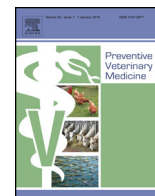




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Use of Bayesian Belief Network techniques to explore the interaction of biosecurity practices on the probability of porcine disease occurrence in Canada



Ruth Cox*, Crawford W. Revie, Daniel Hurnik, Javier Sanchez

Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, PE C1A 4P3, Canada

ARTICLE INFO

Article history:

Received 2 October 2015

Received in revised form 14 March 2016

Accepted 28 June 2016

Keywords:

Bayesian Belief Network

Disease probability

Swine farm

Biosecurity

ABSTRACT

Identification and quantification of pathogen threats need to be a priority for the Canadian swine industry so that resources can be focused where they will be most effective. Here we create a tool based on a Bayesian Belief Network (BBN) to model the interaction between biosecurity practices and the probability of occurrence of four different diseases on Canadian swine farms. The benefits of using this novel approach, in comparison to other methods, is that it enables us to explore both the complex interaction and the relative importance of biosecurity practices on the probability of disease occurrence.

In order to build the BBN we used two datasets. The first dataset detailed biosecurity practices employed on 218 commercial swine farms across Canada in 2010. The second dataset detailed animal health status and disease occurrence on 90 of those farms between 2010 and 2012. We used expert judgement to identify 15 biosecurity practices that were considered the most important in mitigating disease occurrence on farms. These included: proximity to other livestock holdings, the health status of purchased stock, manure disposal methods, as well as the procedures for admitting vehicles and staff. Four diseases were included in the BBN: Porcine reproductive and respiratory syndrome (PRRS), (a prevalent endemic aerosol pathogen), Swine influenza (SI) (a viral respiratory aerosol pathogen), *Mycoplasma pneumonia* (MP) (an endemic respiratory disease spread by close contact and aerosol) and Swine dysentery (SD) (an enteric disease which is re-emerging in North America).

This model indicated that the probability of disease occurrence was influenced by a number of manageable biosecurity practices. Increased probability of PRRS and of MP were associated with spilt feed (feed that did not fall directly in a feeding trough), not being disposed of immediately and with manure being brought onto the farm premises and spread on land adjacent to the pigs. Increased probabilities of SI and SD were associated with the farm allowing access to visiting vehicles without cleaning or disinfection. SD was also more likely to occur when the health status of purchased stock was not known. Finally, we discuss how such a model can be used by the Canadian swine industry to quantify disease risks and to determine practices that may reduce the probability of disease occurrence.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Effective biosecurity strategies need to determine the means by which a pathogen can be introduced and transmitted in order to identify major or minor pathogen threats, so that prevention measures can be placed where they will be most effective. In Canada, not enough money, time, and effort have been invested in identifying and quantifying transmission risks for important swine pathogens (Desrosiers, 2011). Lack of understanding of potential

threats has resulted in unsuccessful control efforts and in losses that can jeopardize both individual businesses and the swine industry. Outbreaks of Swine influenza virus and Porcine Circovirus, for example, have contributed to the considerable reduction in the number of swine farms in Canada (Brisson, 2014), while more recently, the emergence of Porcine Epidemic Diarrhea (PED) has had considerable economic impact (Paarlberg, 2014). In the past, Porcine reproductive and respiratory syndrome (PRRS), was the most costly pig disease for more than two decades, (costing the US swine industry \$560 million per year) (Neumann et al., 2005), yet it took 20 years to identify that airborne transmission was an important means of spread (Desrosiers, 2011).

* Corresponding author.

E-mail address: rucox@upe.ca (R. Cox).

Here we create a tool, known as a Bayesian Belief Network (BBN), which can be used to identify and quantify the probability of disease occurrence in Canadian swine farms. A BBN is a probabilistic graphical model which represents a network of nodes connected by directed links that represent a probability function (Jensen, 2001). BBN models allow users to make informed decisions about a range of possible outcomes using information based on prior evidence (Fenton and Neil, 2013). BBNs have previously been used in the veterinary domain, for example, to aid disease diagnosis (McKendrick et al., 2000; Seidel et al., 2003; Otto and Kristensen, 2004) and to assess associations between biosecurity practices and disease outbreak (Firestone et al., 2014). These studies focused on one specific disease, and BBNs have not, to our knowledge, been used to assess the relative impact of biosecurity practices simultaneously on a group of diseases.

BBNs are ideally suited to the statistical analyses of data from complex epidemiological systems (McCormick et al., 2013). We chose to design a BBN in this context for two reasons. First BBNs provide a method of consolidating evidence in a consistent and mathematically robust manner. Unlike more traditional methods of data analysis, they can incorporate a large number of predictors and a number of interactions (Fenton and Neil, 2013). They can therefore be used to investigate causal relations between events, weigh the consequences of actions and identify unintended side effects. Here we highlight the utility of a BBN, in particular how one might be used to assess on-farm scenarios and to determine the trade-offs that must be made during decision making on any farm operation. Second, an innovative feature of BBNs is that they do not require precise probabilities to calculate the outcome and can provide good results even when only approximate probabilities are available (Ben-Gal, 2007). This is an advantage to the swine industry where precise prior information about emerging disease events is often not available and because stakeholder opinions (e.g. producers, veterinarians and allied industry personnel) can vary widely (Marvin et al., 2010).

Our objectives are to demonstrate the utility of the BBN for evaluating the effectiveness of biosecurity practices on disease occurrence, and for evaluating how biosecurity scenarios could reduce probability of disease. We discuss the limitations of the approach, based on the amount of data available, and we comment on how future data collection could be focused to allow more complete analysis and model development.

2. Method

In order to build a BBN we used data that detailed (1) the biosecurity practices used on farms across Canada and (2) the occurrence of disease on these farms. These data were collected in two different ways as described below.

2.1. Farm features and biosecurity data

In spring 2010 a detailed biosecurity survey was conducted by the Canadian Swine Health Board (CSHB) to acquire knowledge about the management and biosecurity practices in the Canadian swine industry. The survey was conducted at 218 commercial farms that were situated in 5 regions of Canada – British Columbia (BC), the Prairies, Ontario, Quebec and the Maritimes (Table 1). Similar surveys were also conducted at approximately 100 breeder farms and approximately 40 boar stud farms. The surveys were completed by trained assessors who were Canadian Quality Assurance (CQA™) validators, the majority of whom are practicing veterinarians. All assessors attended one of two training sessions to ensure that all questions were asked and recorded in a consistent manner.

The assessor collected the information using a questionnaire containing 145 questions, each with approximately five or six categories. Full results of the survey were reported internally to the Canadian Swine industry (Canadian Swine Health Board, 2010). Specific findings of the survey will not be reported here; rather we will describe how we used the data for BBN construction.

Each survey question was weighted by ten swine industry experts who were members of the Canadian Association of Swine Veterinarians (CASV). The criteria were weighted according to how important the farm feature or biosecurity practice is for the prevention of any and all disease. Each question was given a weight of between 1 and 10 (10 = most important) by each expert. The mean weight was calculated and agreed upon by the experts. Thus the expert elicitation followed a Delphi style approach, where weights were assigned and later agreed upon as a group (O'Hagan et al., 2006). This process occurred at the time of the biosecurity survey (before our work began) and we therefore had no input about the method of expert elicitation. Any question that scored a mean weight of more than 6 was included in our analysis. This gave a total of 14 questions (Table S1), which represent 14 biosecurity practices that were incorporated into the BBN.

2.2. Disease data

2.2.1. Selection of diseases

Following discussion with experts at UPEI and CSHB, six diseases were selected as 'test' diseases for the BBN. They were selected because they were of particular concern to the Canadian swine industry.

Porcine Reproductive and Respiratory Syndrome (PRRS *virus*) (PRRS) – a prevalent endemic disease transmitted via aerosol.

Swine Influenza (*Swine influenza virus*) (SI) – a viral respiratory aerosol pathogen.

Mycoplasma pneumonia (*Mycoplasma hyopneumoniae*) (MP) – an endemic respiratory disease transmitted by close contact and aerosol.

Swine dysentery (*Brachyspira hyodysenteriae* or novel strains) (SD) – an enteric disease which is re-emerging in Canada.

Transmissible Gastro-Enteritis (*Transmissible Gastro-Enteritis Virus*) (TGE) – a highly infectious coronavirus.

