Pig Organ Lesions Recorded in Different Abattoirs: A Statistical Approach to Assess the Comparability of Prevalence

Ariane Horst [1],*, Marvin Gertz [1], Mario Hasler [2] and Joachim Krieter [1]

[1] Institute of Animal Breeding and Husbandry, Kiel University, 24118 Kiel, Germany; mgertz@tierzucht.uni-kiel.de (M.G.); jkrieter@tierzucht.uni-kiel.de (J.K.)

[2] Division of Variation Statistics, Kiel University, 24118 Kiel, Germany; hasler@email.uni-kiel.de

* Correspondence: ahorst@tierzucht.uni-kiel.de; Tel.: +49-431-8801-493

check for updates

Abstract: Documented lesions of slaughtered pigs provide a high-density data-pool that could be valuable for the purpose of animal health monitoring and breeding. However, data quality and structure hamper the application of statistical methods. The present study provides an approach that enables statistical analysis and evaluates the comparability of lesion prevalence among abattoirs. The German Quality and Safety database provided data of recorded lung, pleura, liver, and heart lesions. Filter criteria were used to improve the data structure. Data of $n = 8,004,769$ animals, recorded in nine abattoirs over a period of 18 months, were analyzed. Lesion prevalences were successfully modeled by applying a generalized linear mixed model. To examine prevalence differences, the coefficient of variation (CV) on a six-monthly basis was calculated, and a grand mean test (GMT) of significance was applied. High variations in estimated prevalence occurred on abattoir, six-monthly and organ basis. The highest variation occurred in the lung (CV = 64.7%), whereas liver lesions showed the lowest variation (CV = 21.8%). The GMT enabled the visualization of these variations between abattoirs, organs and over time. Concerning the assessment of the comparability of prevalences, it provides a promising tool to monitor changes in lesion examination and to address divergent abattoirs.

Keywords: abattoir data; meat inspection; lesion recording; data structure; prevalence; pig health

1. Introduction

Animal health is being increasingly focused on consumers and politics as part of the animal welfare discussion. Moreover, it is an important aspect of the pig production process, as affected animals can cause high economic losses [1–3]. Despite this high ethical and economic importance, a standardized surveillance system to monitor pig health is currently not feasible at national or international level. The documentation of animal health data is often characterized by high workload and costs when conducted on a large number of animals [4]. Hence, a nationally data pool comprising animal health data that was recorded under standardized circumstances is missing. In this context, lesions recorded during meat inspection are a useful adjunct to provide information on pig health [5], since the documentation is a regular part of food safety regulation in Europe. This high-density data-pool could be potentially valuable to enable automated monitoring of on-farm health, and even be valuable for breeding purposes [6], as the validity of recorded lesions to monitor on-farm health have been examined and confirmed in literature [7,8]. Nevertheless, challenges concerning data quality and data flow exist, thus the data of recorded lesions is currently used to a small extent only. For example, in Germany, feedback of recorded lesions is given to the farmer in addition to the abattoir settlement [9],

thus farmers and veterinarians can use this information to assess on-farm pig health. However, if a batch of pigs is sent to several abattoirs, contrary information about the prevalence-status within this batch can be received [10], thereby hampering the overall acceptance of such a health-monitoring approach as the reliability of lesion recordings is reduced. This is because several factors alter the lesion scoring outcomes, such as the effect of the meat inspector, abattoir, season, on-farm health status and housing conditions, which is elaborated and discussed in more detail in other studies [1,6]. Concerning the assessment and interpretation of lesion scoring outcomes reported back to the farmer, some studies currently recommend comparing only lesions that are recorded at the same abattoir in order to minimize these effects [3,10].

However, the aggregation of recorded lesions from different abattoirs is indispensable when establishing a valid national surveillance system to monitor pig health or for possible applications in pig-breeding. A standardized procedure (e.g., a "gold standard") that enables the assessment and comparison of lesion recording among abattoirs under practical conditions is currently not available, meaning that a comparative assessment of abattoir data is challenging. This issue has already been the focus of some research [11–13]. As a consequence, present studies examining these data are forced to use only a strongly restricted data-set of recorded lesions in order to enable statistical analysis. Therefore, it is common in the literature to use data originating from only one abattoir [14–16] so no inter-abattoir variation exists that needs to be considered. Some studies used meat inspection data originating from different abattoirs [11–13,17,18]. However, either the examination has been performed by only one or two observers [12,18] or the study aim focuses on batch/farm differences [13,17]. Only one of the listed studies considered inter-abattoir differences in lesion examination, i.e., scale effects, by applying a z-transformation on the data [11].

The present study provides an approach that analyzes recorded lesions from various abattoirs and therefore examines abattoir related differences. By using a mixed model approach for a post-hoc correction of abattoir-specific effects, subsequent statistical analysis of the data is enabled. In addition, the comparability of lesion prevalence of different abattoirs will be discussed regarding their contribution to monitoring pig health on-farm and in breeding.

2. Materials and Methods

2.1. Abattoir Data

Anonymized lesion recordings of single animal data were provided by the German inspection system for food products (QS, Qualität und Sicherheit GmbH). The study period comprised 18 months, from January 2017 to June 2018. The dataset was subdivided into three six-monthly periods (S1, S2, S3) as all subsequent analyses were performed on a six-monthly basis following the salmonellae monitoring and therapy index of QS. The raw data comprised 99 abattoirs, 37,398 supplying farms and $n = 65{,}637{,}324$ observations.

