequent W2F mortality.

Lastly, the result from the high variability explained by the existing multivariable model (68.2%) should not be extrapolated to prospective groups in terms of prediction, as seen that the model utilized in this study was developed mainly with the objective of analyzing retrospective data.

Conclusions

In swine production systems, consolidation and analysis of diverse and disperse data are the backbones of data-driven decision-making. This study demonstrated this by automating the process of integrating whole-herd data related to wean-to-finish mortality, utilizing a commercial statistical software. Also, the multiscale data consolidation supported precision swine health and production management strategies in this production system by providing the capability of

collectively analyze diverse data that was previously dispersed, under their field conditions, enabling a whole-herd approach on investigating swine survivability.

The multivariable model explained 68.2% of the downstream W2F mortality, demonstrating a strong association between sow farm health and productivity parameters on downstream pig health. Under the conditions of this study the major risk factors of W2F mortality were high pre-weaning mortality, high abortion rate during the week of weaning for growing pig cohorts, weaned groups originated from sow farms with PRRSV and *Mhp* epidemic status, marketed cohorts that were weaned during the winter season, and closeouts having a clinical occurrence of enteric coronaviruses throughout the growing phase.

The interactions between health, productivity, and environmental factors change dramatically across production systems and time. This study demonstrates the concept of building a health and production intelligence system that can serve as the foundation for ongoing data analysis in swine production systems, with the ability to include new data streams on risk factors and interventions over time, and the potential to serve as the foundation for future development of causality and forecasting models.

Acknowledgements

The study is part of the Improving Pig Survivability project, which was funded by *The National Pork Board* and *The Foundation for Food and Agriculture Research* grant #18-147.

Conflict of interest

The authors declare that there is no conflict of interests.

References

- Agostini, P.D.S., Manzanilla, E.G., De Blas, C., Fahey, A.G., Da Silva, C.A., Gasa, J., 2015. Managing variability in decision making in swine growing-finishing units. *Ir. Vet. J.* 68. <https://doi.org/10.1186/s13620-015-0048-z>
- Agostini, P.S., Fahey, A.G., Manzanilla, E.G., O'Doherty, J. V., De Blas, C., Gasa, J., 2014. Management factors affecting mortality, feed intake and feed conversion ratio of grow-finishing pigs. *Animal* 8, 1312–1318. <https://doi.org/10.1017/S1751731113001912>
- Akaike, H., 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* 19, 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Altman, D., 2005. Categorizing continuous variables. *Can. J. Psychiatry* 47, 886. <https://doi.org/10.1177/070674370204700916>
- Alvarez, J., Sarradell, J., Kerkaert, B., Bandyopadhyay, D., Torremorell, M., Morrison, R., Perez, A., 2015a. Association of the presence of influenza A virus and porcine reproductive and respiratory syndrome virus in sow farms with post-weaning mortality. *Prev. Vet. Med.* 121, 240–245. <https://doi.org/10.1016/j.prevetmed.2015.07.003>
- Alvarez, J., Sarradell, J., Morrison, R., Perez, A., 2015b. Impact of Porcine Epidemic Diarrhea on Performance of Growing Pigs. *PLoS One* 10, e0120532. <https://doi.org/10.1371/journal.pone.0120532>
- Banhazi, T.M., Black, J.L., 2009. Precision Livestock Farming: A Suite of Electronic Systems to Ensure the Application of Best Practice Management on Livestock Farms. *Aust. J. Multi-Disciplinary Eng.* 7, 1–14. <https://doi.org/10.1080/14488388.2009.11464794>
- Bello, N.M., Ferreira, V.C., Gianola, D., Rosa, G.J.M., 2018. Conceptual framework for