Pleuropneumonia (*Actinobacillus Pleuropneumonia*) (APP) – a respiratory bacterial disease that is spread by aerosol or direct contact.

2.3. Collection of disease data

A questionnaire was designed to collect information about the occurrence of the six diseases on the farms that had taken part in the biosecurity survey in 2010 (Fig. S1). The questionnaire was administered via email as a Microsoft Word 2007 document and online via the Fluidsurveys website in 2013 (Fluidsurveys, 2013). The questionnaire presented the list of six diseases and asked three questions: (1) what is the identification number of the farm? (2) what was the health status of the farm in 2010 at the time when the biosecurity survey was conducted? (3) were there any disease outbreaks on the farm between January 2010 and December 2011? The questionnaire was emailed by CSHB researchers to all veterinarians (n = 40) who completed the biosecurity survey for the 218 farms. The veterinarians were invited to complete the survey for each farm within 2 weeks. One reminder email was sent to non-responders after 2–3 weeks. Vets were instructed that they would receive a payment of \$100 for participation. Farm identification and precise location of the farm was only known by the veteri-

Table 1
Number of commercial farms included in the CSHB biosecurity survey per region of Canada and number of responders to the disease occurrence questionnaire.

Region	Farrow-to-wean		Farrow-to-finish		Finisher		Total	
	No. farms	No. responders (%)	No. farms	No. responders (%)	No. farms	No. responders (%)	No. farms	No. responders (%)
BC	1	1 (100)	16	15 (94)	3	3 (100)	20	19 (95)
Prairies	21	5 (24)	26	7 (27)	24	4 (17)	71	16 (8)
Ontario	14	0 (0)	17	3 (18)	21	5 (24)	52	8 (15)
Quebec	16	9 (56)	10	5 (50)	22	15 (68)	42	29 (69)
Maritimes	9	3 (33)	13	10 (77)	5	5 (100)	27	18 (67)
Total	61	18 (30)	82	40 (49)	75	32 (43)	218	90 (41)

narians and was not disclosed to the authors. Disease reporting was considered to be accurate because the veterinarians were able to consult their records that they made at the time of diagnosis. Data about farm biosecurity practices and disease status were collated in one database. The farm identification number allowed us to match anonymously the disease information from our survey with the biosecurity information on that farm.

2.4. BBN construction

In brief, a BBN is a model which describes the direct dependencies between a set of variables. It is represented as a directed acyclic graph, in which nodes (representing variables) are connected by arrows that represent directed causal relations. Each node, which has a number of states, contains a conditional probability table (CPT), which specifies the conditional probability of the node being in a specific state given the state of its parent nodes. For example, the BBN built here represents the relationship between diseases and biosecurity practices. Given specific biosecurity practices, the BBN uses Bayes' rule to calculate posterior probabilities of disease occurrence. Thus the term 'Bayesian' in BBN refers to the prominent role of Bayes' rule of probability. Detailed description of Bayesian networks is provided in Jensen (2001) and Fenton and Neil (2013).

The BBN model was constructed using the freely available software GeNIe (<http://genie.sis.pitt.edu/>). The following steps were adopted when constructing the BBN models, following Fenton and Neil (2013), Chapter 6:

- 1) Identification of relevant variables;
- 2) Creation of the BBN structure;
- 3) Identification of variables that require direct links and specification of the relevant conditional probability table;
- 4) Model validation and testing.

2.4.1. Identification of relevant variables

There were 14 biosecurity practices that were considered to be the most important for the prevention of any and all disease by members of the CASV (Table S1). Thirteen of these were included in the BBN, while one ("frequency with which carcasses are removed from pens") was excluded because all farms adopted the same practice. Two additional nodes were included representing region of Canada and farm type since this information is relevant to disease occurrence. One additional node represented disease occurrence.

2.4.2. Creation of the BBN structure

A BBN model was built based on a naïve Bayes structure (Fig. S2) in which the disease node was linked to all other nodes, while none of the other nodes were linked directly to each other. This is because each node directly influences disease occurrence. While alternative, non-naïve, causal structures that assess node interactions and latent variables (Otto and Kristensen, 2004) could be explored, we began with a naïve Bayes approach for simplicity (Fenton and Neil, 2013).

The possible states of each node are listed in Table S1. For example, the central disease node represents the probability of disease occurrence; it has five states i.e. one for each disease and one for no disease (ND). Some biosecurity practices had up to eight states in the original questionnaire, however, for simplicity we combined states where biologically possible (particularly when information about disease was not available for some states) to improve accuracy (Marcot et al., 2006).

2.4.3. Identification of variables that require direct links and specification of the CPT

In the BBN, all except two of the arrows are directed from the disease node (parent node) towards the biosecurity nodes (child node). This indicates the direction of inference: the occurrence of 'Disease 1' infers the state of 'Biosecurity practice A', rather than of causality: 'Biosecurity practice A' causes 'Disease 1'. Information can flow in both directions, and mathematically, a model with arrows running in the opposite direction would be equivalent (Fenton and Neil, 2013). There are, however, two reasons for selecting this arrow direction. First, the direction allows us to specify a CPT for each biosecurity node, which is conditional on disease (rather than specify an unconditional probability table for each biosecurity node). Second, the direction allows us to specify an unconditional probability table for the disease node, instead of having to specify a probability table for the disease node that is conditioned on all biosecurity nodes.

Although it may appear more natural to use the direction from cause to effect, in this case it makes sense to use the direction from effect to cause, because the necessary priors are more straightforward to elicit (Fenton and Neil, 2013). This is because we can assess records of disease on farms and then count the proportions that employ a biosecurity practice. In contrast if we sample biosecurity practices, it will be more difficult to then count how many farms become infected with a disease. There were two exceptions: the nodes for region and for farm type were parent nodes of the disease node. This is because they represent a fixed state for each farm and cannot be altered in order to reduce disease occurrence.

2.4.3.1. Specification of CPTs. Conditional probabilities were calculated according to Bayes' theorem using the data about disease occurrence and biosecurity practices on each farm. For example, we calculated $P(\text{a swine farm is within 1 km} | \text{PRRS})$ – i.e. the probability of a neighbouring swine farm being within 1 km, given that a farm has PRRS. Conditional probabilities were input into the CPT of each node. Occasionally, some cells of a CPT were 0 when there was a lack of data (e.g. no cases of a particular disease in a given region) and for these cases a value of 0.01 was specified, according to the accepted Laplace Estimator convention (Witten et al., 2011). While not every cell in a CPT must have a non-zero entry (Marcot et al., 2006), we adopted this practice so that none of the states would have an 'impossible' outcome.

2.4.3.2. *Prior knowledge.* We tested two types of priors for the prevalence of each disease.

- (i) Informative priors were computed conditional on farm type and region. For example, the conditional probability of (PRRS|Farrow to wean farm type), (PRRS|Farrow to finish farm type), and $P(\text{PRRS}|\text{Finishing farm type})$ for each region. We did not receive any questionnaires from Farrow to wean farms in Ontario, and in this case we assigned equal weights to each disease state (uninformative priors). When there were no cases of a disease on a farm type in a region we assigned priors of 0.01, according to the *Laplace Estimator* convention (Witten et al., 2011). For example, Farrow to wean farms in BC reported no cases of any disease, and so we assigned priors of 0.96 for no disease and 0.01 for each of the four diseases.
- (ii) Informative priors were computed conditional on farm type and region and in addition incorporated information about the historical health status of the farm (following Gustafson et al., 1998, 2005). Data about historical health status of the farm came from question 2 of our questionnaire, in which the status of each farm was reported as either naïve, positive, stable, or not tested. We combined reports of positive or stable into one 'positive' class and we excluded any farms that had not been tested. We calculated the conditional probability of a positive or negative health status given the region and farm type. From this we calculated the prior odds of health status, which were then multiplied by the likelihood ratio of disease prevalence to estimate the odds of disease (for each region and farm type). Odds were then converted to normalised conditional probabilities.

The BBN included an on/off switch node. When 'off' the BBN used the informative priors without information about historical health status and we call this model the BBN without history (BBN_{-h}). When switched 'on' the BBN used the informative priors which included historical health status and we call this model BBN with history (BBN_{+h}).