Out of all lesions currently recorded according to the German administrative regulation of food hygiene [9], this study focuses on recordings solely from lung, pleura, liver and heart disorders. In Germany, these lesions have been recorded since the early 1990s [19], and in addition are mainly focused on in the literature as they have an ecological and animal health-related impact on the farm [11,14,15,20]. Since the renewal for the German meat inspection regulation in the midyear of 2016, heart and liver lesions are documented by using a binary code (0 = negative, 1 = positive), whereas lung and pleura lesions are usually recorded as discrete, polytomous variables (−1 = negative, 0 = marginal, 1 = moderate, 2 = severe). In some abattoirs, however, the code −1 for lung and pleura lesions is not documented but instead the code 0 is sometimes used simultaneously for "no lesions", as it was common before the renewal. To enable a uniform examination on a binomial scale, the data of lung and pleura lesions were merged to a binary code. Negative (−1) and marginal (0) lesions are now aggregated to negative (0) and moderate (1) and severe (2) lesions to positive (1). The following prevalences were calculated based on the number of positive lesion recordings in the data: $\text{Lung}_{S1} = 10.6\%$,

lung$_{S2}$ = 12.9%, lung$_{S3}$ = 13.9%, pleura$_{S1}$ = 7.74%, pleura$_{S2}$ = 8.37%, pleura$_{S3}$ = 8.84%, liver$_{S1}$ = 8.71%, liver$_{S2}$ = 10.9%, liver$_{S3}$ = 8.78%, heart$_{S1}$ = 4.51%, heart$_{S2}$ = 4.78%, and heart$_{S3}$ = 4.88%.

2.2. Preliminary Examination of the Data Structure

To investigate the structure of the raw data, a network analysis was performed using the Python 3.4.9 module networkx 1.11. The data structure in this context refers to the number of farms that deliver animals regularly to more than one abattoir and therefore enables a statistical connection of abattoirs. When creating a projected and undirected monopartite network, the abattoirs (knots) are connected by farms (edges). To assess the overall connectedness of a network, parameter density (D) can be used, as this parameter represents the number of existing connections out of all possible connections of the network. If all possible connections are formed, the density is D = 1. In addition, fragmentation (F) provides information about isolated network components. If all knots are connected and present in one single network, i.e., one coherent network component, fragmentation is F = 0. Furthermore, the average edges per knots can be calculated which is known as average degree (AD). Focusing on the edge weight (i.e., number of farms) between knots, the average effective degree (AE) gives information about the average effective, valid connection of knots of the network.

2.3. Improving Data Quality and Structure

When analyzing recorded lesions under current practical conditions, a sometimes low reliability hampers the application of statistical methods. However, applying a data filter enhances the data structure and thus might also improve data quality. Therefore, abattoirs were excluded from the dataset if they showed missing recordings (i.e., no documented lesions) during three consecutive weeks of the study period. This step excluded 62.6% (*n* = 62) of the abattoirs from the dataset. Some abattoirs, in addition, exhibited an "overall prevalence per abattoir" of exactly 0% or >80% in the considered organs. These cases were excluded since the values were assumed to be implausible.

In a second step, the data structure was improved at the farm-level. The following conditions were identified as filter thresholds: A delivery per farm to at least two different abattoirs in order to increase the connection between abattoirs and farms. Furthermore, a minimal sample size of 40 animals for each supplying farm during a six-month period was considered. This threshold was found in a preliminary investigation (study in submission), assessing the agreement of a farm's prevalence, depending on different sample sizes over time.

The filtered data-set comprised nine abattoirs (A1–A9), 5321 farms and *n* = 8,004,768 observations (minimal abattoir sample size n_{min} = 50,141, maximal abattoir sample size n_{max} = 1,775,753). The filtered data was used for the subsequent statistical analysis, as it represented an appropriate, connected data structure, which was necessary in the case of potential statistical issues based on the insufficient data quality.

2.4. Statistical Model

Due to the binomial nature of the data, a generalized linear mixed model was applied using R package lme4 [21] with logit as the link function.

During model development, fixed and random effects were added to the model in a stepwise manner, and the Akaike's information criteria (AIC) as well as the Bayesian information criteria (BIC) were used as references to compare the different models. The final model for recorded lesions included the abattoir (A1–A9) as a fixed effect, and slaughter day, nested in abattoir, and farm as random effects.

2.5. Statistical Analysis

The least square means (LSM) and random estimates of slaughter day, nested in abattoir, were extracted from the model to enable an overview of the absolute intra-abattoir differences as well as variation over time.

To enable for a visual inspection of mixed model results, the random estimates of only two abattoirs with comparable sample sizes were illustrated in the following, serving as examples (other results included in the Supplementary Materials). Random estimates for lung and heart prevalence, therefore, were selected, as these organs were assumed to exhibit different results that serve for a better visualization of the issue. Supported by the application of the penalized basis spline (degree = 3, PBSPLINE statement of PROC SGPLOT in SAS® 9.4), serving as moving average, the extent of the slaughter day effect was illustrated.

Subsequently, intra-and inter-abattoir variation were analyzed in more detail: The LSM of the model provide estimated prevalence on a six-monthly basis for each organ and abattoir. The average mean of the abattoirs LSM form the grand mean for each six-monthly period of the data-set. In addition, the LSM of the abattoirs were used to calculate the coefficient of variation of prevalence within organ and six-monthly period to assess the variation of examinations of the considered organs.

Finally, a grand mean test (GMT) of significance was applied, to assess which abattoir's mean differs significantly from the grand mean of the data. The grand mean was defined as the weighted mean of the estimated prevalence of all considered abattoirs. As no "gold standard" for evaluation was available, the grand mean was considered as a statistically grounded reference value for the "baseline" prevalence. Therefore, a grand mean for each six-monthly period was applied.

Data preparation and statistical analysis were applied with SAS® 9.4 and the statistic programming language R 3.6.0. Since these values were based on logit scale, random estimates, LSM as well as the results of the GMT of significance have been retransformed to prevalence scale to enhance the comprehensibility of the results.