- investigating causal effects from observational data in livestock. *J. Anim. Sci.* 96, 4045–4062. <https://doi.org/10.1093/jas/sky277>
- Bennette, C., Vickers, A., 2012. Against quantiles: Categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med. Res. Methodol.* 12. <https://doi.org/10.1186/1471-2288-12-21>
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences* Second Edition.
- Collins, C.L., Pluske, J.R., Morrison, R.S., McDonald, T.N., Smits, R.J., Henman, D.J., Stensland, I., Dunshea, F.R., 2017. Post-weaning and whole-of-life performance of pigs is determined by live weight at weaning and the complexity of the diet fed after weaning. *Anim. Nutr.* 3, 372–379. <https://doi.org/10.1016/j.aninu.2017.01.001>
- Davis, M.E., Sears, S.C., Apple, J.K., Maxwell, C. V., Johnson, Z.B., 2006. Effect of weaning age and commingling after the nursery phase of pigs in a wean-to-finish facility on growth, and humoral and behavioral indicators of well-being. *J. Anim. Sci.* 84, 743–756. <https://doi.org/10.2527/2006.843743x>
- de Grau, A., Dewey, C., Friendship, R., de Lange, K., 2005. Observational study of factors associated with nursery pig performance. *Can. J. Vet. Res.* 69, 241–245.
- Declerck, I., Dewulf, J., Sarrazin, S., Maes, D., 2016. Long-term effects of colostrum intake in piglet mortality and performance1. *J. Anim. Sci.* 94, 1633–1643. <https://doi.org/10.2527/jas.2015-9564>
- Dewey, C.E., Johnston, W.T., Gould, L., Whiting, T.L., 2006. Postweaning mortality in Manitoba swine. *Can. J. Vet. Res.* 70, 161–167.
- Dohoo, I.R., Ducrot, C., Fourichon, C., Donald, A., Hurnik, D., 1997. An overview of techniques

- for dealing with large numbers of independent variables in epidemiologic studies. *Prev. Vet. Med.* 29, 221–239. [https://doi.org/10.1016/S0167-5877\(96\)01074-4](https://doi.org/10.1016/S0167-5877(96)01074-4)
- Dohoo, I.R., Martin, W., Stryhn, H.E., 2003. *Veterinary Epidemiologic Research*. Charlottetown, P.E.I.
- Fablet, C., Rose, N., Grasland, B., Robert, N., Lewandowski, E., Gosselin, M., 2018. Factors associated with the growing-finishing performances of swine herds: An exploratory study on serological and herd level indicators. *Porc. Heal. Manag.* 4, 1–11. <https://doi.org/10.1186/s40813-018-0082-9>
- Fourichon, C., 1991. Application of ecopathological methods to the investigation of health problems on farms. *Rev. Sci. Tech.* 10, 151–177. <https://doi.org/10.20506/rst.10.1.542>
- Gebhardt, J.T., Tokach, M.D., Dritz, S.S., DeRouche, J.M., Woodworth, J.C., Goodband, R.D., Henry, S.C., 2020. Postweaning mortality in commercial swine production. I: Review of non-infectious contributing factors. *Transl. Anim. Sci.* 4, 462–484. <https://doi.org/10.1093/TAS/TXAA068>
- Goumon, S., Faucitano, L., 2017. Influence of loading handling and facilities on the subsequent response to pre-slaughter stress in pigs. *Livest. Sci.* 200, 6–13. <https://doi.org/10.1016/j.livsci.2017.03.021>
- Holtkamp, D.J., Kliebenstein, J.B., Neumann, E.J., Zimmerman, J.J., Rotto, H.F., Yoder, T.K., Wang, C., Yeske, P.E., Mowrer, C.L., Haley, C.A., 2013. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *J. Swine Heal. Prod.* 21, 72–84.
- Larriestra, A.J., Maes, D.G., Deen, J., Morrison, R.B., 2005a. Mixed models applied to the study

of variation of grower-finisher mortality and culling rates of a large swine production system. *Can. J. Vet. Res.* 69, 26–31.

Larriestra, A.J., Morrison, R.B., Deen, J., Abvp, D., 2005b. Peer reviewed Production tool A decision-making framework for evaluating interventions used at weaning to reduce mortality in lightweight pigs and. *J. Swine Heal. Prod.* 13, 143–149.