2.4.3.3. *Model output.* When the BBN was 'run' it indicated the 'steady state' scenario between the biosecurity practices and the probability of disease occurrence. Adding evidence to the BBN, by changing the probability of the state of one biosecurity practice, resulted in the BBN revising the states of the other nodes according to Bayes' theorem. Thus changes in the probability for the states at one node were reflected in changes in the probability for the states at other nodes. The relative risk (RR) of each disease was calculated by dividing the probability of disease occurrence (with evidence) by the probability of disease occurrence in the 'steady state'.

2.4.4. Model validation and testing

2.4.4.1. *Model goodness of fit.* The goodness of fit of BBN_{-h} and BBN_{+h} was assessed by inputting information about each farm in turn and comparing the predicted disease status with the observed disease status. The BBN predicted disease status was interpreted in two different ways as follows.

- (i) The disease status of the farm was the disease state with the *highest percentage*. For example, if the disease node indicated 42% PRRS, 35% SI, 7% MP, 6% SD and 10% no disease (ND), then the BBN predicted an occurrence of PRRS. Thus, there could only be one outcome (i.e. prediction) for each farm.
- (ii) It is possible that more than one disease can occur on a farm at the same time. To account for this, the BBN prediction was based on a *threshold value* of disease status. The threshold values explored were 20, 30, 33, 40, 50 and 60%. If the threshold were any of 20, 30 or 33%, then for the scenario described in (i) above, the BBN predicted two diseases: PRRS (42%) and SI (35%).

A threshold of 40% would predict one disease: PRRS. A threshold of 50% would predict no outcome. Note that the number of outcomes changes according to the chosen threshold.

We used the observed and predicted probability of disease occurrence on each farm to calculate the accuracy (percentage of correctly classified outcomes), number of predicted outcomes, sensitivity, positive predictive value (PPV) and f-measure of the BBN. The f-measure, which is a weighted average commonly used to assess classification accuracy = $2 \times ((\text{PPV} \times \text{sensitivity}) / (\text{PPV} + \text{sensitivity}))$ (Powers, 2011). These calculations were made using the full dataset of 90 farms for training, and also using 10-fold cross validation (using 90% of the data for training and 10% for testing) in the data mining software WEKA (Hall et al., 2009).

2.4.4.2. *Model scenario analysis.* The influence of biosecurity practices under different scenarios was assessed using the BBN described previously. Firstly, we assessed how the state of each biosecurity practice independently altered the probability of disease. To do this, the state of each biosecurity practice was altered in turn (i.e. setting one state to 100% at a time) and the probability of disease was recorded. We then calculated the relative risk (RR) of each disease compared to the steady state BBN; for example, a change in the probability of PRRS from the steady state of 15% to 18% equals a RR of 1.2.

Secondly, in order to demonstrate a key strength of the BBN approach, we assessed how combinations of biosecurity practices altered the probability of disease. To do this, we selected a few biosecurity practices that tended to result in the greatest degree of change (increase or decrease) in RR for the various diseases. We then altered the state of those biosecurity nodes simultaneously and recorded the probability of disease.

We compared the BBN approach with the results from a logistic regression using presence or absence of disease as the outcome of interest. It was not possible to complete a multinomial analysis because there was not sufficient data about each disease.

3. Results

3.1. Farm health status and disease occurrence

A total of 20 veterinarians responded to the disease questionnaire, and provided information about 116 farms (96 commercial, 15 breeding and 5 boar farms). Further analysis included 90 of the commercial farms due to small sample sizes for breeding and boar farms. Six commercial farms were excluded because biosecurity information about those farms was not available. The commercial farms were either farrow-to-wean (birth to 3 months of age; n = 18), farrow-to-finish (birth to six months of age; n = 40) or finisher farms (3–6 months of age; n = 32) (Table 1).

Questionnaire responses (question 2) indicated that the health status of the majority of farms was known and that most were naïve for the diseases included (Table 2A). Responses to question 3 of our questionnaire indicated that of the 90 farms, 35 had at least one disease occurrence between January 2010 and December 2011. There were a total of 51 disease occurrences (Table 2B); 22 farms reported one occurrence of disease, 10 farms reported two and three farms reported three. There were no cases of TGE and one case of APP and so these diseases were excluded from all further analysis. Remaining analysis focused on 50 disease occurrences, which were PRRS (32%), SI (42%), MP (18%) and SD (8%).

Table 2A
The health status of 90 Canadian commercial swine farms in 2010 when the CSHB biosecurity survey was conducted.

Disease	Naïve	Positive	Stable	No testing
Porcine Reproductive and Respiratory Syndrome (PRRS)	43	15	25	7
Swine Influenza (SI)	38	12	15	25
Mycoplasma Pneumonia (MP)	34	19	29	8
Swine Dysentery (SD)	58	3	2	27
Transmissible Gastro-Enteritis (TGE)	71	0	0	19
Pleuropneumonia (APP)	62	2	8	18

Definitions: Naïve: the pathogen has been looked for (by observation and testing) but has never been detected on this farm. Positive: The animals on the farm are known to be clinically infected with the pathogen. Stable: The animals are serologically positive (natural or vaccinated), however there are no clinical signs. No testing: No testing has been done.

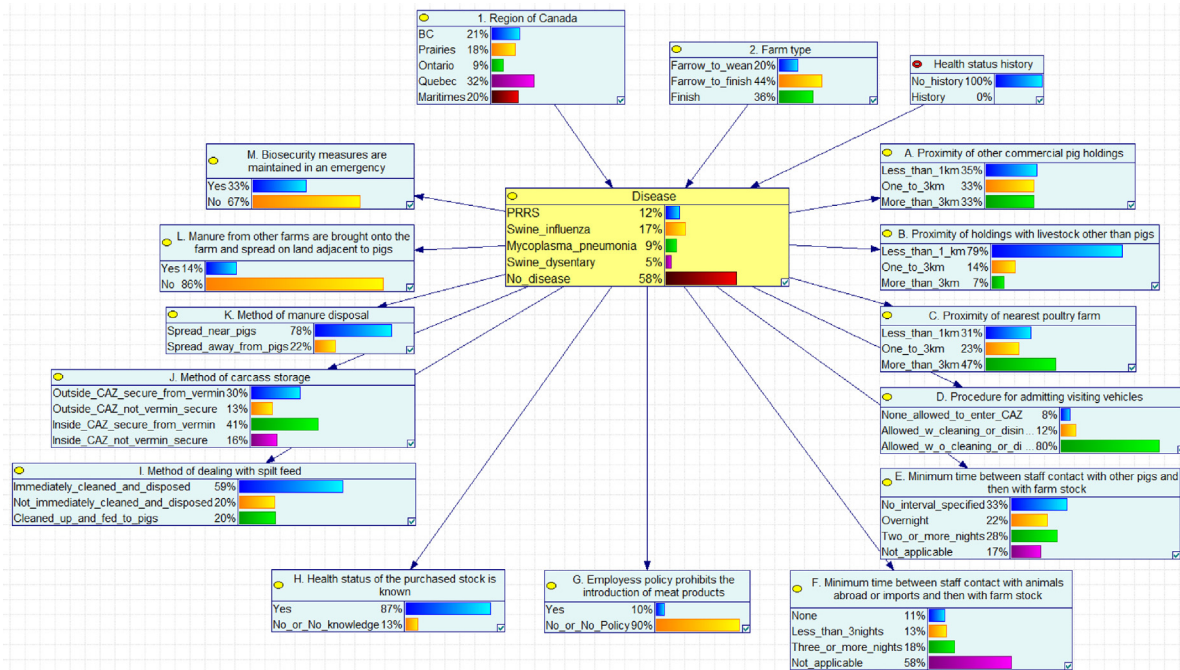


Fig. 1. Bayesian Belief Network (BBN) constructed to examine the probability of occurrence of four diseases on commercial swine farms in Canada. The central disease node (yellow) was connected to all other nodes via conditional probabilities. Within each node, the possible states (left hand side of the node) and the corresponding probability of each state is shown as a percentage and by the coloured bars. The node to switch between ‘no history’ and ‘history’ is shown on the right hand side and in this example is switched ‘off’. This figure represents the “steady state” of BBN_{-h}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2B
The occurrence of disease on 90 commercial swine farms between January 2010 and December 2011.

