BIOSECURITY MEASURES TO REDUCE HEPATITIS E VIRUS TRANSMISSION WITHIN PIG FARMS

Marina Meester
Student MSc Epidemiology Postgraduate – student nr. 4069439

LAY SUMMARY

Hepatitis E virus (HEV) can infect humans and cause inflammation of the liver. For people living in industrialized countries, pigs are the main source of human infection. Pigs that become infected do not fall ill, but they do shed HEV in their feces, and the virus can reside in their liver. Therefore, transmission of HEV from pigs to humans may occur via contact with pig manure or consumption of raw or undercooked pork liver.

HEV is present in most pig farms, and it is currently not known how to stop infection in pigs, to indirectly protect humans against HEV. Thus, we aimed to find out what measures pig farmers can take, to prevent pigs on their farm from becoming infected.

In the present study, 73 farms of two classes were compared; 1: farms on which every group of pigs becomes infected with HEV, 2: farms in which some groups of pigs remain free from HEV. Measures that are taken more often in one of the two farm classes, may protect against HEV infection, or the other way around, may increase the chance of HEV infection. Farms were classified based on blood samples collected from several groups of pigs from those farms at slaughter. The measures taken by farmers were identified by conducting a questionnaire and an audit on the farms. For both the questionnaire and audit, emphasis was put on measures regarding internal biosecurity, which entails all things a farmer can do to stop spread of a virus within the farm, like cleaning an empty barn, rodent control and changing or cleaning boots between groups of pigs.

In the statistical analysis, the answers to all questions and all checkpoints in the audit were taken into account and a model was used to distinguish the measures that had the strongest relationship with having a group of pigs that remained free from HEV.

We have found that the floor material present in barns with fattening pigs have the strongest relationship with an HEV free group of pigs. Namely, rubber flooring increased the odds of having an HEV free group by 6 fold, and steel by 7 fold. We hypothesize that rubber and steel can be cleaned more effectively than the common floor material concrete, so that rubber and steel floors indirectly prevent new pigs in the barn from becoming infected. Furthermore, cleaning driving boards, that are used to move pigs from one barn to another, protects against HEV infection of pigs. Possibly, the driving boards are a source of transmission between pigs. Also, controlling flies is a protective measure, especially when done by releasing predatory flies that eat the eggs of barn flies. A farm characteristic that may increase the chance of HEV infection is a long duration of the fattening period, which may be the result of pigs becoming older on those farms and thereby increasing the time in which they are at risk of becoming infected.

In conclusion, several measures related to cleaning and cleanability of pig farms can be taken, as well as improving control of flies, to prevent groups of pigs from becoming infected by hepatitis E virus.

ABSTRACT

Hepatitis E virus (HEV) can be transmitted from pigs and pork to humans and cause liver inflammation. Many pig farms are HEV positive and deliver almost every group (batch) of pigs HEV positive to slaughter, indicating efficient within-farm transmission of HEV. Reducing the number of HEV positive pigs at slaughter is necessary to lower human exposure, yet how to do so is unknown. Case-control studies can give insight in measures that reduce within-farm transmission of HEV and thereby reduce HEV positive slaughter pigs. We conducted a case-control study with 73 pig farms, that were selected from a previous study, based on serological (ELISA) and virological (PCR) results of multiple slaughter batches. Case farms had at least one HEV PCR and ELISA negative (PCR^-ELISA^-) batch, indicating the ability to prevent within-farm transmission, and control farms had the highest proportion of PCR^+ELISA^+ batches, indicating high within-farm transmission. Farm biosecurity measures were identified via a questionnaire and an audit. The outcome was the ratio of PCR^-ELISA^- to positive batches per farm and 35 farms had between 0.1 and 0.6 PCR^-ELISA^- batches. Variable selection was performed by bootstrapped grouped logistic Least Absolute Shrinkage and Selection Operator (LASSO) regression to account for multicollinearity and overfitting the model to the data, and to assess stability of the selected variables. By logistic regression, the final odds ratios (OR) with 95% confidence intervals (95%CI) of the selected variables were determined. Rubber and steel floor material in fattening pens increased the odds of a PCR^-ELISA^- batch (OR 5.87 (95%CI 3.03-11.6 and 7.13 (95%CI 3.05-16.9) respectively). Cleaning driving boards at least once a week (OR 1.99 (95%CI 1.07-3.80)), and fly control with predatory flies (OR 4.52 (95%CI 1.59-13.5)) were protective factors, whereas a fattening period longer than 105 days was a risk factor. No external biosecurity measures were significantly associated with HEV slaughter batch negativity. This study shows that HEV control within pig farms should focus on cleaning and cleanability of floors and fomites and adequate fly control.

Keywords: HEV; zoonosis; foodborne; within-farm transmission; case-control study; mitigation

INTRODUCTION

Yearly, an estimated 20 million hepatitis E virus (HEV) infections occur worldwide, leading to an estimated 3.3 million symptomatic cases of hepatitis E, an inflammation of the liver [1]. HEV is a single-stranded quasi-enveloped RNA virus of which at least eight genotypes have been discovered [2]. HEV genotypes 1 and 2 have been found solely in humans and predominantly in low- and middle-income countries where limited sanitary conditions and access to clean drinking water can lead to large waterborne HEV outbreaks [1]. HEV genotypes 3 and 4 have been found in humans and animals, including pigs, wild boar, deer and a growing record of other animal species (e.g. [3-5]). In Europe (EU), HEV genotype 3 is most common and HEV seroprevalence in blood donors from EU countries ranges from 1.1% (Spain) to 52% (France) [6]. HEV infection is mostly asymptomatic and self-limiting, however humans may acquire symptoms of acute liver inflammation, acute liver failure or chronic liver cirrhosis [7] and extrahepatic manifestations like Guillain-Barre syndrome [8]. It appears that foodborne transmission, by raw or undercooked pork meat or pork liver sausages, is the major pathway for HEV genotype 3 infections in humans [9].

The seroprevalence of HEV in pigs is reported for many EU countries and ranges from 8 to 93% in individual pigs [10]. Pigs shed HEV in feces and possibly urine [11], and HEV may persist in the liver, yet pigs do not show any clinical signs due to HEV infection [12]. A systematic review found that in 14 out of 15 studies, farm-level HEV seroprevalences were higher than 60% [10], which is line with our finding that 100% of Dutch pig farms, delivered at least one seropositive pig to slaughter (Meester et al., submitted). On average 15% of pigs at slaughter are found HEV positive in feces in the Netherlands [13]. The wide dissemination of HEV in pig farms and the delivery of HEV positive pigs to slaughter signify the need to understand how HEV persists in farms and how to reduce the transmission of HEV on farm level.