3. Results

3.1. Network Analysis of the Data Structure

Based on the abattoirs of the raw data-set, the following network parameters were calculated: D = 0.31, F = 0.02, AD = 30.5 and AE = 9.59. One abattoir can be found isolated from the coherent network; the other abattoirs are connected by edges to at least two other abattoirs. According to the density, the connection between abattoirs is moderate, as only 44% of the abattoirs exhibit greater than or equal to 33 edges (which represents a third of all possible connections), and only 15% of the abattoirs are connected to greater than or equal to half of the other abattoirs in the data-set.

A repetition of the network analysis with the filtered data set produced the following network parameters: D = 0.94, F = 0, AD = 7.56 and AE = 4.22. In contrast to the raw data, the filtered data showed one coherent network component of abattoirs with no isolated abattoirs (F = 0). The majority of all possible connections among abattoirs were constructed according to the calculated density. Considering that the filtered data-set comprised fewer abattoirs than the raw data, a decreased average degree as well as a decreased average effective degree were calculated. The network of the abattoirs (squares 1–9) is visualized in Figure 1. The edges represent the number of farms (*n* = 3243) connecting the abattoirs. Thereby, six edges (thick line) containing a minimum of 513 and a maximum of 1020 farms, 10 edges (solid line) are formed by a minimum of 167 and a maximum of 324 farms whereas 18 edges (dashed line) are connecting the abattoirs by a minimum of one farm and a maximum of 99 farms.
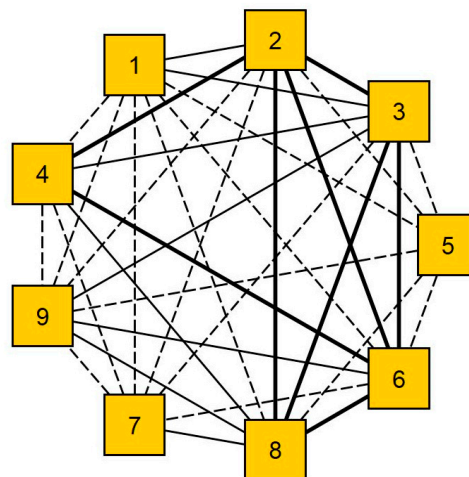
Figure 1. A network of the filtered data set comprising nine abattoirs (squares 1–9). The edges represent the number of linking farms. Thick line: 500–1500 farms, solid line: 100–499 farms, dashed line: 1–99 farms.

3.2. Intra-Abattoir Variation

The extent of daily fluctuations during the study period is indicated exemplarily for two abattoirs (A2 and A4) and two lesions in Figure 2 (further results can be found in Supplementary Figure S1). Variation in the lung and heart prevalence is illustrated according to the effect of slaughter day, nested in abattoir. A4 exhibits a larger extent of variation between the minimum and maximum random estimates than abattoir A2. The effect of slaughter day variation in random estimates is higher in the lung than in heart prevalence. Furthermore, the moving average of lung prevalence exhibits fluctuations in the course of the study period whereas the average heart prevalence shows an approximately steady course.
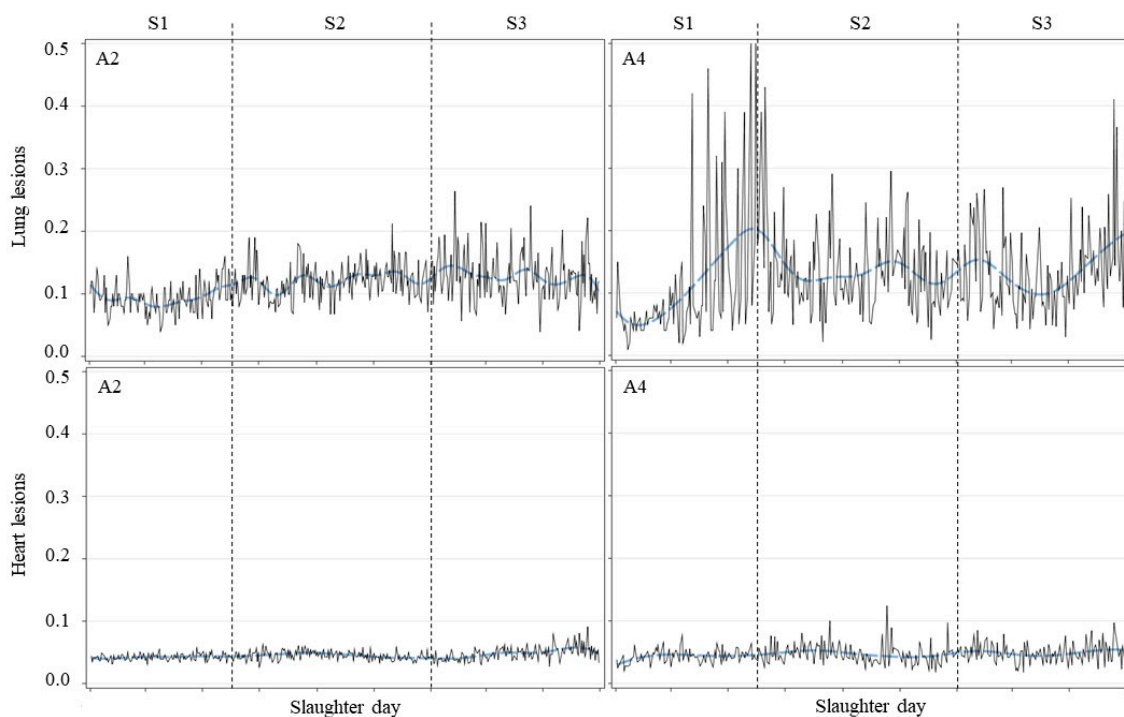
Figure 2. Retransformed random estimates (%) of slaughter day (solid line), nested in abattoir, in the course of six-monthly periods (S1–S3) and the moving average (dashed line); exemplarily illustrated for documented lung and heart lesions of two selected abattoirs (A2 and A4).