Disease	Yes	No
Porcine Reproductive and Respiratory Syndrome (PRRS)	16	74
Swine Influenza (SI)	21	69
Mycoplasma Pneumonia (MP)	9	81
Swine Dysentery (SD)	4	86
Transmissible Gastro-Enteritis (TGE)	0	90
Pleuropneumonia (APP)	1	89

3.2. BBN steady state

The steady state of BBN_{-h} was 12% PRRS, 17% SI, 9% MP, 5% SD and 58% ND (Fig. 1). The steady state of BBN_{+h} was 16% PRRS, 16% SI, 12% MP, 2% SD and 54% ND. Inclusion of prior information about the health status of a farm tended to increase the ‘steady state’ probability of PRRS and MP, and decrease the probability of SD and ND, while having little influence on SI.

Preliminary assessment of the data suggested that the region node of the BBN might have a misleading influence on the BBN predictions. This is because the majority of swine farms included in the biosecurity survey were situated in the Prairies (n = 71, 33%)

and Ontario (n=52, 24%), but we only received disease status information from 16 (23%) and 8 (15%) farms in these provinces respectively. In contrast, we received information from 95% of the 20 farms in BC and so the disease prevalence data are heavily skewed by these inputs. We decided to re-assess the model by excluding the region node. We now use the subscripts BBN_{+r} and BBN_{-r} when region was included or excluded respectively. The probability of disease in the steady state of BBN_{-h-r} was 15% PRRS, 20% SI, 8% MP, 4% SD and 53% ND. The steady state of BBN_{+h-r} was 17% PRRS, 18% SI, 12% MP, 1% SD and 52% ND. In comparison, the disease data received for each type of farm appeared to be a representative sample of the farms that were included in the biosecurity survey: we received responses from 30% of farrow-to-wean farms, 48% of farrow-to-finish farms and 43% of finisher farms. The farm type node thus remained in the BBN.

3.3. BBN goodness of fit

We compared the fit of BBN_{-h+r}, BBN_{-h-r}, BBN_{+h+r} and BBN_{+h-r}. Comparison of model predictions that were based on the state with the highest percentage in the disease node, showed that the most accurate models were BBN_{-h+r} and BBN_{-h-r} (Fig. 2). These models generated an accuracy of 67% for BBN_{-h+r} and 65% for

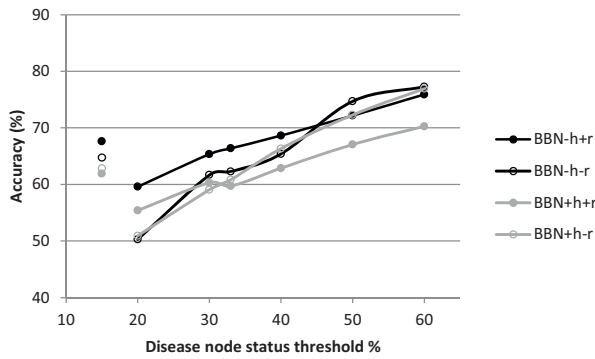


Fig. 2. Accuracy of the BBN (percentage of correctly classified outcomes) at increasing values of disease node status threshold. The BBN predicted outcome (i.e. disease occurrence) was the disease status that exceeded the threshold value. Data points that are not joined by the line show the accuracy when the BBN prediction was the disease status with the highest percentage.

BBN_{-h-r} using the full dataset. Ten-fold cross validation resulted in a more honest estimate of 48% of occurrences correctly classified using BBN_{-h+r} and 50% using BBN_{-h-r}.

When we used a threshold value to make predictions (based on the entire dataset), model accuracy increased as the threshold value increased (Fig. 2). At a low threshold (20%) the models predicted between 146 and 163 outcomes (Fig. 3) with an accuracy of between 50 and 60%. A threshold of 50% generated an accuracy of between 67 and 74% and predicted between 83 and 97 outcomes. BBN_{-h+r} was the most accurate up to the threshold of 40%, while BBN_{-h-r} was the most accurate at 50% (and above), with 75% of occurrences correctly classified, although it only predicted a total of 83 outcomes (the full dataset contained 105 outcomes).

The models with the greatest f-measures (based on the state with the highest percentage in the disease node) were BBN_{-h+r} and BBN_{-h-r} (Fig. 4). The f-measure of each model increased with increasing disease node status threshold (Fig. 4). The model with the highest f-measure tended to be BBN_{-h+r}. At the 50% threshold, the f measure of BBN_{-h+r} and BBN_{-h-r} did not vary considerably (0.706 and 0.700 respectively).

The model that was most accurate, predicted the number of outcomes and had the greatest f-measure varied according to the threshold value. However BBN_{-h+r} and BBN_{-h-r} tended to be the optimum models according to the goodness of fit results. In light of these results, as well as the bias in the regional data that we described in the previous section, we chose BBN_{-h-r} as the most

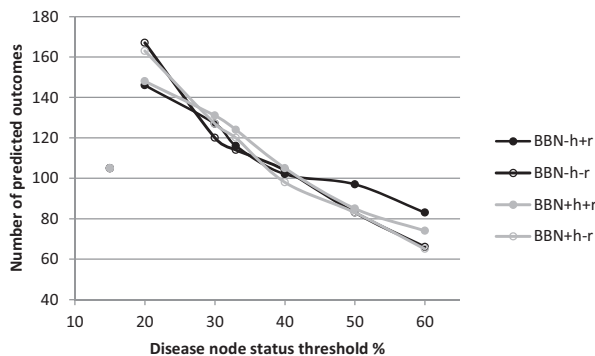


Fig. 3. BBN predicted number of outcomes at increasing values of disease node status threshold. The BBN predicted disease outcome was the disease status that exceeded the threshold value. Data points that are not joined by the line, show the number of outcomes, when the BBN prediction was the disease status with the highest percentage. (In this case all models predicted one outcome per scenario and therefore predicted 105 outcomes in total, because there were 105 outcomes in the real dataset).

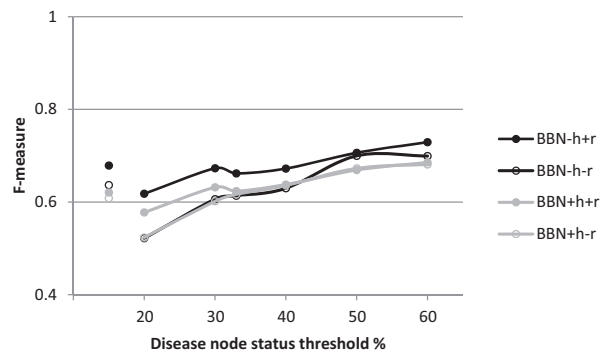


Fig. 4. F-measure of the BBN at increasing values of disease node status threshold. The BBN predicted disease outcome was the disease status that exceeded the threshold value. Data points that are not joined by the line, show the f-measure when the BBN prediction was the disease status with the highest percentage.

parsimonious representation of the data. We use this model in all further analysis.

3.3.1. Logistic regression

The biosecurity practices that were associated ($p < 0.1$) with presence of disease were: Proximity of other commercial pig holdings, Procedure for admitting visiting vehicles and Method of dealing with spilt feed. The logistic regression model was statistically significant, ($\chi^2 = 23.4$, $df = 6$, $p < 0.001$; The Hosmer Lemeshow Goodness of fit test indicated that the model fit the data ($\chi^2 = 1.91$, $p = 0.96$). Farms were less likely to have a disease outbreak when the distance to the nearest neighbouring pig holding was more than 3 km. Farms within 1–3 km of a neighbour were 4 times more likely to have disease. Farms were less likely to have disease when visiting vehicles were allowed to enter the peripheral zone (compared to not allowed). They were also less likely to have disease when spilt feed was cleaned up and fed to the pigs, rather than cleaned up and disposed of immediately or not immediately). The accuracy (70%) was slightly greater than that of BBN_{-h-r} (65%), while the f-measures were similar (0.630 and 0.637 respectively).

3.3.2. BBN scenario analysis: influence of biosecurity practices independently on probability of disease

The BBN_{-h-r} was explored by changing the state of each node (set to 100%) in turn and recording the probability of disease (Table 3). For instance, in the steady state, the probability of disease was 15% PRRS, 20% SI, 8% MP, 4% SD and 53% ND; while if the nearest pig farm was within 1 km then the probability of disease became 18% PRRS, 31% SI, 8% MP, 3% SD and 40% ND. We express this in Table 3 as the RR compared to the steady state of each disease. In the table we highlight RR of 1.5 or more and 0.5 or less to demonstrate the biosecurity practices that had the most influence on disease probability and we focus the description of results on these changes.