So far, limited knowledge is available on measures that could reduce HEV transmission or the proportion of HEV positive pigs at slaughter. Presumed measures belong to internal and external biosecurity [12]. A few risk factor

studies that compare the management on pig farms with high versus low prevalence of HEV have been performed in the EU and countries with similar pig farming [14-16]. Two of these studies did not have sufficient power to perform multivariable analyses [15, 16]. Hence, more studies are necessary to identify factors associated with HEV prevalence on pig farms. The aim of the current study is therefore to identify factors that are associated with a lower risk of within-farm HEV transmission on Dutch pig farms.

Materials and Methods

Study design and sampling

A case-control stud was performed on pig farms that had previously been sampled at slaughter for a farm prevalence study (Meester et al., submitted). A full description of the sampling strategy and inclusion criteria in that study is given in Meester et al. (submitted). In short, 215 pig farms that delivered pigs to three Dutch abattoirs of one company were selected. Selection was done randomly for organic and conventional pig farms. Repeated cross-sectional sampling of batches delivered to slaughter was performed to collect ~ six blood samples per batch, between January and August 2019. The number of sampled batches differed per farm, which is accounted for in the risk factor analysis. Every individual blood sample was analyzed for HEV antibodies by an enzyme-linked immunosorbent assay (ELISA), and the blood samples were pooled per batch to test for HEV RNA by reverse transcription polymerase chain reaction (RT-PCR). Samples were stored at -20°C until analysis.

Farm selection criteria

The 215 pig farms in the prevalence study had an average farm-level seroprevalence of 76%. All farms delivered at least one seropositive pig to slaughter. Although no HEV negative farms could be identified, analyses did retrieve four farm clusters based on differences in within-farm transmission patterns (Meester et al., submitted).

The clusters entailed cluster 1 with almost solely HEV seropositive, yet PCR negative batches (PCR^-ELISA^+), cluster 2 and 3 with a combination of PCR^-ELISA^+ and PCR^+ELISA^+ batches (i.e. always seropositive, some batches also PCR positive) and finally cluster 4, that consisted of farms that were able to deliver at least one batch to slaughter that was both PCR negative and with at least five out of six pigs seronegative (PCR^-ELISA^- batch) (Meester et al., submitted). It is hypothesized that farms with PCR^-ELISA^- batches are able to prevent within-farm transmission of HEV and that certain biosecurity factors aid this prevention. On that account, for the current case-control study on the one hand farms with low within-farm transmission were selected, being farms with the most PCR^-ELISA^- batches (fig 1., green boxes). One the other hand farms with high within-farm transmission were selected, being farms with the most PCR positive as well as ELISA positive batches (PCR^+ELISA^+, cluster 2 and 3) (fig 1., red boxes). The aim was to include 100 farms, 50 with a low and 50 with a high within-farm transmission. The criteria had to be loosened twice because of insufficient

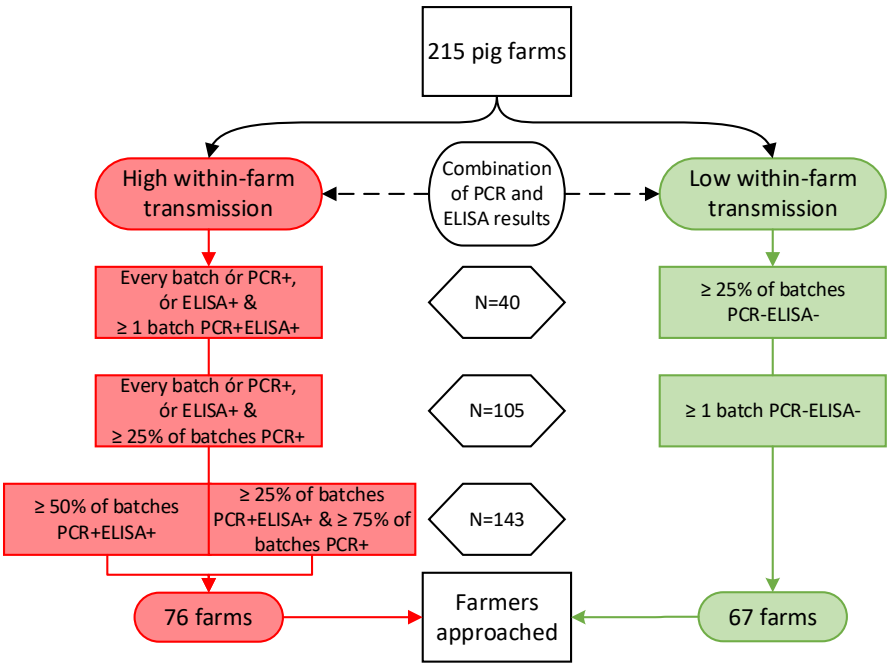

Figure 1 Selection criteria for farms

willingness of farmers to participate (fig 1). In total 143 farms were finally selected and approached for participation. The first approach was done by the slaughter company. Farmers that confirmed willingness to participate were contacted by the research team within one week after confirmation to make an appointment.

Questionnaire and audit design

To investigate factors related to a high risk of within-farm transmission, first a questionnaire (Q) and audit (A) were designed. A mind map was made with all topics of interest and connections between the topics. The Q and A of Walachowski et al. to investigate risk factors for a high HEV farm seroprevalence and the presence of HEV PCR positive livers at slaughter were used as example [14]. All risk factors for a high HEV farm prevalence published in previous studies were included in the Q and A [14-16]. Special attention was given to internal biosecurity, because that concerns measures that reduce spread of pathogens within farms. Questions about the farm management in general, as well as questions specific for different production stages (farrowing for farrow-to-finish (f-t-f), weaning for f-t-f and weaning-to-finish (w-t-f), and fattening for all farms) were included. The main topics of the Q and A are displayed in table 1. Second, the specific questions per topic were comprised and comprehensively discussed with a diplomate of the European College of Porcine Health Management and the first Q and A design were provided with feedback by i.a. pig veterinarians and the slaughter company.

Lastly, the Q and A were pretested at three pig farms, to check the feasibility, understanding and interpretation of questions by the farmers and the amount of time it would take to conduct them.

The final Q was developed in Microsoft Access [17] and existed of 388 questions, of which 210 were binary (true or false); 98 categorical; 33 continuous and 47 open. The A existed of 238 questions/checkpoints, of which 122 were binary, 43 categorical, 50 continuous and six open. Open questions were meant for writing down the answer to a question when the farmer could not choose between the given answer options ('other answer, namely'), so the answer could be categorized later on.