3.3. Inter-Abattoir Variation

Figure 3 illustrates the retransformed LSM for each abattoir, organ, and six-monthly period. It is noticeable that some abattoirs (e.g., A2, A4, A5) show strongly fluctuating LSM of lung prevalence in the course of the study period. Especially in S2 and S3, these abattoirs exhibit a strongly increased LSM in comparison to S1. Abattoir A5 also shows an increase in liver, pleura, and heart prevalence in S3. The maximum difference between the lung prevalence of two abattoirs is given in A4 and A9 with 21.8 percentage points (S2), whereas heart prevalence shows a maximum difference between A5 and A9 of 4.60 percentage points (S3). Pleura and liver prevalence show maximum differences in S1 and S3 with 5.81 and 6.82 percentage points.
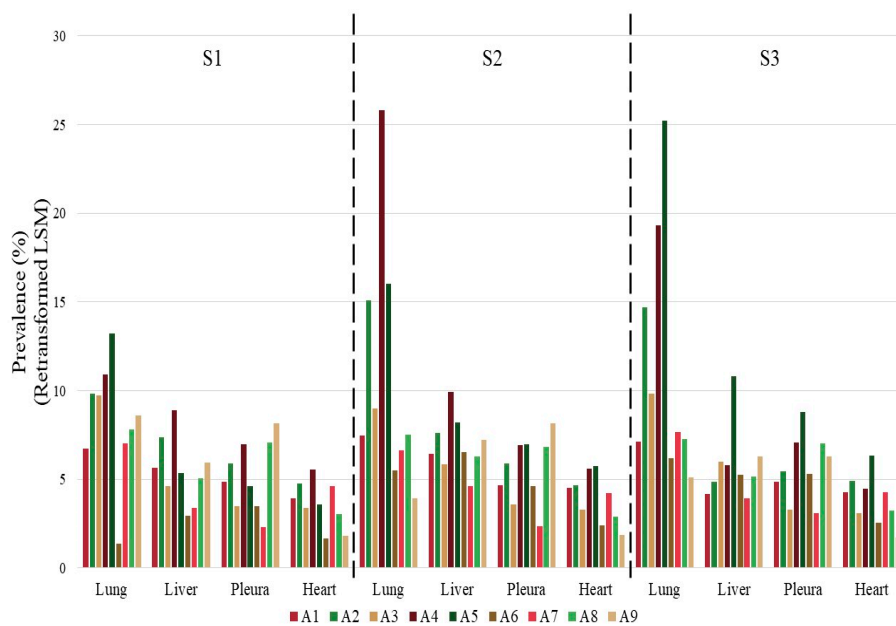

Figure 3. Prevalence (%) of each abattoir (A1–A9) and organ for the respective six-monthly period (S1–S3), based on the retransformed least square means (LSM) of the model.

Concerning the coefficient of variation (CV) to assess the variation of examinations of the considered organs, the following CVs were calculated for S1: $CV_{lung} = 39.6\%$, $CV_{heart} = 36.5\%$, $CV_{pleura} = 37.6\%$, $CV_{liver} = 33.7\%$; S2: $CV_{lung} = 64.7\%$, $CV_{heart} = 35.3\%$, $CV_{pleura} = 34.0\%$, $CV_{liver} = 21.8\%$; S3: $CV_{lung} = 60.8\%$, $CV_{heart} = 35.2\%$, $CV_{pleura} = 32.5\%$, $CV_{liver} = 34.8\%$. The highest CV occurs in lung prevalence in S2, whereas liver lesions show the lowest variation in S3. Variation of heart and pleura prevalence can be ranged in between. The coefficient of variation fluctuates during study period: Except for lung and liver prevalence, heart and pleura prevalence show a decreasing variation.

3.4. Grand Mean Test of Significance

Using the LSM from the model, the deviation of each abattoir's mean (including their 90% confidence interval) to the grand mean of the respective six-monthly period is illustrated in Figure 4. An abattoir can be assumed to be not significantly different from the grand mean if its mean (dot) and/or its 90% confidence interval (bracket) touches the grand mean (dashed line). The GMT of significance reflects the variations shown and discussed based on the LSM of the model. Generally, the abattoirs' means of liver, pleura, and heart prevalences vary less around the grand mean than the means of lung prevalence. The deviations of each abattoirs' mean of lung prevalence vary up to 12.0 percentage points (e.g., A4, S2), whereas deviations of heart prevalence differ lower than 2.47 percentage points (e.g., A5, S3) from the grand mean. Deviations of pleura and liver prevalence can be ranged in between. Only one observation (A1, S1) can be found in lung prevalence that does not significantly deviate