Compared to the steady state model, a farrow-to-wean farm had an increased probability of PRRS, while a farrow-to-finish farm had a decreased probability of PRRS. A finishing farm had an increased probability of MP.

The proximity to a neighbouring commercial pig farm influenced the probability of disease. The probability of SI increased (RR 1.55) if there was another pig holding within 1 km. The RR of SD was 1.75 if there was a pig holding or another livestock holding within 1–3 km, while the RR decreased to 0.75 if the distance was beyond 3 km. The probability of PRRS decreased when the nearest pig holding or livestock holding was 1–3 km away (RR 0.47), or more than 3 km away (RR 0.60). The RR of ND was 1.32 if the nearest pig holding was more than 3 km away.

Table 3
The steady state of BBN_{p-r} and the impact of each biosecurity practice on the relative risk of each disease. The first row shows the probability of each disease and of no disease when no evidence has been entered into the model. Entries in each row show the RR of disease compared to the steady state model if that state is selected. An increase in disease probability with a RR of more than 1.5 is highlighted in red; a decreased disease probability with a RR of less than 0.5 is highlighted in blue.

Risk factor	Disease				
	PRRS	SI	MP	SD	ND
No Evidence: steady state	15%	20%	8%	4%	53%
Farm type					
Farrow-to-wean	1.73	1.30	1.13	1.00	0.66
Farrow-to-finish	0.47	0.95	0.63	1.25	1.23
Finisher	1.20	0.90	1.63	0.75	0.92
Proximity of other commercial pig holdings					
Nearest pig holding < 1km	1.20	1.55	1.00	0.75	0.75
Nearest pig holding 1-3km	0.47	0.70	0.88	1.75	1.23
Nearest pig holding >3km	0.60	0.60	0.75	0.75	1.32
Proximity of other holdings with livestock other than pigs (including slaughterhouses, hobby farms and sale barns)					
Nearest other livestock holding <1km	1.00	1.10	1.00	1.00	0.96
Nearest other livestock holding 1-3km	0.47	0.70	0.88	1.75	1.23
Nearest other livestock holding >3km	1.60	0.60	1.50	0.25	0.96
Proximity of the nearest poultry farm					
Nearest poultry farm < 1km	0.80	1.10	0.75	0.75	1.08
Nearest poultry farm 1-3 km	1.60	0.60	1.00	1.00	0.96
Nearest poultry farm >3km	0.80	1.10	1.25	1.00	0.98
Procedure for admitting visiting vehicles					
Visiting vehicle access allowed without cleaning and/or disinfection.	0.93	0.80	1.13	1.00	1.09
Visiting vehicle access allowed with cleaning and/or disinfection.	0.47	1.90	0.13	2.00	0.87
None allowed to enter the peripheral biosecure area (CAZ).	2.13	1.65	1.38	0.00	0.43
Minimum specified time interval that must elapse between farm staff having contact with other pigs and then with farm stock					
No interval specified	0.93	1.00	1.00	0.00	1.09
Overnight	0.60	0.90	0.50	2.25	1.13
2 or more nights	1.13	0.85	1.25	0.75	1.00
Not applicable	1.33	1.30	1.25	1.25	0.72
Minimum specified time interval that must elapse between farm staff having contact with animals abroad / imported animals and then with farm stock					
None	0.07	0.85	0.13	2.25	1.36
<3 nights	0.53	0.10	1.00	2.00	1.40
3 or more nights	0.07	2.05	0.63	1.25	0.91
Not applicable	1.53	0.85	1.38	0.50	0.89
Employee policy prohibits the introduction of meat (cooked or uncooked) products					
No or No policy	1.07	0.90	1.00	0.50	1.06
Yes	0.13	1.95	1.25	5.00	0.57
Health status of purchased stock is known*					
No or Vague	0.47	0.75	0.88	3.75	1.02
Yes	1.07	1.05	1.00	0.50	1.00
Method of dealing with spilt feed*					
Cleaned up and fed to pigs	0.33	0.25	0.63	0.00	1.60
Not cleaned up and not disposed of immediately	2.00	1.10	2.13	0.00	0.58
Cleaned up and disposed of immediately	0.80	1.20	0.75	1.50	0.96
Method of carcass storage					
Inside the CAZ, not secure from vermin, scavengers and pests.	1.20	0.30	1.50	1.50	1.08
Inside the CAZ, secure from vermin, scavengers and pests.	1.27	1.35	0.88	0.00	0.91
Outside of the CAZ, not secure from vermin, scavengers and pests	0.47	0.75	0.88	2.00	1.17
Outside of the CAZ, secure from vermin, scavengers and pests.	0.60	0.95	1.13	1.50	1.06
Method of manure disposal*					
Spread (treated or untreated) adjacent to pig barns	0.87	1.05	0.88	1.00	1.06
Spread (treated or untreated) away from pig barns.	1.40	0.85	1.63	1.00	0.83
Manure or organic waste from other farms are brought onto the farm and spread on land adjacent to the pigs*					
Yes	2.47	0.65	0.75	1.50	0.72
No	0.73	1.05	1.13	0.75	1.06
Biosecurity measures are maintained in an emergency					
No	1.13	0.80	1.25	0.75	1.04
Yes	0.73	1.40	0.75	1.50	0.92

*These biosecurity practices were used in our demonstration of how the BBN can be employed to assess a combination of practices on the probability of disease.

In some instances the RR of disease increased when it might be expected that there would be a decrease; we review the reasons for such counter-intuitive results in our discussion. For example, it was found that when visiting vehicles were not allowed to enter the peripheral biosecure area (CAZ) the probability of SD and of ND decreased (RR of 0 and 0.43 respectively); however, an unex-

pected result was that PRRS, and SI tended to increase (RR 2.13 and 1.65 respectively). Furthermore, if visiting vehicles were allowed after cleaning and disinfection then the probability of SI and SD increased, and the RR of PRRS and MP decreased, while there was a small decrease in ND.

When there was no minimum specified time interval between farm staff having contact with other pigs and then with farm stock, the probability of all diseases and ND did not change. The only exception was that SD decreased to 0. However, if the interval was overnight then the RR of SD was 2.25 (and the probability of other diseases tended to decrease a little). Note that the states of this node did not alter the probability of ND considerably.

When the health status of purchased stock was unknown (as opposed to known), the RR of SD increased by 3.75. In contrast, the RR of PRRS was 0.47. The RR of other disease and of ND didn't tend to change considerably according to this biosecurity practice.

When spilt feed was not cleaned up and disposed of immediately the RR of PRRS and MP was 2 and 2.13 respectively. If spilt feed was cleaned up and fed to the pigs then there was a reduction in the probability of all four diseases, while the probability of ND increased (RR of 1.6). There were 18 farms that fed spilt feed to the pigs, although only three reported occurrence of disease (PRRS, SI and MP).

Storage of carcasses for disposal outside of the CAZ generated a decreased probability of PRRS (RR 0.47). Probability of ND was higher if stored outside of the CAZ than inside. Within the CAZ, storage of carcasses in an unsecure place was a lower risk than if stored in a secure place for SI. However, note that this surprising result is based on a small sample size of only 14 farms (5 occurrences of disease) on which the carcasses were stored in an unsecure area.

If manure or organic waste was brought onto the farm and spread adjacent to the pig barns then the RR of PRRS was 2.47 and of SD was 1.5. Other diseases did not alter considerably, although the RR of ND was 0.72.

3.3.3. BBN scenario analysis: influence of biosecurity practices in combination on probability of disease

A few biosecurity practices that tended to result in the greatest degree of change (increase or decrease) in RR for the various diseases were selected, namely: Health status of purchased stock; Method of dealing with spilt feed; Method of manure disposal; Whether manure or organic waste was brought from other farms onto the farm or spread on land adjacent to the pigs. We generated a few scenarios that represent combinations of these biosecurity practices (Table 4) and we use the model to illustrate how the probability of disease varies according to each scenario (Fig. 5). Note that these scenarios illustrate the utility of the BBN and that if a detailed dataset were available it would be possible to have confidence in more detailed predictions.