Farm visits

Farm visits were done by eight people (the researcher and seven students of Utrecht University veterinary medicine and the HAS university for applied agricultural sciences) in duos, between March and October 2020. The researcher trained the students by jointly practicing the Q and A at three farms that were not included in the study. During the first visits on included farms, the researcher joined students to give feedback about their visit. After that, students were considered able to do farm visits without onsite supervision. The composition of the duos alternated to limit observer bias. Farmers, students and researcher in the project were blind for farm HEV prevalence and seroprevalence before and during the farm visits.

Data cleaning

The data cleaning process consisted of 6 steps. At first, the questionnaire and audit were manually checked in MS Excel for obvious errors like typos and forgetting decimal points [18]. Besides, 'other answer, namely'-variables were categorized and combined with the categorical question they belonged to. In R [19], variables were checked one by one for a) having any variation in answers; b) having sufficient answers per answer category; c) ability to be combined with other variables. Continuous variables (n=83) were all categorized, because the association with the outcome does not necessarily have to be linear. Questions with less than five answers per category were recategorized by combining categories and questions that (after recategorization in case of >2 categories) had no or hardly any variation in answers were removed from the dataset.

Handling of associated and missing data

Variables that were asked both in the Q and A, or that were very similar, for example the number of fulltime equivalents (FTE) and the number of employees, were assessed for association by a χ^2 test, or Fisher exact test

in case the χ^2 assumptions were not met, to reduce the number of noise variables in the model. In tests with a χ^2 p-value <0.05, one of the two variables was taken out of the dataset. Associations between other potentially related variables were not assessed prior to multivariable analysis, as the method for analysis handles multicollinearity by itself (see paragraph about the statistical analysis hereafter). After that, the number of variables with missing values (NAs) and the proportion of NAs per variables were explored. Variables with more than 15% NAs (arbitrary cut-off) were excluded from further analysis. Because of the limited number of participating farms, multiple imputation by chained equations (MICE) was performed to prevent loss of power in the multivariable analysis [20]. MICE assumes that the probability a value is missing depends only on the observed values, and runs a number of regression models in which every variable is modeled conditional upon the other variables [20]. It considers each missing value to follow a specific distribution (e.g. binary variables modeled using logistic regression and continuous using linear regression), and draws a plausible value from that to replace the missing value [21]. For more information, we refer to Azur et al. and the CRAN package MICE [21, 22]. This method retrieved five multiple imputed datasets (MIDS) that were used in the further analysis.

Statistical analysis

Although farms were selected as being a case or a control based on assumptions of low or high within-farm transmission, a binomial outcome variable would negatively simplify the outcome as there is information available on the number of PCR$^-$ELISA$^-$ batches and the number of positive batches. Therefore, grouped logistic regression with the number of PCR$^-$ELISA$^-$ tested batches relative to the number of other batches per farm was performed, which results in odds ratios for having a PCR$^-$ELISA$^-$ batch. The association between all potential factors and the outcome was assessed by least absolute shrinkage and selection operator (LASSO) regression with the CRAN package glmnet [23]. LASSO regression is a multivariable regularization model that provides sparser models than traditional regression models and is able to handle multicollinearity [24, 25]. Traditional regression can create an infinite number of competing model solutions (with relatively many nonzero parameter estimates) that would all maximize the binomial log-likelihood and would have a high probability of overfitting the model to the data [26]. Overfitting especially occurs in cases where the number of predictors (p) exceeds the number of observations (N) (wide data). LASSO provides sparser, or more conservative, model estimates with less final selected parameters, because it regularizes coefficient estimation. The regularization leads coefficients to have reduced absolute values and some coefficients to be shrunken to zero. Regularization is controlled by a metaparameter λ (formula 1). A variable is selected when the absolute value of its correlation with the outcome is larger than λ [27]. Formula 1 shows how logistic LASSO regression minimizes the negative log-likelihood along with regularization by λ. Here, y is the N-vector of outcomes and X is the N × p matrix of predictors, and the specific form log-likelihood L varies according to the generalized linear model [25].

$$\underset{\beta_0,\beta}{minimize}\left\{-\frac{1}{N}\,L(\beta_0,\beta;y,X)\;+\;\lambda||\beta||_1\right\} \qquad (1)$$

The height of λ for this dataset was determined by 10 fold cross-validation, in which the data is randomly split up in ten parts, and the LASSO regression is performed on all combinations with 9/10th of the data with a wide range of λ values and compared with the remaining 1/10th of the data to see which λ gives the optimal predicting model for the 1/10th of the data. The optimal λ has the minimum mean cross-validated prediction error and is used in the final LASSO regression model with the full dataset [25].

Despite applying cross-validation, the optimal value of λ can still vary between model runs. Also, especially in case of small effect sizes, LASSO still selects false positive variables according to several simulation studies [28, 29]. Nesting λ selection in bootstrap sampling and determining bootstrap confidence intervals of the coefficients of all variables may improve LASSO models through a second step of variable selection [28] and based on the number of times a variable has a coefficient larger than zero (i.e. is selected in the LASSO regression), one can rank variables in terms of stability [30]. So, cross-validation and λ determination was nested within bootstrap sampling, and grouped logistic LASSO regression was performed 500 times.

Table 1 Main themes, subthemes, topics and number (N) of questions in the questionnaire (Q) and audit (A)

Theme	Subtheme	Questions	Q N	A N
General farm characteristics				
	Farm type	Organic or conventional; production stages; Own breeding gilts; genetics	16	
	Personnel	Number; function; For specific tasks	15	
	Hepatitis E virus	Knowledge; Importance	2	
Animals, size, production parameters				
	Buildings	Specific for production stage; Age; Number	4	
	Farm size	Number of accommodations per production stage; Number of batches per year	8	
	Diseases and vaccinations	Salmonella; PRRSv; Influenza; Defined daily dose; Vaccinations per production stage	37	
	Production weaners	Mortality; Age at weaning; Weight	6	1
	Production fatteners	Mortality; Age at fattening; average daily gain; feed conversion rate; difference in age between pigs at slaughter	7	1
Feed, water, manure				
	Feed	Feed type; system; acidification; feed remainders	20	2
	Water origin	private source or municipality; age of private source	5	
	Water cleaning	additional substances in water; cleaning of water system and water	7	
	Manure	Frequency of emptying manure pit; frequency of pen befouling, closed floors	11	24
External biosecurity				
	Hygiene lock	Shower; clothing; boots; contact other pig farms	12	21
	Loading and unloading place	Same place for loading and unloading; cleaning; walking route passes loading place	8	9
	Quarantine	Presence; usage; separate manure pit and air supply	4	7
Internal biosecurity				
	Other animal species	Other farm animals; pets; pigs of other farms	18	3
	Pest control	Presence; protocol; company or private; successfulness; method for control of flies	11	7
	Cleaning	Frequency of cleaning pens; corridors; ceilings; boots; clothes; boards; Method for cleaning pens; corridors; boots	69	
	Disinfection	Frequency of disinfecting pens; corridors; boots; clothes; boards; Type of disinfectants; time between cleaning and disinfection	26	
	Cleanliness of materials	Overalls; boots; corridors; pens; outdoor pens; General score for cleanliness inside and outside farm; floor material		47
	Direct contact between pigs			
	Farrowing	cross-fostering; all in - all out (AIAO)	18	8
	Weaning	mingling during weaning; pen density; transferring pigs to weaning compartment; AIAO	17	24
	Fattening	mingling during fattening; pen density; AIAO	19	22
	Sick-bay	Presence; Emptiness; return from sick-bay to other compartments; specific compartment	6	22
	Indirect contact between pigs			
	Between farm compartments	Treatment round; gloves; walking routes; hygiene lock per production stage	17	36
	Within farm compartments	Period of emptiness compartment; showering of sows; gilt acclimatization	6	4
	Farm equipment, materials, carcasses	Needles; enrichment; equipment per production stage; carcass removal and storage	19	