from the grand mean. In pleura prevalence, seven of these kinds of observations occur whereas six observations can be found in heart prevalence and five observations can be found in liver prevalence that do not significantly deviate from the grand mean.
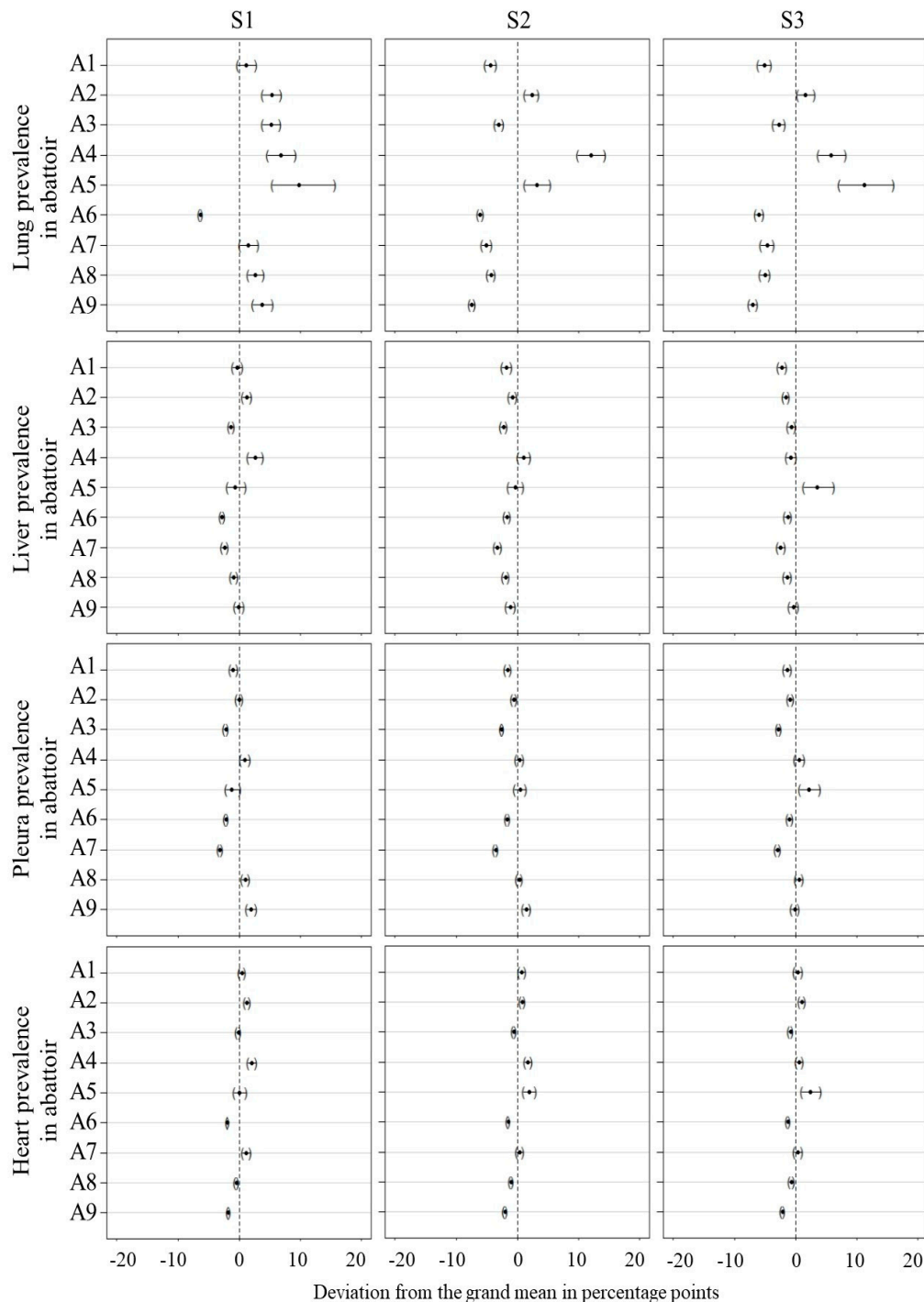
Figure 4. Deviations in lung, liver, pleura, and heart prevalence of abattoirs mean compared to the grand mean of the respective six-monthly period (S1–S3). Dot = abattoir deviation; bracket = 90% confidence interval; dashed line = grand mean. Calculated grand means (GM) based on the LSM of each abattoir in S1: GM_{lung} = 8.35%, GM_{heart} = 3.59%, GM_{pleura} = 5.19%, GM_{liver} = 5.46%; S2: GM_{lung} = 10.8%, GM_{heart} = 3.91%, GM_{pleura} = 5.55%, GM_{liver} = 6.97%; S3: GM_{lung} = 11.4%, GM_{heart} = 3.88%, GM_{pleura} = 5.68%, GM_{liver} = 5.80%).

Each abattoir has to be assumed as significantly different from the grand mean at least once during the study period. Nevertheless, three abattoirs can be found that do not differ from the grand mean in two consecutive six-monthly periods, which are A5 in liver and pleura prevalence in S1 and S2, as well as A4 in pleura prevalence and A7 in heart prevalence in S2 and S3. Focusing on the total observations per abattoir that show no significant difference from the grand mean, the maximum number of these kinds of observations is given for A5 ($n = 5$) whereas A6 shows the minimal number of zero observations. An average number of $n = 2.11$ non-significant observations per abattoir have been calculated.

Comparing the abattoir deviations from the grand mean of lung prevalence with the other deviation results, a conspicuous difference is, first, that in S1 every abattoir except for A6 shows positive deviations from the grand mean. Second, in S2 and S3, all abattoirs except for A2, A4, and A5 show negative deviations from the grand mean. Therefore, A4 and A5 exhibit the highest deviation.

4. Discussion

4.1. Data Quality and Structure

The number of abattoirs excluded from the study was high when focusing on a continuous lesion recording. This can be explained by the circumstance of the cessation of business in some cases during the study period. Another reason is given by small abattoirs not slaughtering daily. If these abattoirs also slaughter other species, it is reasonable that a continuous recording of solely pig lesions is missing.

The partially inaccurate data quality and data structure of recorded lesions is known and has often been discussed in the literature [1,6,15], and have been confirmed by the applied network analysis of the raw data. To allow an assessment of the usability of recorded lesions, an adjustment of the data is necessary to enable statistical analysis. Data filtering and filter criteria based on abattoirs and the farm-level have been applied in other studies too [13,17]. However, since there is no standardized procedure for how to determine filter thresholds, the specification depends on the researcher's decision. Nonetheless, for the present study a general comparison of the overall organ prevalence of the filtered data (lung = 10.6%, pleura = 7.79%, liver = 9.92%, heart = 4.31%) and the raw data set (lung = 9.56%, pleura = 6.83%, liver = 10.5%, heart = 3.93%) can be used to ensure that the data reduction occurred more or less randomly.