For example, under Scenario 1, when the health status of purchased stock is not known, feed is cleaned up and fed to pigs, manure is spread adjacent to barns and manure is brought onto the farm, the probability of PRRS increases from 15 to 31% (RR of 2.1), MP decreases from 8 to 4% (RR of 0.5) and ND decreases from 53% to 41% (RR 0.77).

Alternative strategies (2 and 3) also decrease the probability of ND. Under Scenario 2 the probability of PRRS and MP increased (to 31% (RR 2.1) and 25% (RR 3.2) respectively), while there was little change in the probability of SI and a decrease in the probability of SD. When this scenario was altered to allow manure from other farms to be spread on land adjacent to the pigs, the greatest risks for disease were PRRS, which increased to 69% (RR 4.6) and MP which increased to 12% (RR 1.5).

Scenario 4 could well be considered to be a low risk strategy since the BBN model predicted that the probability of ND increased to 85% (RR 1.6), while the probability of PRRS, SI and SD all decreased (to 3% (RR 0.2), 4% (RR 0.2) and 1% (RR 0.25) respectively).

We also investigated the predictions based on Scenario 1 and Scenario 4 (described above) for the three different farm types (Fig. 6). In general, farrow-to-finish farms were most likely to have ND, while farrow-to-wean farms tended to have the highest proba-

bility of disease. The probability of PRRS and SD tended to increase in Scenario 1, while the probability of SI and MP tended to decrease on each farm type. Under Scenario 4 the probability of ND increased for each farm type. On farrow-to-wean farms it increased to 74% (RR 2.11), on farrow-to-finish farms to 91% (RR 1.4) and on finisher farms to 80% (RR 1.63). The probability of all diseases decreased for this scenario.

4. Discussion

4.1. BBN model predictions

We illustrated the utility of a naïve BBN for the Canadian swine industry in the assessment of management practices on the probability of disease occurrence. Results indicated that farm type influenced the probability of disease, with farrow-to-wean farms most likely to experience disease occurrences. Although we excluded information about region from our model, due to limitations in the data, it is likely that region has a considerable influence on disease occurrence since farm type and management strategies vary between regions. For example, the size of farm changes according to region, with Manitoba, Quebec and Saskatchewan having the largest herds and BC and the Atlantic provinces the smallest (Brisson, 2014). Unfortunately we were unable to consider farm size in our analysis because that data was not available. Differences in management strategies include movement of animals between provinces: Alberta, Saskatchewan and Ontario export pigs to other provinces for slaughter, while BC, Manitoba and Quebec import animals from other provinces for slaughter (University of Guelph, 2010).

We included four diseases in the BBN, which were representative of different disease types. In general, there are two main types of swine pathogens; those that are introduced mainly through direct pig contacts (where animals that are sub-clinically infected with a pathogen are unknowingly introduced into a previously uninfected herd); and those that are introduced through indirect contact (pathogens that find their way into herds via routes other than animals) (Desrosiers, 2011). In general, PRRS and MP can transmit easily by indirect contact, while SD and SI tend to be transmitted primarily by direct contact. The aim of our work was to demonstrate the value of a BBN tool, we therefore briefly discuss how our findings compare to the transmission dynamics of the diseases studied.

PRRS virus can spread by direct or indirect contact although the majority of cases in Canada (81–100%) are associated with indirect transmission (Laroche et al., 2003; Desrosiers, 2004). The virus can be detected in saliva, urine, milk and faeces; routes of indirect transmission include via fomites (boots, overalls, etc.), farm personnel, transport vehicles and insects. Significant risk factors for spread between farms include purchase of animals from herds incubating infection (World Organisation for Animal Health, 2008) and proximity to infected neighbouring herds, particularly because transmission can occur via aerosol (Otake et al., 2010). Biosecurity protocols to reduce the probability of PRRS include sanitation of vehicles, and implementation of strategies for personnel/fomite entry (World Organisation for Animal Health, 2008), however control programs are frequently unsuccessful at preventing virus introduction (Desrosiers, 2011). In our BBN_{h-r}, the greatest risk of PRRS was related to when manure was brought onto the farm and spread on land adjacent to the pigs. Our model also showed that when the distance to neighbouring pig farms was more than 1 km, the probability of PRRS decreased compared to being within 1 km. This is in agreement with studies that have shown that herds located within 1.5 km of a neighbouring pig farm were more likely to be PRRS-positive (Lambert et al., 2010). The slightly

Table 4
Four different scenarios, which represent combinations of biosecurity practices. Scenarios were simulated in the BBN and the resulting probability of disease for each scenario presented in Fig. 5.

Biosecurity practices	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Health status of purchased stock is known	No or vague	No or vague	No or vague	Yes
Method of dealing with spilt feed	Cleaned up and fed to pigs	Not cleaned up and not disposed of immediately	Not cleaned up and not disposed of immediately	Cleaned up and disposed of immediately
Method of manure disposal	Spread (treated or untreated) adjacent to pig barns	Spread (treated or untreated) away from pig barns	Spread (treated or untreated) away from pig barns	Spread (treated or untreated) away from pig barns
Manure or organic waste from other farms are brought onto the farm and spread on land adjacent to the pigs	Yes	No	Yes	No

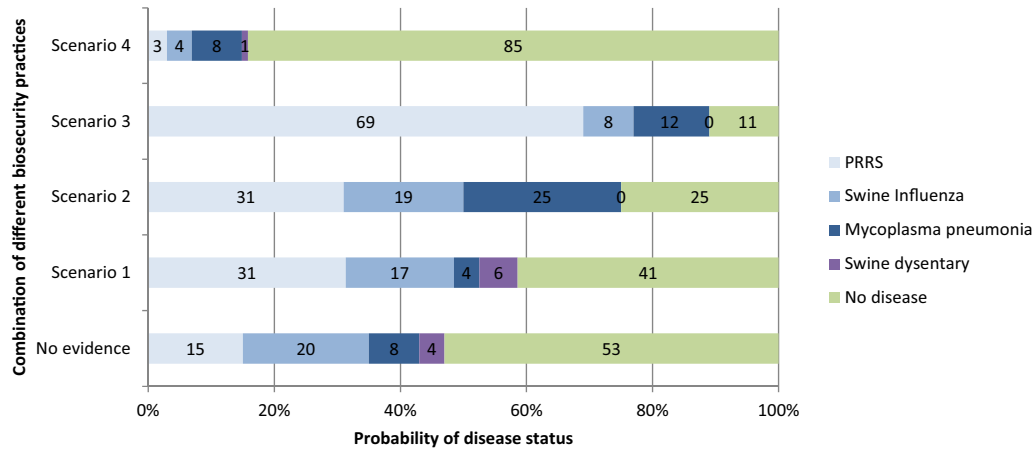


Fig. 5. Assessment of a combination of biosecurity practices on the probability of disease occurrence. The first line shows the disease state without evidence (i.e. 'steady state'). Other lines show the disease state under four different scenarios, which represent a different combination of biosecurity practices, as described in Table 4.

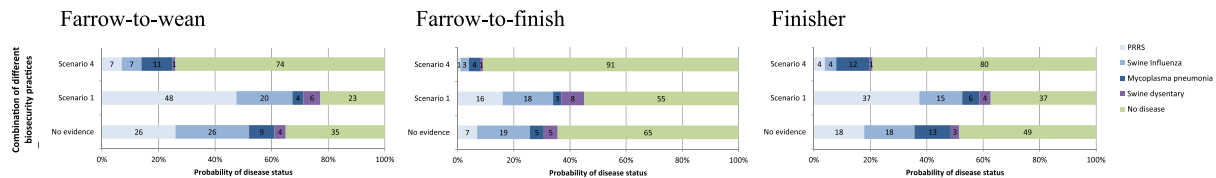


Fig. 6. Assessment of a combination of biosecurity practices on the probability of disease occurrence on three different swine farm types. The first line shows the disease state without evidence. Other lines show the disease state under two different scenarios (as described in Table 4).

higher RR at a further distance (RR 1–3 km = 0.47, RR >3 km = 0.6) is likely due to the small sample size of farms that had PRRS within each distance category.

SI is highly contagious and was the most common cause of disease reported in our survey (21 occurrences). The virus can be introduced into a herd by infected pigs, humans, and other species (World Health Organisation, 2010). Biosecurity practices highlighted by our model included allowing vehicle access without cleaning or disinfecting and other pig holdings situated within 1 km.