Leliveld, L.M.C., Riemensperger, A. V., Gardiner, G.E., O’Doherty, J. V., Lynch, P.B., Lawlor, P.G., 2013. Effect of weaning age and postweaning feeding programme on the growth performance of pigs to 10 weeks of age. *Livest. Sci.* 157, 225–233.
<https://doi.org/10.1016/j.livsci.2013.06.030>

Lenzerini, M., 2002. Data integration: A theoretical perspective. *Proc. ACM SIGACT-SIGMOD-SIGART Symp. Princ. Database Syst.* 233–246.

Liu, Z.X., Wei, H.K., Zhou, Y.F., Peng, J., 2018. Multi-level mixed models for evaluating factors affecting the mortality and weaning weight of piglets in large-scale commercial farms in central China. *Anim. Sci. J.* 89, 760–769. <https://doi.org/10.1111/asj.12963>

López-Vergé, S., Gasa, J., Coma, J., Bonet, J., Solà-Oriol, D., 2019. Effect of lactation length caused by the management production system on piglet performance until slaughter. *Livest. Sci.* 224, 26–30. <https://doi.org/10.1016/j.livsci.2019.04.003>

Losinger, W.C., Bush, E.J., Smith, M.A., Corso, B.A., 1998. An analysis of mortality in the grower/finisher phase of swine production in the United States. *Prev. Vet. Med.* 33, 121–145. [https://doi.org/10.1016/S0167-5877\(97\)00052-4](https://doi.org/10.1016/S0167-5877(97)00052-4)

Madson, D.M., Magstadt, D.R., Arruda, P.H.E., Hoang, H., Sun, D., Bower, L.P., Bhandari, M., Burrough, E.R., Gauger, P.C., Pillatzki, A.E., Stevenson, G.W., Wilberts, B.L., Brodie, J.,

- Harmon, K.M., Wang, C., Main, R.G., Zhang, J., Yoon, K.J., 2014. Pathogenesis of porcine epidemic diarrhea virus isolate (US/Iowa/18984/2013) in 3-week-old weaned pigs. *Vet. Microbiol.* 174, 60–68. <https://doi.org/10.1016/j.vetmic.2014.09.002>
- Maes, D.G.D., Duchateau, L., Larriestra, A., Deen, J., Morrison, R.B., De Kruif, A., 2004. Risk Factors for Mortality in Grow-finishing Pigs in Belgium. *J. Vet. Med. Ser. B* 51, 321–326. <https://doi.org/10.1111/j.1439-0450.2004.00780.x>
- Main, R.G., Dritz, S.S., Tokach, M.D., Goodband, R.D., Nelssen, J.L., 2004. Increasing weaning age improves pig performance in a multisite production system. *J. Anim. Sci.* 82, 1499–1507. <https://doi.org/10.2527/2004.8251499x>
- Moura, C.A.A.A., Johnson, C., Baker, S.R., Holtkamp, D.J., Wang, C., Linhares, D.C.L.L., 2019. Assessment of immediate production impact following attenuated PRRS type 2 virus vaccination in swine breeding herds. *Porc. Heal. Manag.* 5, 2–7. <https://doi.org/10.1186/s40813-019-0120-2>
- Oliveira, J., Yus, E., Guitián, F.J., 2009. Effects of management, environmental and temporal factors on mortality and feed consumption in integrated swine fattening farms. *Livest. Sci.* 123, 221–229. <https://doi.org/10.1016/j.livsci.2008.11.016>
- Pak, S. Il, Spahr, U., Jemmi, T., Salman, M.D., 2002. Risk factors for *L. monocytogenes* contamination of dairy products in Switzerland, 1990–1999. *Prev. Vet. Med.* 53, 55–65. [https://doi.org/10.1016/S0167-5877\(01\)00274-4](https://doi.org/10.1016/S0167-5877(01)00274-4)
- Pardo, C.E., Kreuzer, M., Bee, G., 2013. Effect of average litter weight in pigs on growth performance, carcass characteristics and meat quality of the offspring as depending on birth weight. *Animal* 7, 1884–1892. <https://doi.org/10.1017/S1751731113001419>

Passafaro, T.L., Van De Stroet, D., Bello, N.M., Williams, N.H., Rosa, G.J.M., 2019.