Similar to this study and in order to achieve acceptable data quality, Meyns et al. (2011) [17] used data from the ten largest abattoirs in the country. Hulsegge and Greef (2028) [13], in addition, selected two large abattoirs with complete datasets (without missing data points) and a steady inspection system during the study period. However, applying a data filter results in a conflict between data accuracy and data reduction [13]. For example, while this study excluded farms with <40 animals per six-monthly period, Meyns et al. (2011) [17] determined a minimum threshold of 80 animals per farm during a four months' study period. The determination of filter criteria therefore regulates the number of abattoirs and farms being part of the data-set, thus the data structure. Regarding the usage of a surveillance program to monitor on-farm pig health, small farms and abattoirs will be neglected if the criteria applied are too strictly. The rejection of small farms, however, is not the purpose of a surveillance program, as it should concern every farm to be accepted nationally. The applied filter criteria in this study enabled a clear estimation of abattoir, slaughter day and farm effect, while in turn containing only a small data volume of abattoirs and farms. Thus, further studies should examine to what extent the data structure deteriorates in the case of including small farms in the dataset. For this purpose, potential changes in the estimates for the abattoir and farm effect could be used to enable a realistic compromise between data accuracy and data structure.

4.2. Data Modeling

Following the application of filter thresholds, recorded lesion disorders were successfully modeled by applying a generalized linear mixed model. Preliminary examination indicated that the available

raw data is exposed irregularities concerning distribution and heteroscedasticity. When considering farm and slaughter day in the model, data gaps occurred on certain factor-levels. These are obtained by some factor-levels exposing extreme-value observations with solely zeros or ones, resulting in heteroscedasticity among the abattoirs. The applied data filter successfully decreased these data-gaps.

4.3. Intra-Abattoir Variation

High variations in estimated prevalence occurred at abattoir-, six-monthly period- and organ-levels. Abattoir-related differences can be classified into intra- and inter-abattoir variation. Intra-abattoir variation is shown in the random estimates of slaughter day. Steinmann et al. (2014) [22] also found large deviations in a daily assessment period of two weeks in a German abattoir. Meat inspector (e.g., motivation to work, individual sensitivity of judging) and farm of origin (e.g., on-farm management, housing conditions) are known factors causing this kind of variation and are listed and discussed more precisely by Horst et al. (2019) [6]. Since the data of the present study are provided in anonymized form, neither any information about the meat inspector performing the examination nor on-farm conditions are available. Considering slaughter day and fattening farm as random effects to the model currently represents the next possible level to account for these effects.

Depending on abattoir and organ prevalence, some abattoirs show an intra-abattoir variation in prevalence in the course of the study period (see penalized basis spline in Figure 2 and Figure S1 (supplementary material)). Especially noticeable is the variation in the lung estimates of slaughter day (e.g., A1, A4, A6 and A9). Since no information about modifications in the abattoir's process design was known, and no respiratory epidemic occurred during the study period, one explanation could be changed circumstances concerning the lesion assessment (e.g., meat inspector training and/or internal adjustments of the assessment procedure). This can be confirmed by Enøe et al. (2003) [12] who focused on differences in prevalence over time and indicated that variations do result from changes in the sensitivity and specificity of lesion assessment rather than changes in the true prevalence.

4.4. Inter-Abattoir Variation

Inter-abattoir differences occur when comparing the LSM of the model for each abattoir and organ grouped into the respective six-monthly period (see Figure 3). A large proportion of these variations, besides the subjective recording, mainly arise from the individualized process designs of the abattoirs (e.g., line speed, light, work organization) [23,24].

Noticeable are the lung LSM for abattoir A6 in S1 as well as A4 and A5 in S2 and S3. These values can be recognized as outliers, as they show considerable deviations from the overall mean and/or differ noticeably from the abattoir's prevalence estimated for the previous/subsequent six-monthly period. The plausibility of these prevalences, regarding a valid lesion examination, should be considered on a low level.

In this study, the coefficient of variation was used to assess the variation of prevalence compared to its mean for the respective organ disorder. As expected, lung prevalence showed the highest coefficient of variation among the abattoirs (average CV_{lung} = 55.0%), which can be confirmed by Hoischnen-Taubner et al. (2011) [25], Schleicher et al. (2013) [14], and Eckhardt et al. (2009) [26]. The lowest coefficient of variation was calculated for liver prevalence (average CV_{liver} = 30.1%). A low variation, i.e., a high agreement, for liver lesions was calculated by Schleicher et al. (2013) [14] and Hoischen-Taubner et al. (2011) [25], too. In contrast, heart prevalence is known to show only a very small amount of meat inspector variation [14]. Thus, this organ was assumed to exhibit just a low variation relative to its mean; however, contrary to this assumption, heart prevalence exhibits a higher variation (average CV_{heart} = 35.7%) than liver and pleura prevalence (average CV_{pleura} = 34.7%). Similar results can be found in Teixeira et al. (2016) [27], who found a higher standard deviation in pericarditis prevalence compared to its mean, based on batch level.

4.5. Grand Mean as Reference Value

The current data quality and data structure are not sufficient to enable the definition of a gold standard based on abattoir lesion data, which has been expounded in the previous sections. Organ lesions are recorded under practical conditions, consequently it is not possible to generate repeated measures which could serve as comparative values. Furthermore, reference results by other studies are currently missing that could be used as a gold standard [14,20]. Even though the determination of a gold standard is theoretically possible, the justification of such a value is not simple. The emphasis thereby is, to identify a justified institution (e.g., veterinary inspection office, agricultural communities of interest, department of agriculture) determining an "optimal prevalence value/range" or "gold standard abattoir". However, each institution represents different positions and opinions, which lead to plausible but different decisions. Moreover, the determination must be made on the basis of well-founded arguments/scientific results and accepted by the parties involved (which in turn produces a vicious circle). Nonetheless, a reference value is indispensable in terms of statistical evaluation. Regarding the current data condition, the grand mean of the data seems to be the most fairly and statistically justified reference value, as no arbitrary determination is necessary. Furthermore, the grand mean comprises data from all considered abattoirs, so no abattoir is assumed to be the gold standard or will be disadvantaged.