SD incidence in North America was declining following eradication and effective manure removal techniques, however, in 2009 it re-emerged in Western Canada (Harding et al., 2013) and approximately 50% of farms became infected through contaminated pigs. Other specific security breaches are poorly understood, although exposure to contaminated faeces or manure (Harding et al., 2013) and wildlife (Desrosiers, 2011; Harding et al., 2013) are likely mechanisms. Increased probability of SD predicted in our model tended to occur when vehicles were allowed access to the CAZ, when farm staff had contacted other herds in the past 3 days, when the health

status of purchased stock was not known and when manure was brought onto the farm and spread on land adjacent to the pig barns. SD tends to be a disease of growing pigs and this may be one reason why our model indicated an increased risk on farrow-to-finish farms compared to finisher farms (note these predictions were based on a limited number of SD cases).

MP is not well suited to indirect transmission (Desrosiers, 2011) and there are no reports of this disease occurring in species other than swine. Aerosol transmission is likely the main driver of cases that are not caused by direct contact (Desrosiers, 2002). Proximity of other pig farms is therefore, a risk factor, with the MP pathogen, having been identified in air samples collected more than 9 km from their source (Otake et al., 2010). Of the nine cases of MP reported in our survey, all occurred on farms that were within 10 km of another swine farm, while none occurred on the 14 farms whose nearest neighbour was more than 10 km away. MP can affect pigs early after weaning, but more commonly occurs in grower and finisher stages. This is likely why there was a higher probability of occurrence on finisher farms in the BBN_{h-r} model.

4.2. Model design

Industry-wide quantitative data collected directly from Canadian swine farms was used to specify the CPT of each node in the BBN. This is a unique feature of our study because it is rare that BBN design is based on directly collected data. Indeed, calculation of the CPT for each node has been described as ‘usually the hardest part of the modelling process’ (Fenton and Neil, 2013) and in most cases published literature or expert opinion are relied upon to approximate CPTs (Newton et al., 2007). Although some BBNs have been built by calculating conditional probabilities from published odds ratios for risk factors (Otto and Kristensen, 2004), often there is insufficient information available for such estimation.

We selected the most important biosecurity factors based on expert judgement. While there may be other factors that are relevant to specific diseases or regions of Canada, for instance, we believe that using those identified by a range of experts provided a solid starting point. The node relating to region of Canada was removed because of the potential biases introduced by level of reporting. In particular, there was an under-representation of farms in Ontario (none reporting any disease) and an over-representation of farms in BC.

4.3. Sensitivity, PPV and f-measure

We assessed the performance of the BBN by calculating the accuracy, sensitivity, PPV and f-measure for models that excluded or included information about prior health status and geographic region. We also defined the disease prediction of the model using different threshold values to account for multiple diseases. The BBN had an accuracy of 65%; a value that is within the range reported for other naïve BBNs that were designed for disease diagnosis and is considered of comparable performance to the optimally efficient diagnostic rule on available data (Geenen et al., 2011). In general, choosing a threshold can depend on the use of the model and the implications of false positive and/or false negative errors (Marcot et al., 2006). Prior information about the health status of the farm did not improve the fit of the model. One reason may be that a farm with a history of disease occurrence may have altered its biosecurity practices in response to that outbreak. This temporal misalignment between collection of biosecurity data and disease outbreak data is one limitation of our study, (indeed any retrospective cross-sectional study), which we are unable to account for. Since it is possible that diseased farms changed their biosecurity practices, misclassification of exposure is likely to be differential. This may explain some of the contradictory results, although a more robust dataset is required to estimate the magnitude of this effect.

4.4. BBN structure and correlation of biosecurity practices

Uncertainty is a common feature associated with emerging disease events. BBNs are particularly advantageous in these situations because if the state of one node is unknown, then the BBN can still be used to calculate a probability dependent on the known states of other nodes – i.e. based on the evidence available. This is an important point because although four diseases were included in the BBN, the evidence for SD was only based on 4 outbreaks. Despite this, there are improvements that could be made to the model, particularly if a dataset containing more farms was available. More data would allow us to make stronger inference about interactions between biosecurity practices, and would give more confidence in predicting which scenarios reduce disease occurrence.

Lack of data may have caused some of the counter-intuitive results that we observed. For example, we noted that the probability of SD was 0 if there was no minimum specified time interval between farm staff contact with other pigs and then with farm

stock, but was greater (RR was 2.15) if the time interval was overnight. This result occurred because the farms that have no specified time interval between contacts did not report any cases of SD. Similarly when there was no specified time interval between farm staff contact with animals abroad and animals on the farm, the probability of PRRS and MP decreased (compared to a time interval of one night or more). When the time interval was more than three nights then the probability of SI increased. These results reflect the small sample size for these categories. There were 11 farms that employed the ‘no specified time interval’ policy and none reported a disease. Another noteworthy counter-intuitive result relates to vehicle access to the CAZ. When vehicle access to the CAZ was not allowed, there was an increased probability of PRRS and SI, compared to when vehicles were allowed to enter without cleaning or disinfecting. We suggest that the influence of this node is likely biased by the fact that the majority of farms included in the survey (n = 70) allowed vehicle access without cleaning or disinfecting.

Correlation or confounding of biosecurity practices (included in the model or not) could contribute to the counter-intuitive results described above. Although we did not formally test for correlation between nodes, it is not unreasonable to suggest that some biosecurity practices will be dependent on one another. Confounding factors that we were not able to consider include type of visiting vehicles (e.g. feed, manure or waste trucks) and methods of pig transport (e.g. methods of loading pigs, clean up around loading, method of load out). Furthermore, the frequency of each event is highly relevant. For example, movement of pigs on to a finisher farm is more frequent than onto other farm types, while movement of pigs off farrowing farms is more frequent than other farm types (Thakur et al., 2014). It is for this reason that nursery farms have been highlighted as targets for disease surveillance (Dorjee et al., 2013).

Here we chose to create a BBN using a naïve structure as a first step in the analysis of the relatively complex data. Such a structure is excellent for integrating information while maintaining a model of limited size (Gustafson et al., 2005; Needham et al., 2007). Although, BBN’s attempt to ensure conditional independence between nodes, we did not assess dependence of nodes for this simple model. Despite the underlying independence assumption in a naïve BBN, they have proven to be powerful probabilistic models for solving classification problems in a variety of domains (Geenen et al., 2011) and have been shown to outperform more sophisticated models even for classification problems in which the assumption does not hold (Domingos and Pazzani, 1997). Furthermore, in comparison to classical statistical analysis, our approach has a number of advantages. The logistic regression only allowed us to assess risk factors for presence or absence of disease, while the BBN allowed us to assess the occurrence of four different diseases and the impact of biosecurity practices simultaneously. Further, we were able to explicitly model causal factors, and we demonstrated how the tool can be used to arrive at decisions with visible, auditable reasoning.

Building realistic and accurate BBNs with a proper graphical structure and all the required CPTs remains a major challenge (Zhou et al., 2014). Ongoing analysis (not presented here) aims to account for dependence among nodes in order to tease apart the complex interactions of biosecurity practices. This includes revising the BBN by creating influence diagrams of the causal web (e.g. Marcot et al., 2006). We have also been using machine learning to explore network structures and to improve model fit (Cox et al., 2015).

4.5. Potential use of the swine BBN

While there are some limitations to the current BBN, we have presented a method that is robust and we believe that anomalous results would likely be resolved by the collection of additional data.

Our BBN provides a visual tool that allowed us to carry out an interactive assessment of biosecurity practices on the RR of disease occurrence. Importantly, such a BBN can be readily updated as new evidence becomes available, which means that it provides a ready-made platform which can be modified as new threats emerge and/or working practices in the Canadian pork industry are revised.

Demonstrating the possible means by which a given pathogen can be introduced into swine herds is a necessary step in understanding its epidemiology, while being able to weigh the significance of each of these means is crucial (Desrosiers, 2011). Thus we suggest a number of end users of a BBN model. On a small scale, this type of tool could be used by industry and vets within a province to create farm-specific risk profiles. On a wider scale, it could contribute to improving industry biosecurity practices nationally. At present, surveillance for swine endemic diseases occurs at a provincial level in Canada. An improvement to this system could be collation of farm data, along with laboratory and slaughterhouse data, by a coordinating body such as the CSHB. Creation of such a veterinary network and collation of syndromic data would enable the Canadian industry's mission of re-focusing efforts in animal health on a national basis. If such data existed then a BBN model could become a fundamental tool for identifying and monitoring pathogen threats.