This retrieved the number of times each variable had a coefficient larger than zero in the model, the mean coefficient over all 500 models and the bootstrap 95% confidence intervals (CI) for the variables. The results were compared between the five MIDS. Then, a multivariable grouped logistic regression model was used to determine the odds ratios for having an HEV PCR$^-$ELISA$^-$ batch at slaughter, for the most stable variables.

RESULTS

Farms

Of 143 approached farmers, 73 were willing to participate (overall participation rate 51%), consisting of 35 farms with at least one PCR$^-$ELISA$^-$ batch and 38 without any PCR$^-$ELISA$^-$ batches. The number of sampled batches per farm ranged from two to 23 and the proportion of PCR$^-$ELISA$^-$ batches from 0.1 to 0.6 (fig. 2). All farms were visited between July and October 2020. Table 2 shows the baseline results of the low and high within-farm transmission farms in the study and the average results for all 215 farms in the original prevalence study (Meester et al., submitted).

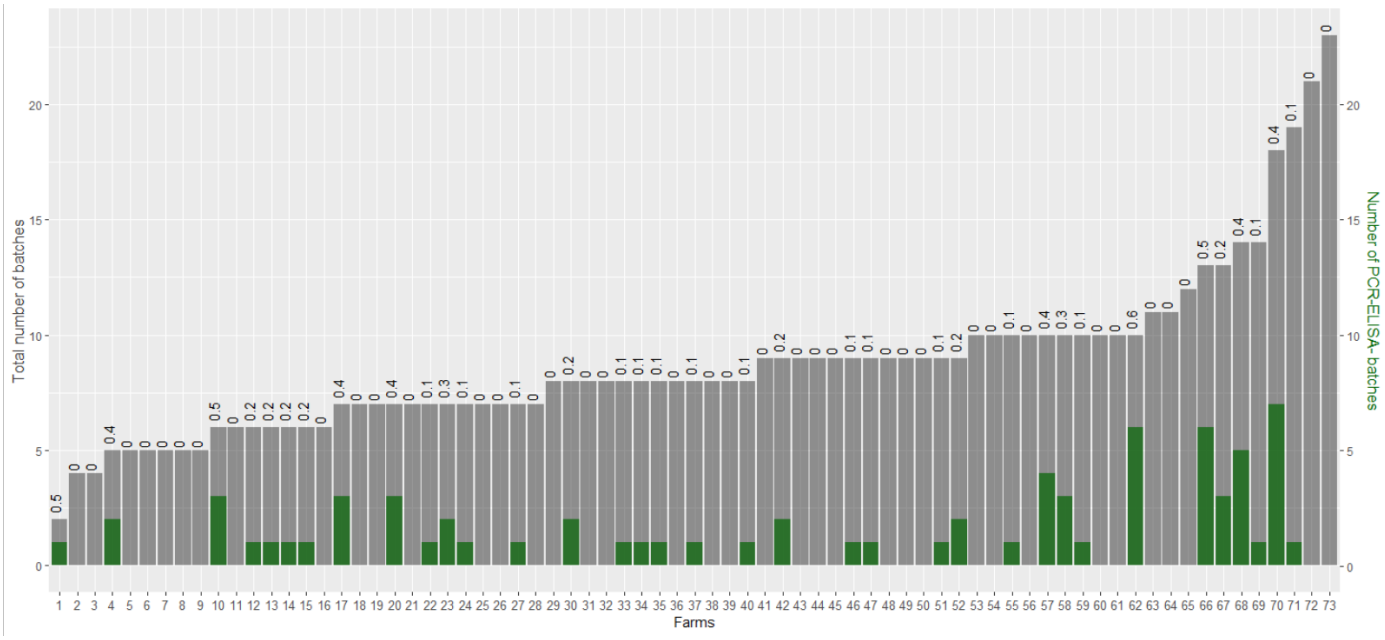

Figure 2 Barplot of the number and proportion of PCR-ELISA- batches per farm compared to the number of total batches sampled

Table 2 Serological and PCR results of included farms with and without PCR-ELISA-batches and all 215 farms in the previous study

	Farms with at least 1 PCR$^-$ELISA$^-$ batch	Farms without PCR$^-$ELISA$^-$ batches	Average results of 215 farms in previous prevalence study
Seroprevalence mean (IQR)	57.0% (39.0 –72.6%)	84.3% (78.6 – 90.4%)	73.6% (66.7 – 87.2%)
PCR positive batch proportion mean (IQR)	27.8% (11.8 – 44.4%)	47.0% (31.2 –69.2%)	40.2% (25.0 – 57.1%)

Data cleaning

Figure 3 shows a flowchart of all steps in the analysis, starting with the data cleaning process. Variables that were excluded for not having any variation had less than eight answers (for questions with 73 answers) in one of the answer categories. The limit of eight was set after attempts with lower limits in which the logistic

regression model could not converge. Seven variables in the questionnaire were dropped for not being interpretable by the farmer or by the researcher in hindsight. Other variables were recategorized (N=114) or combined with each other. For instance, three binary variables about feed type could be combined to one variable with three feed categories. Especially for the audit a lot of variables were combined (104 combined to 27 variables) because for both weaning and fattening compartments four different pens were audited and results were averaged over the four pens per production stage. After data cleaning, 128 questionnaire and 90 audit variables remained (fig. 3).