4.6. Applicability of the Grand Mean Test

Regarding the quality of lesion recording, the results of the GMT indicate that particularly the examination and documentation of lung disorders needs improvement, which can be confirmed by other studies, too [14,25].

The test was capable of visualizing inter-abattoir variations as well as changes in organ examination during the study period on abattoir-level. Hence, it could be used to assess the (positive) effect of regular training of meat inspectors at national level. Assuming that national, standardized training results in less inter-abattoir variation, a change in the GMT of significance is visible in smaller deviations of the abattoirs means from the grand mean in the subsequent six-monthly period.

Assuming that the variability in estimated prevalence might be affected by changes in the sensitivity and specificity of the examination [12], the GMT provides an opportunity to visualize and monitor alterations due to meat inspector training in practice. Furthermore, this test could be used to identify abattoirs showing conspicuous deviations in organ prevalences. Lesion recording in these abattoirs should be considered for improvement, and the strongly divergent prevalence based on this lesion recording must be recognized as inappropriate to monitor animal health. In these cases, for example, the veterinary inspection office and/or the QS system can directly address and interact with the abattoir in order to improve the organ lesion assessment.

When considering conspicuous abattoirs nevertheless in statistical methods (e.g., the GMT), their data condition will bias the output depending on their sample size. To generate a purposive data base with improved data quality, these noticeable abattoirs should be excluded in the context of statistical analysis. One example is exemplarily illustrated in Supplementary Figure S2 for lung lesions. Abattoir A6 in S1 was excluded as it was considered as an outlier. Resulting from this step, a higher grand mean for lung prevalence in S1 was calculated and decreased deviations of the remaining abattoir means from the grand mean occur (results are shown in Figure S2 (supplementary material)). Another example is given in S3 when excluding abattoir A5 as it shows noticeable deviations from the other abattoirs. When excluding this abattoir from the dataset and applying the GMT for all focused organs, lower grand means were calculated (results are shown in Supplementary Figure S3). Similar to the previous example, the deviations from the grand mean for the remaining abattoirs decrease. Hence, it may be concluded that the data quality of recorded lesions is improvable by excluding and addressing strongly divergent abattoirs until the quality management of their recording is checked and adjusted if necessary.

5. Conclusions

When analyzing lesion data recorded under the current practical conditions, disadvantaged data quality and structure hamper the application of statistical methods. Due to the improvement of the data structure by applying filter criteria, variation-causing effects were modeled, and the prevalences of lesion disorders were estimated successfully. However, data filtering is accompanied by data loss, which is especially detrimental in terms of animal health monitoring. Therefore, further research is necessary to investigate an appropriate balance between data accuracy and data volume.

The GMT of significance is an expedient method to visualize the variation between abattoirs but also the different extents of variation in organ prevalence. While lung prevalence exhibits high variation, the presented results indicate liver, pleura, and heart prevalence to be suitable for the purpose of health monitoring or breeding. Furthermore, the test provides a promising tool to identify conspicuous abattoirs and to monitor changes due to regular meat inspector training at national level.

References

1. Harley, S.; More, S.; Boyle, L.; Connell, N.O.; Hanlon, A. Good animal welfare makes economic sense: Potential of pig abattoir meat inspection as a welfare surveillance tool. *Ir. Vet. J.* **2012**, *65*, 11. [CrossRef] [PubMed]

2. Heinonen, M.; Bergman, P.; Fredriksson-Ahomaa, M.; Virtala, A.-M.; Munsterhjelm, C.; Valros, A.; Oliviero, C.; Peltoniemi, O.; Hälli, O. Sow mortality is associated with meat inspection findings. *Livest. Sci.* **2018**, *208*, 90–95. [CrossRef]

3. Vial, F. Editorial: Slaughterhouses as Sources of Data for Animal Health Intelligence. *Front. Vet. Sci.* **2019**, *5*, 332. [CrossRef] [PubMed]

4. Merks, J.W.M.; Mathur, P.K.; Knol, E.F. New phenotypes for new breeding goals in pigs. *Animal* **2012**, *6*, 535–543. [CrossRef]

5. EFSA Panel on Biological Hazards (BIOHAZ), EFSA Panel on Contaminants in the Food Chain (CONTAM) and EFSA Panel on Animal Health and Welfare (AHAW). Scientific Opinion on the public health hazards to be covered by inspection of meat (swine). *EFSA J.* **2011**, *9*, 1–198.

6. Horst, A.; Gertz, M.; Krieter, J. Challenges and opportunities of using meat inspection data to improve pig health traits by breeding: A review. *Livest. Sci.* **2019**, *221*, 155–162. [CrossRef]

7. Dickhaus, C.P.; Meemken, D.; Blaha, T. Attempts to Quantify the Health Status of Pig Herds: Developing and validating a Herd Health Score (HHS). In *Sustainable Animal Production: The challenges and Potential Developments for Professional Farming*; Aland, A., Ed.; Wageningen Acad. Publ.: Wageningen, The Netherlands, 2009; pp. 191–201.

8. van Staaveren, N.; Doyle, B.; Manzanilla, E.G.; Calderón Díaz, J.A.; Hanlon, A.; Boyle, L.A. Validation of carcass lesions as indicators for on-farm health and welfare of pigs. *J. Anim. Sci.* **2017**, *95*, 1528–1536. [CrossRef]

9. Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz. *Allgemeine Verwaltungsvorschrift über die Durchführung der amtlichen Überwachung der Einhaltung von Hygienevorschriften für Lebensmittel und zum Verfahren zur Prüfung von Leitlinien für eine gute Verfahrenspraxis*; AVV LmH: Berlin, Germany, 2009.