Acknowledgements

The Canadian Swine Health Board provided funding for this work and allowed access to the biosecurity data. Lucie Verdon helped to design and administer the questionnaire. We thank the veterinarians and farmers for participation in the survey. We also thank William Chalmers for technical assistance with the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.prevetmed.2016.06.015>.

References

- Ben-Gal, I., 2007. Bayesian networks. In: Ruggeria, F., Faltin, F., Kennet, R. (Eds.), *Encyclopedia of Statistics in Quality and Reliability*. Wiley & SonsUK.
- Brisson, Y., 2014. The Changing Face of the Canadian Hog Industry. <http://www.statcan.gc.ca/pub/96-325-x/2014001/article/14027-eng.htm>.
- Canadian Swine Health Board, 2010. National Biosecurity Benchmarking Survey final project report. December 2010. Confidential Report to the Canadian Pork Council.
- Cox, R., Sanchez, J., Revie, C.W., 2015. Development of a Bayesian Belief Network by machine-learning as a tool for evidence-based management of disease emergence on swine farms in Canada. In: 14th International Symposium on Veterinary Epidemiology and Economics, Merida, Mexico.
- Desrosiers, R., 2002. Aerosol transmission of *Mycoplasma hyopneumoniae* in North America. *Int. Piglett.* 21, 68–70.
- Desrosiers, R., 2004. Epidemiology, diagnosis and control of swine diseases. *Proc. Am. Assoc. Swine. Vets.* 12, 1–30.
- Desrosiers, R., 2011. Transmission of swine pathogens: different means, different needs. *Anim. Health. Res. Rev.* 12, 1–13.
- Domingos, P., Pazzani, M., 1997. On the optimality of the simple Bayesian network classifier. *Mach. Learn.* 29, 103–130.
- Dorjee, S., Revie, C.W., Poljak, Z., McNab, W.B., Sanchez, J., 2013. Network analysis of swine shipments in Ontario, Canada, to support disease spread modelling and risk-based disease management. *Prev. Vet. Med.* 112, 118–127.
- Fenton, N., Neil, M., 2013. Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press, Taylor & Francis Group Boca Raton, USA.
- Firestone, S.M., Lewis, F.I., Schemann, K., Ward, M.P., Toribio, J.A., Taylor, M.R., Dhand, N.K., 2014. Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers' perceptions of a subsequent outbreak. *Prev. Vet. Med.* 116, 243–251.
- Fluidsurveys, 2013. Fluidsurveys. www.fluidsurveys.com.
- Geenen, P.L., van der Gaag, L.C., Loeffen, W.L., Elbers, A.R., 2011. Constructing naive Bayesian classifiers for veterinary medicine: a case study in the clinical diagnosis of classical swine fever. *Res. Vet. Sci.* 91, 64–70.
- Gustafson, L.L., Franklin, W.L., Sarno, R.J., Hunter, R.L., Young, K.M., Johnson, W.E., Behl, M.J., 1998. Predicting early mortality of newborn guanacos by birth mass and hematological parameters: a provisional model. *J. Wildl. Manage.* 62, 24–35.
- Gustafson, L.L., Ellis, S.K., Bartlett, C.A., 2005. Using expert opinion to identify risk factors important to infectious salmon-anemia (ISA) outbreaks on salmon farms in Maine, USA and New Brunswick, Canada. *Prev. Vet. Med.* 70, 17–28.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. *SIGKDD Explor.*
- Harding, J., Fernando, C., Costa, M., Patterson, A., Rubin, J., Hill, E.E., 2013. What have we learned about swine dysentery in Canada? *Adv. Pork Prod.* 24, 175–179.
- Jensen, F.V., 2001. *Bayesian Networks and Decision Graphs*. Springer, Verlag, New York.
- Lambert, M.E., Poljak, Z., D'Allaire, S., 2010. Investigation of factors associated with Porcine Reproductive and Respiratory Syndrome (PRRS) herd status for production sites in a low density area of swine production in Quebec, Canada. *Proc. Int. Pig. Vet. Soc.* 1, 1–150.
- Larochelle, R., D'Allaire, S., Magar, R., 2003. Molecular epidemiology of Porcine reproductive and respiratory syndrome virus (PRRSV) in Quebec. *Virus Res.* 96, 3–14.
- Marcot, B.G., Steventon, J.D., Sutherland, G.D., McCann, R.K., 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Can. J. For. Res.* 36, 3063–3074.
- Marvin, D.M., Dewey, C.E., Rajic, A., Poljak, Z., Young, B., 2010. Knowledge of zoonoses among those affiliated with the Ontario Swine industry: a questionnaire administered to selected producers, allied personnel, and veterinarians. *Foodborne Pathog. Dis.* 7, 159–166.
- McCormick, B.J., Sanchez-Vazquez, M.J., Lewis, F.I., 2013. Using Bayesian networks to explore the role of weather as a potential determinant of disease in pigs. *Prev. Vet. Med.* 110, 54–63.
- McKendrick, I.J., Gettinby, G., Gu, Y., Reid, S.W.J., Revie, C.W., 2000. Using Bayesian Belief Networks to aid differential diagnosis of tropical bovine diseases. *Prev. Vet. Med.* 47, 141–156.
- Needham, C.J., Bradford, J.R., Bulpitt, A.J., Westhead, D.R., 2007. A primer on learning in bayesian networks for computational biology. *PLoS Comp. Biol.* 3, 1409–1416.
- Neumann, E.J., Kliebenstein, J.B., Johnson, C.D., Mabry, J.W., Bush, E.J., Seitzinger, A.H., Green, A.L., Zimmerman, J.J., 2005. Assessment of the economic impact of Porcine Reproductive and Respiratory Syndrome on swine production in the United States. *J. Am. Vet. Med. Assoc.* 227, 385–392.
- Newton, A.C., Stewart, G.B., Diaz, A., Golicer, D.P.A.S., 2007. Networks as a tool for evidence-based conservation management. *J. Nat. Conserv.* 15, 144–160.
- O'Hagan, A., Buck, C.E., Daneshkhal, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain judgements. In: *Eliciting Experts' Probabilities*. John Wiley & Sons, Ltd., Chichester, UK.
- Otake, S., Dee, S., Corzo, C., Oliveira, S., Deen, J., 2010. Long-distance airborne transport of infectious PRRSV and *Mycoplasma hyopneumoniae* from a swine population infected with multiple viral variants. *Vet. Microbiol.* 145, 198–208.
- Otto, L., Kristensen, C.S., 2004. A biological network describing infection with *Mycoplasma hyopneumoniae* in swine herds. *Prev. Vet. Med.* 66, 141–161.
- Paarlberg, P.L., 2014. Updated Estimated Economic Welfare Impacts of Porcine Epidemic Diarrhea Virus (PEDV). University of Purdue, USA, pp. 1–38.
- Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* 2, 37–63.
- Seidel, M., Breslin, C., Christley, R.M., Gettinby, G., Reid, S.W.J., Revie, C.W., 2003. Comparing diagnoses from expert systems and human experts. *Agric. Syst.* 76, 527–538.
- Thakur, K.K., Revie, C.W., Hurnik, D., Poljak, Z., Sanchez, J., 2014. Analysis of swine movement in four Canadian regions: network structure and implications for disease spread. *Transbound Emerg. Dis.*, <http://dx.doi.org/10.1111/tbed.12225>.
- University of Guelph, 2010. Hog Processing in Canada. Report Prepared by Economics Research Group. University of Guelph, Canada 1–26.
- Witten, I.H., Frank, E., Hall, M.A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco.
- World Organisation for Animal Health, 2008. PRRS: the disease, its diagnosis, prevention and control. Report of the ad hoc group on Porcine Reproductive Respiratory Syndrome, Paris, 9–11 June 2008, 1–7. World Health Organisation, 2010. Swine influenza. OIE Terrestrial Manual 2.8.8, 1–11.
- World Health Organisation, 2010. Swine influenza. OIE Terrestrial Manual 2.8.8, 1–11.
- Zhou, Y., Fenton, N., Neil, M., 2014. Bayesian network approach to multinomial parameter learning using data and expert judgements. *Int. J. Approx. Reason.* 55, 1252–1268.