Generalized additive mixed model on the analysis of total transport losses of market-weight pigs. *J. Anim. Sci.* 97, 2025–2034. <https://doi.org/10.1093/jas/skz087>

Pierozan, C.R., Callegari, M.A., Dias, C.P., De Souza, K.L., Gasa, J., Da Silva, C.A., 2020.

Herd-level factors associated with piglet weight at weaning, kilograms of piglets weaned per sow per year and sow feed conversion. *Animal* 14, 1283–1292.

<https://doi.org/10.1017/S175173111900346X>

Pieters, M.G., Maes, D., 2019. Mycoplasmosis, in: *Diseases of Swine*. Wiley, pp. 863–883.

<https://doi.org/10.1002/9781119350927.ch56>

Piñeiro, C., Morales, J., Rodríguez, M., Aparicio, M., Manzanilla, E.G., Koketsu, Y., 2019. Big

(pig) data and the internet of the swine things: A new paradigm in the industry. *Anim.*

Front. 9, 6–15. <https://doi.org/10.1093/af/vfz002>

Santman-Berends, I.M.G.A., Nijhoving, G.H., van Wuijckhuise, L., Muskens, J., Bos, I., van

Schaik, G., 2021. Evaluation of the association between the introduction of data-driven tools to support calf rearing and reduced calf mortality in dairy herds in the Netherlands.

Prev. Vet. Med. 191, 105344. <https://doi.org/10.1016/J.PREVETMED.2021.105344>

Schwabe, C.W., 1993. The current epidemiological revolution in veterinary medicine. Part II.

Prev. Vet. Med. 18, 3–16. [https://doi.org/10.1016/0167-5877\(93\)90088-B](https://doi.org/10.1016/0167-5877(93)90088-B)

Silva, G.S., Yeske, P., Morrison, R.B., Linhares, D.C.L., 2019. Benefit-cost analysis to estimate

the payback time and the economic value of two *Mycoplasma hyopneumoniae* elimination methods in breeding herds. *Prev. Vet. Med.* 168, 95–102.

Škorput, D., Dujmović, Z., Karolyi, D., Luković, Z., 2018. Variability of birth weight and

growth of piglets in highly prolific sows. *J. Cent. Eur. Agric.* 19, 823–828.

<https://doi.org/10.5513/JCEA01/19.4.2355>

Smith, A.L., Stalder, K.J., Timo, ;, Serenius, V., Baas, T.J., Mabry, J.W., 2007. Effect of piglet birth weight on weights at weaning and 42 days post weaning, *Journal of Swine Health and Production*.

Stein, T., Martineau, G.P., Morris, R., Charette, R., 1987. A new approach to managing health in Swine operations. *Can. Vet. J. = La Rev. Vet. Can.* 28, 355–35562.

Yamane, I., Yamazaki, H., Ishizeki, S., Watanabe, Y., Okumura, H., Okubo, M., Kure, K., Hayakawa, Y., Furukawa, M., Ooi, M., Mizukami, Y., Ito, M., 2016. Impact of a porcine epidemic diarrhea outbreak on swine productivity in Japan: A retrospective cohort study. *J. Vet. Med. Sci.* 78, 1385–1389. <https://doi.org/10.1292/jvms.15-0723>

Zimmerman, J.J., Dee, S.A., Holtkamp, D.J., Murtaugh, M.P., Stadejek, T., Stevenson, G.W., Torremorell, M., Yang, H., Zhang, J., 2019. Porcine Reproductive and Respiratory Syndrome Viruses (Porcine Arteriviruses), in: *Diseases of Swine*. Wiley, pp. 685–708. <https://doi.org/10.1002/9781119350927.ch41>

Zotti, E., Resmini, F.A., Schutz, L.G., Volz, N., Milani, R.P., Bridi, A.M., Alfieri, A.A., da Silva, C.A., 2017. Impact of piglet birthweight and sow parity on mortality rates, growth performance, and carcass traits in pigs. *Rev. Bras. Zootec.* 46, 856–862.

<https://doi.org/10.1590/S1806-92902017001100004>