10. Blaha, T. Stand der Umsetzung des Lebensmittelrechts in der Schweineproduktion. In *4. Leipziger Tierärztekongress, Schwerpunkt Schwein*; Aschenbach, J.R., Gäbel, G., Daugschies, A.H., Eds.; Leipziger Blaue Hefte: Leipzig, Germany, 2008; pp. 549–553.

11. Nienhaus, F.; Meemken, D.; Schoneberg, C.; Hartmann, M.; Kornhoff, T.; May, T.; Heß, S.; Kreienbrock, L.; Wendt, A. Health scores for farmed animals: Screening pig health with register data from public and private databases. *PLoS ONE* **2020**, *15*, e0228497. [CrossRef]

12. Enøe, C.; Christensen, G.; Andersen, S.; Willeberg, P. The need for built-in validation of surveillance data so that changes in diagnostic performance of post-mortem meat inspection can be detected. *Prev. Vet. Med.* **2003**, *57*, 117–125.

13. Hulsegge, B.; Greef de, K.H. A time-series approach for clustering farms based on slaughterhouse health aberration data. *Prev. vet. med.* **2018**, *153*, 64–70. [CrossRef]

14. Alban, L.; Petersen, J.V.; Busch, M.E. A comparison between lesions found during meat inspection of finishing pigs raised under organic/free-range conditions and conventional, indoor conditions. *Porcine Health Manag.* **2015**, *1*. [CrossRef] [PubMed]

15. Schleicher, C.; Scheriau, S.; Kopacka, I.; Wanda, S.; Hofrichter, J.; Köfer, J. Analysis of the variation in meat inspection of pigs using variance partitioning. *Prev. Vet. Med.* **2013**, *111*, 278–285. [CrossRef] [PubMed]

16. Scollo, A.; Gottardo, F.; Contiero, B.; Mazzoni, C.; Leneveu, P.; Edwards, S.A. Benchmarking of pluck lesions at slaughter as a health monitoring tool for pigs slaughtered at 170kg (heavy pigs). *Prev. Vet. Med.* **2017**, *144*, 20–28. [CrossRef] [PubMed]

17. Meyns, T.; van Steelant, J.; Rolly, E.; Dewulf, J.; Haesebrouck, F.; Maes, D. A cross-sectional study of risk factors associated with pulmonary lesions in pigs at slaughter. *Vet. J.* **2011**, *187*, 388–392. [CrossRef]

18. Merialdi, G.; Dottori, M.; Bonilauri, P.; Luppi, A.; Gozio, S.; Pozzi, P.; Spaggiari, B.; Martelli, P. Survey of pleuritis and pulmonary lesions in pigs at abattoir with a focus on the extent of the condition and herd risk factors. *Vet. J.* **2012**, *193*, 234–239. [CrossRef]

19. Blaha, T. Erfassung von Schlachttierbefunden für die Qualitätssicherung bei der Fleischproduktion. *Deutsche tierärztliche Wochenschrift* **1994**, *101*, 264–267.

20. Bonde, M.; Toft, N.; Thomsen, P.T.; Sørensen, J.T. Evaluation of sensitivity and specificity of routine meat inspection of Danish slaughter pigs using Latent Class Analysis. *Prev. Vet. Med.* **2010**, *94*, 165–169. [CrossRef]

21. Bates, D.; Maechler, M.; Bolker, B.; Walker, S.; Christensen, R.H.B.; Singmann, H.; Dai, B.; Scheipl, F.; Grothendieck, G.; Green, P.; et al. Package 'lme4': Linear Mixed-Effects Models using 'Eigen' and S4. Available online: https://cran.r-project.org/web/packages/lme4/lme4.pdf (accessed on 7 April 2020).

22. Steinmann, T.; Blaha, T.; Meemken, D. A simplified evaluation system of surface-related lung lesions of pigs for official meat inspection under industrial slaughter conditions in Germany. *BMC Vet. Res.* **2014**, *10*, 98. [CrossRef]

23. Correia-Gomes, C.; Smith, R.P.; Eze, J.I.; Henry, M.K.; Gunn, G.J.; Williamson, S.; Tongue, S.C. Pig Abattoir Inspection Data: Can It Be Used for Surveillance Purposes? *PLoS ONE* **2016**, *11*, e0161990. [CrossRef]

24. Stärk, K.D.C.; Alonso, S.; Dadios, N.; Dupuy, C.; Ellerbroek, L.; Georgiev, M.; Hardstaff, J.; Huneau-Salaün, A.; Laugier, C.; Mateus, A.; et al. Strengths and weaknesses of meat inspection as a contribution to animal health and welfare surveillance. *Food Control* **2014**, *39*, 154–162. [CrossRef]

25. Hoischen-Taubner, S.; Blaha, T.; Werner, C.; Sundrum, A. Zur Reproduzierbarkeit der Befunderfassung am Schlachthof für Merkmale der Tiergesundheit. Repeatability of anatomical-pathological findings at the abattoir for characteristics of animal health. *J. Food Saf. Food Qual.* **2011**, *62*, 82–87.

26. Eckhardt, P.; Fuchs, K.; Kornberger, B.; Köfer, J. Untersuchungen über die Raliabilität der im Zuge der Fleischuntersuchung erhobenen Befunde von Schlachtschweinen. Wiener Tierärztliche Monatsschrift. Available online: https://center.ssi.at/smart_users/uni/user94/explorer/43/WTM/Archiv/2009/WTM_05-06-2009_Artikel_6.pdf (accessed on 1 November 2017).

27. Teixeira, D.L.; Harley, S.; Hanlon, A.; O'Connell, N.E.; More, S.J.; Manzanilla, E.G.; Boyle, L.A. Study on the Association between Tail Lesion Score, Cold Carcass Weight, and Viscera Condemnations in Slaughter Pigs. *Front. Vet. Sci.* **2016**, *3*, 24. [CrossRef] [PubMed]