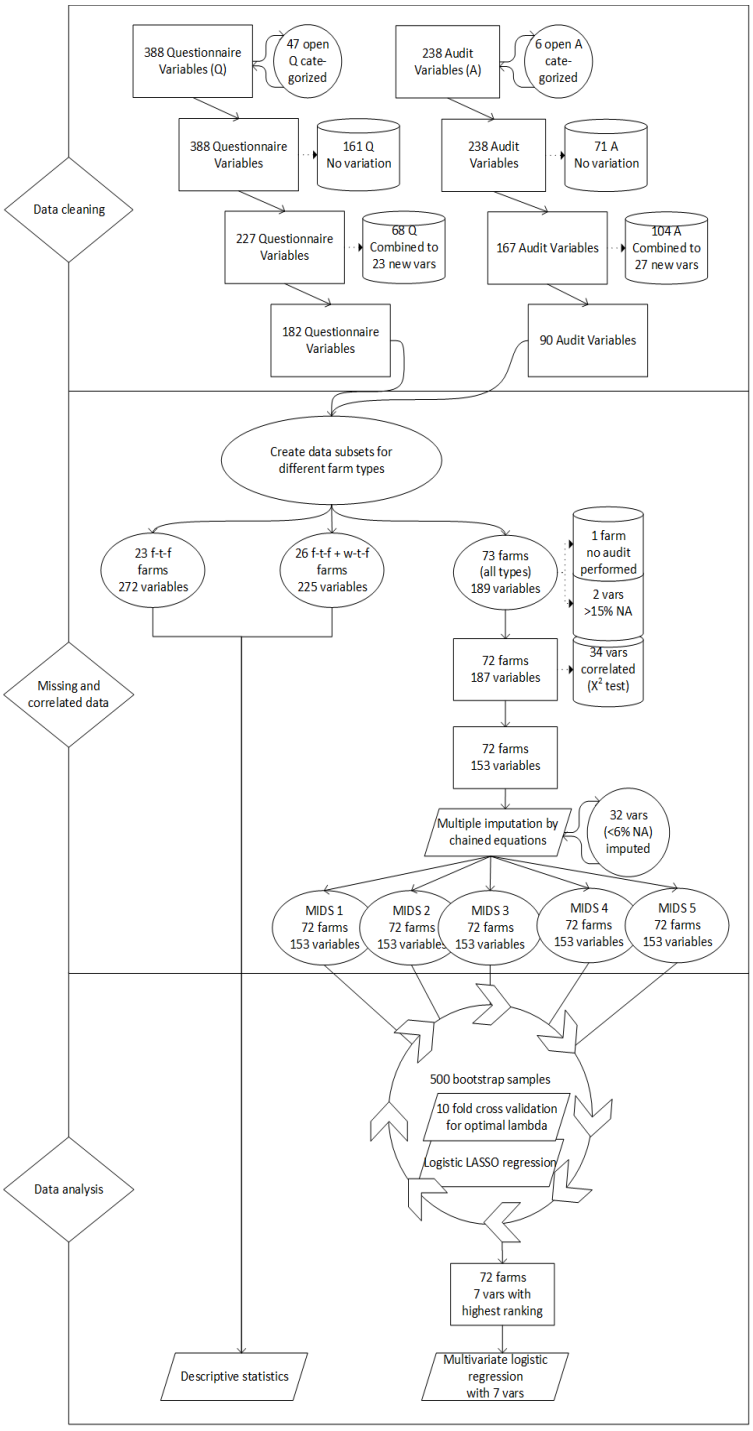

Figure 3 Flowchart of the steps in data cleaning, imputation and analysis

Subsets, related variables and missing data

Because specific questions were asked per production stage, subsets of the dataset were made, giving a f-t-f dataset with 23 farms, a f-t-f and w-t-f dataset with 26 farms and a dataset with general farming and fattening questions for all 73 farms (fig 3., second block). The subsets for f-t-f and w-t-f did not contain sufficient farms for analysis. Descriptive statistics for those subsets are provided in additional file 1.

For the subset of variables for all farms, one farm was taken out because the audit could not be performed there. Of 189 remaining variables, 34 were dropped because of a significant association with another variable. Furthermore, two variables were dropped because more than 15% of answers was missing. That left 32 variables in the subset with between 1.4 and 6.9% missing values that were imputed five times.

Lasso regression

The bootstrap logistic LASSO regression with cross-validated λ selection was performed on a dataset with 153 potential risk factors, for all 5 MIDS. Variables were selected between 0 and 445 times out of 500 bootstrap samples without large differences between the MIDS. Figure 4 displays how often the 15 highest ranked variables were selected on average. We decided to select the first seven variables for the final model, because the marginal difference in how often variables are selected is higher between variable 7 (unloading place next to air inlet) and 8 (functional hygiene lock) than between variable 6 (boots have profiled soles) and 7. Besides, in all five MIDS bootstrap LASSOs the first seven variables belonged to the highest seven, showing the

stability of the variables, whereas the 8th to 10th variables were only in the highest ranking in three or four MIDS. The bootstrap 95% CI of the coefficients of these variables were all below or above one so the final model could not be made sparser by looking at CIs like suggested by Laurin et al [28]. Therefore the final selected variables that are associated with the odds of having a PCR-ELISA- batch are rubber and steel floor material in fattening pens, cleaning of pig boards, a long average fattening period, fly control by predatory flies, boots with profiled soles and a loading or unloading place next to the air inlet of a barn. The multivariable logistic regression model with these variables retrieved odds ratios and 95% CIs that can be found in table 3.

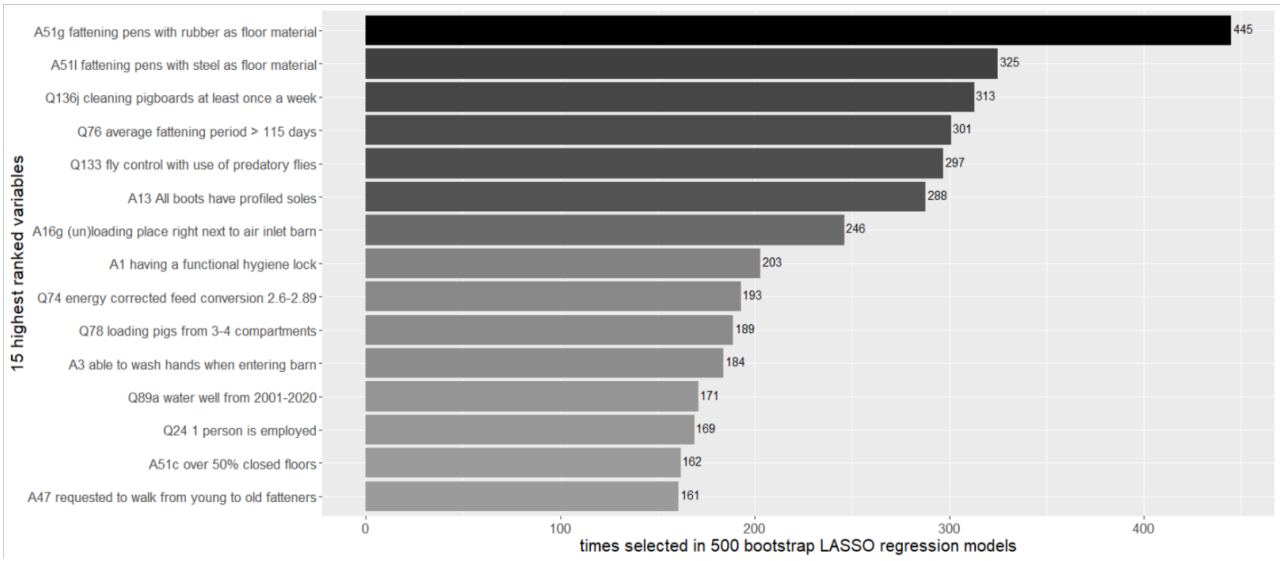

Figure 4 15 highest ranked variables according to 500 bootstrap logistic LASSO regression models, by how often variables were selected

Table 3 Odds ratios and 95% confidence intervals in final multivariable logistic regression model, for having an HEV PCR-ELISA- batch of pigs delivered to slaughter

Variable	Audit (A) or Questionnaire (Q)	Odds Ratio	2.5%	97.5%
Fattening pens without rubber as floor material	A	Reference	-	-
Fattening pens with rubber as floor material	A	5.87*	3.03	11.6
Fattening pens without steel as floor material	A	Reference	-	-
Fattening pens with steel as floor material	A	7.13*	3.05	16.9
Cleaning pig boards never or less than once a week	Q	Reference	-	-
Cleaning pig boards once a week or after every pig contact	Q	1.99*	1.07	3.80
Average fattening period (2019) ≤ 105 days	Q	Reference	-	-
Average fattening period (2019) >105 ≤115 days	Q	0.26*	0.12	0.58
Average fattening period (2019) > 115 days	Q	0.21*	0.09	0.45
No fly control	Q	Reference	-	-
Fly control with pesticides sprayed on walls	Q	1.34	0.54	3.34
Fly control with pesticides in manure pit	Q	1.75	0.73	4.35
Fly control with use of predatory flies	Q	4.52*	1.59	13.5
All boots have smooth soles	A	Reference	-	-
Some boots have smooth, others have profiled soles	A	0.65	0.27	1.57
All boots have profiled soles	A	0.81	0.32	2.27
(Un)loading place away from air inlet barn	A	Reference	-	-
(Un)loading place right next to air inlet barn	A	0.80	0.41	1.54

DISCUSSION

The aim of this study was to find factors that are associated with the ability of pig farms to deliver an HEV free batch of pigs to slaughter (PCR^-ELISA^-) to understand how to reduce transmission of HEV within farms and thereby reduce the proportion of HEV positive pigs at slaughter.

Rubber and steel on the floor of fattening pens were the factors with the highest ranking in the bootstrap grouped logistic LASSO regression and were significantly associated with the odds of a PCR^-ELISA^- batch of pigs with ORs of 5.9 and 7.1 respectively. Steel as floor material is often used as slatted floor, for feces and urine to run down into the manure pit. Steel slats are narrower than concrete slats, so having steel may reduce the chance that pigs have contact with HEV contaminated feces of pen mates. Rubber floor material is used at spots were the floor may otherwise quickly deteriorate, for instance in front of the feed through (Dr. Peter van der Wolf, European Specialist Porcine Health Management, personal communication). Moreover, having rubber or steel as part of the pen floor, by definition means less floor surface is made of concrete. Concrete is a porous material and may be harder to clean sufficiently. Therefore, rubber and steel as floor materials may reduce transmission of HEV between consecutive batches of pigs within compartments. Almost all pig farms had fattening pens with partly concrete floors, so the question whether the pen floor contained concrete had to be removed from the dataset.

Three variables associated with the outcome may prevent transmission between groups of pigs that are simultaneously present in the barn, namely the frequency of cleaning pig driving boards, the type of measure used to control barn flies and the type of boots used (insignificant). Pig driving boards are used to move pigs between barns and if not cleaned properly and frequently, they may mechanically spread HEV between pigs of several ages and locations within the farm. Boots with profiled soles are less easily cleaned than smooth soles so farmers may carry contaminated feces from compartment to compartment, leading to a lower odds of a HEV PCR^-ELISA^- in case profiled boots were present. Three measures to reduce barn flies all increase the odds of delivering PCR^-ELISA^- batches of pigs to slaughter compared to not applying any measure. Flies may be carriers of viruses like HEV and spread the virus within farms. This mechanical transport has been described for several viruses like porcine circovirus 2b, rotavirus and porcine reproductive and respiratory syndrome virus [31-34]. The most effective way of fly control to reduce HEV transmission according to this study is the use of predatory flies. Predatory flies are mainly applied by professional exterminators because it takes knowledge and experience to maintain a stable number of predatory flies without those flies becoming a plague (Dr. Ir. Joost van den Borne, personal communication). It is possible that fly control by predatory flies is associated with how professional and thereby successful fly control on the farm is.

A longer fattening period reduces the odds of delivering a PCR^-ELISA^- batch of pigs to slaughter. A previous risk factor study showed that a large age gap between the youngest and oldest pig in a batch of pigs increases the risk of HEV positive livers at slaughter [14]. Both risk factors correspond to age dependent HEV results reported in other studies, namely that HEV seroprevalence rises with age (e.g. [35-37]) and the virological prevalence falls with age [38]. Other risk factors found in previous studies were included in the Q and A but have not been found to be associated with HEV in the current study. For instance, external biosecurity factors like having a quarantine period, a sanitary ford, or contact between pigs and other domestic species [15] as well as demanding showering and wearing farm-specific boots for visitors before coming into the farm [16] have been asked in the questionnaire but are not associated with the outcome. The difference between previous and our results may be due to a different type of outcome. Whereas other studies have HEV virological or serological prevalence of the farm as a whole as outcome, we have looked at a combination of virological and serological prevalence of farm deliveries to slaughter, presumably demonstrating ability to prevent HEV transmission between compartments. Factors that cannot differ between compartments, like having a hygiene lock, a certain breed of pigs or the source of the water, could in theory not, and were not associated with our outcome.

The last factor selected in the final model is the location of the loading and/or unloading place. The odds of having a PCR^-ELISA^- batch decrease when the loading or unloading place was adjacent to an air inlet of a barn, yet the OR is insignificant. The tentative association may indicate that HEV positive pigs that are loaded or unloaded and shed virus (maybe increased shedding due to stress of moving) expose susceptible pigs inside the barn to HEV via the air inlet. This variable should be included in future risk factor studies as well, although it was dubious whether this 7th most stable variable had to be included in the final logistic regression model, as it was selected 246 times in the bootstrap LASSO, compared to 288 times for the 6th most stable variable (fig. 4). LASSO regression with a stability analysis by bootstrapping brings about an arbitrary choice of the number of selected variables, which is a disadvantage of the method. Green et al. (in press) have suggested to determine the cut-off of stability (a stability threshold) by random permutation of the outcome variable and repeating the bootstrap LASSO with that permuted outcome, to compare the frequency of selection of a variable with the actual, and with the permuted outcome [39]. Although this is beyond the scope of the current study, this could be an interesting addition to future risk factor analyses with wide data.

Using an audit besides a standardized questionnaire in this study has shown great value for multiple reasons. Firstly, a part of the variables could only be objectively scored with the audit, like the amount of flies in the barn. Moreover, incongruence between answers of farmers and farm auditors could be assessed, for instance to the question whether pets are allowed inside the barn. Lastly, farmers sometimes realized that answers given in the questionnaire did not strike with findings in the audit, after which they confessed that some answers were socially desired. Therefore, announcing an audit on forehand may prevent response bias. An audit alone would not suffice as it gives an objective view of only one moment in time whereas the farmer can explain his or her management through time.

The batches were sampled a year before the farm visits were conducted. All questions regarding management that could have changed in a year's time included the words 'in 2019' to ensure that the answer applied to the sampled batches. We also asked whether the farmer had changed his or her management in the last year and what changes those were, to try to account for it. Still, the interval between sampling and collecting farm management data may very well have led to recall bias and misclassification bias. It is expected that both biases are non-differential between farms with and without PCR^-ELISA^- batches, because farmers were unaware of the HEV results during the farm visit.

Data analysis by bootstrap grouped logistic LASSO regression with nested cross-validation for λ selection is advisable in case the number of parameters exceeds the number of observations to prevent overfitting of the model on the data and selection of a lot of false positive variables. However, LASSO regression analysis with a random intercept or random slope is not yet available. Although a random farm intercept in the current study could have been applied because of the grouped binomial outcome and may have explained a part of the variation in the data, all questions in the questionnaire and audit were on farm level, instead of batch level, so a random farm intercept was not deemed necessary.

To summarize, internal biosecurity measures like the efficacy of cleaning fattening pens, frequency of cleaning fomites like pig boards, and barn fly control, in particular by applying predatory flies, may reduce transmission of HEV between farm compartments and as a result increase the number of HEV free batches of pigs at slaughter. Implementing these measures on pig farms may hence reduce the risk of HEV exposure to pork consumers. An intervention study aiming to keep a farm compartment within an HEV affected farm free from HEV by implementation of the measures found is necessary to demonstrate inference.

Acknowledgements

This study could not have been performed without the help of some amazing people. I would like to thank Jan van den Broek for fruitful discussions about the analysis. My favorite quote of his is 'if you have simple data, you can do a simple analysis'. Thanks as well to Joris van Beers of Vion Food Group for approaching farmers and

convincing them to participate, all participating farmers who shared their management and showed their farm, Peter van der Wolf (Ceva), Manon Houben (Royal GD) and Joost van den Borne (HAS lector Healthy Farming) for their feedback on the questionnaire, and students Carmijn Meulenbroek, Kaïn Saygan, Nick van Doorn, Lisa Verhoeven, Laura van den Ameele, Teus van Alphen and Aart van Dam for conducting the farm visits with great conviction.

I wish to express gratitude to my supervisors, Prof. Dr. Arjan Stegeman, Prof. Dr. Wim van der Poel, Dr. Martijn Bouwknegt and Dr. Tijs Tobias. Arjan, for being able to conclude on things in five minutes that I have been in doubt about for five weeks, Wim for sharing his great knowledge about virology and lab techniques, Martijn for always being critical about my statistical analysis and in particular about AICs vs. p-values, and last but not least, Tijs, for dragging me through the long process of data cleaning by motivational talks and good questions and his everlasting optimism even when trying to conduct questionnaires in covid times.

Funding

This work was part of the project "HEVentie: hepatitis E virus intervention in primary pig production". HEVentie receives financial support of the Topsector AgriFood (TKI AF 18119). Within TKI Agri&Food private partners, research institutes and government cooperate to innovations for safe and healthy food for 9 billion people on a resilient globe.

Conflict of interest

The authors declare that they have no competing interests.

REFERENCES

1. World Health Organization: **Hepatitis E Fact Sheet**. 27 July 2021, accessed 25 Februari 2022. Retrieved from https://www.who.int/news-room/fact-sheets/detail/hepatitis-e.
2. Nan Y, Zhang YJ: **Molecular Biology and Infection of Hepatitis E Virus**. *Frontiers in microbiology* 2016, **7**:1419.
3. Rui P, Zhao F, Yan S, Wang C, Fu Q, Hao J, Zhou X, Zhong H, Tang M, Hui W *et al*: **Detection of hepatitis E virus genotypes 3 and 4 in donkeys in northern China**. *Equine Veterinary Journal* 2020, **52**(3):415-419.
4. Dziedzinska R, Krzyzankova M, Bena M, Vasickova P: **Evidence of Hepatitis E Virus in Goat and Sheep Milk**. *Viruses* 2020, **12**(12).
5. Bigoraj E, Kozyra I, Kwit E, Rzeżutka A: **Detection of hepatitis E virus (rabbit genotype) in farmed rabbits entering the food chain**. *International journal of food microbiology* 2020, **319**:108507.
6. Lapa D, Capobianchi MR, Garbuglia AR: **Epidemiology of Hepatitis E Virus in European Countries**. *International Journal of Molecular Science* 2015, **16**(10):25711-25743.
7. Kamar N, Pischke S: **Acute and Persistent Hepatitis E Virus Genotype 3 and 4 Infection: Clinical Features, Pathogenesis, and Treatment**. *Cold Spring Harbor perspectives in medicine* 2018.
8. Kamar N, Marion O, Abravanel F, Izopet J, Dalton HR: **Extrahepatic manifestations of hepatitis E virus**. *Liver international: official journal of the International Association for the Study of the Liver* 2016, **36**(4):467-472.
9. EFSA BIOHAZ Risk Assesment Panel (EFSA Panel on Biological Hazards): **Scientific Opinion on the public health risks associated with hepatitis E virus (HEV) as a food-borne pathogen**. *EFSA Journal* 2017, **15**(7).
10. Salines M, Andraud M, Rose N: **From the epidemiology of hepatitis E virus (HEV) within the swine reservoir to public health risk mitigation strategies: a comprehensive review**. *Veterinary research* 2017, **48**(1):31.

11. Bouwknegt M, Rutjes SA, Reusken CB, Stockhofe-Zurwieden N, Frankena K, de Jong MC, de Roda Husman AM, Poel WH: **The course of hepatitis E virus infection in pigs after contact-infection and intravenous inoculation**. *Veterinary Research* 2009, **5**:7.
12. Meester M, Tobias TJ, Bouwknegt M, Kusters NE, Stegeman JA, van der Poel WHM: **Infection dynamics and persistence of hepatitis E virus on pig farms - a review**. *Porcine Health Management* 2021, **7**(1):16.
13. Hakze-van der Honing RW, van Coillie E, Antonis AF, van der Poel WH: **First isolation of hepatitis E virus genotype 4 in Europe through swine surveillance in the Netherlands and Belgium**. *PloS one* 2011, **6**(8):e22673.
14. Walachowski S, Dorenlor V, Lefevre J, Lunazzi A, Eono F, Merbah T, Eveno E, Pavio N, Rose N: **Risk factors associated with the presence of hepatitis E virus in livers and seroprevalence in slaughter-age pigs: a retrospective study of 90 swine farms in France**. *Epidemiology and infection* 2014, **142**(9):1934-1944.
15. Lopez-Lopez P, Risalde MLA, Frias M, Garcia-Bocanegra I, Brieva T, Caballero-Gomez J, Camacho A, Fernandez-Molera V, Machuca I, Gomez-Villamandos JC *et al*: **Risk factors associated with hepatitis E virus in pigs from different production systems**. *Veterinary microbiology* 2018, **224**:88-92.
16. Wilhelm B, Leblanc D, Leger D, Gow S, Deckert A, Pearl DL, Friendship R, Rajic A, Houde A, McEwen S: **Farm-level prevalence and risk factors for detection of hepatitis E virus, porcine enteric calicivirus, and rotavirus in Canadian finisher pigs**. *Canadian journal of veterinary research* 2016, **80**(2):95-105.
17. Microsost Cooperation: **Microsoft Access**. 2102 edn; 2018.
18. Microsost Cooperation: **Microsoft Excel**. 2102 edn; 2018.
19. R Core Team: **R: A Language and Environment for Statistical Computing**. In. Veinna, Austria: R Foundation for Statistical Computing; 2020.
20. Schafer JL, Graham JW: **Missing data: our view of the state of the art**. *Psychological Methods* 2002, **7**(2):147-177.
21. Azur MJ, Stuart EA, Frangakis C, Leaf PJ: **Multiple imputation by chained equations: what is it and how does it work?** *International Journal of Methods in Psychiatry Research* 2011, **20**(1):40-49.
22. Van Buuren S, Groothuis-Oudshoorn K: **mice: Multivariate Imputation by Chained Equations in R**. *Journal of Statistical Software* 2011, **45**(3):1-67.
23. Friedman J, Hastie T, Tibshirani R: R**egularization Paths for Generalized Linear Models via coordinate descent**. *Journal of Statistical Software* 2011, **39**(5):1-13.
24. Chong IG, Jun CH: **Performance of some variable selection methods when multicollinearity is present. Chemometr Intell Lab Syst 78: 103-112**. *Chemometrics and Intelligent Laboratory Systems* 2005, **78**:103-112.
25. Hastie T, Tibshirani R, Wainwright M: **Statistical learning with sparsity: the lasso and generalizations**: Chapman & Hall/CRC press; 2015.
26. Tibshirani R: **Regression Shrinkage and Selection via the Lasso**. *Journal of the Royal Statistical Society Series B (Methodological)* 1996, **58**(1):267-288.
27. Efron B, Hastie T, Johnstone I, Tibshirani R: **Least angle regression**. *The Annals of Statistics* 2004, **32**:407-499.
28. Laurin C, Boomsma D, Lubke G: **The use of vector bootstrapping to improve variable selection precision in Lasso models**. *Statistical Applications in Genetics and Molecular Biology* 2016, **15**(4):305-320.
29. Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ: **Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events**. *Statistics in Medicine* 2016, **35**(7):1159-1177.
30. Lima E, Davies P, Kaler J, Lovatt F, Green M: **Variable selection for inferential models with relatively high-dimensional data: Between method heterogeneity and covariate stability as adjuncts to robust selection**. *Science Reports* 2020, **10**(1):8002.

31. Blunt R, McOrist S, McKillen J, McNair I, Jiang T, Mellits K: **House fly vector for porcine circovirus 2b on commercial pig farms**. *Veterinary microbiology* 2011, **149**(3-4):452-455.
32. Tan SW, Yap KL, Lee HL: **Mechanical transport of rotavirus by the legs and wings of Musca domestica (Diptera: Muscidae)**. *Journal of Medical Entomology* 1997, **34**(5):527-531.
33. Otake S, Dee SA, Moon RD, Rossow KD, Trincado C, Pijoan C: **Studies on the carriage and transmission of porcine reproductive and respiratory syndrome virus by individual houseflies (Musca domestica)**. *The Veterinary record* 2004, **154**(3):80-85.
34. Rochon K, Baker RB, Almond GW, Gimeno IM, Pérez de León AA, Watson DW: **Persistence and Retention of Porcine Reproductive and Respiratory Syndrome Virus in Stable Flies (Diptera: Muscidae)**. *Journal of Medical Entomology* 2015, **52**(5):1117-1123.
35. Seminati C, Mateu E, Peralta B, de Deus N, Martin M: **Distribution of hepatitis E virus infection and its prevalence in pigs on commercial farms in Spain**. *Veterinary journal* 2008, **175**(1):130-132.
36. De Deus N, Casas M, Peralta B, Nofrarias M, Pina S, Martin M, Segales J: **Hepatitis E virus infection dynamics and organic distribution in naturally infected pigs in a farrow-to-finish farm**. *Veterinary microbiology* 2008, **132**(1-2):19-28.
37. Tsachev I, Baymakova M, Ciccozzi M, Pepovich R, Kundurzhiev T, Marutsov P, Dimitrov KK, Gospodinova K, Pishmisheva M, Pekova L: **Seroprevalence of Hepatitis E Virus Infection in Pigs from Southern Bulgaria**. *Vector Borne Zoonotic Diseases* 2019.
38. Capai L, Maestrini O, Casabianca F, Villechenaud N, Masse S, Bosseur F, de Lamballerie X, Charrel RN, Falchi A: **Drastic decline of hepatitis E virus detection in domestic pigs after the age of 6 months, Corsica, France**. *Transboundary and emerging diseases* 2019, **66**(6):2462-2473.
39. Green M, Lima E, Hyde R: **Selection Stability in High Dimensional Statistical Modelling: Defining a Threshold for Robust Model Inference**. In press.