UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

# THE NETWORK OF ASSOCIATIONS BETWEEN ANTIMICROBIAL CONSUMPTION, BIOSECURITY, VACCINATION AND PRODUCTION: AN ANALYSIS OF DANISH SOW HERDS USING ADDITIVE BAYESIAN NETWORKS

RICARDO ALEXANDRE PINTO LOPES

CONSTITUIÇÃO DO JURI

Doutor Virgílio da Silva Almeida

Doutor Fernando Jorge Silvano Boinas

Mestre Telmo Renato Landeiro Raposo Pina Nunes

ORIENTADORA

Doutora Lis Alban

CO-ORIENTADOR

Mestre Telmo Pina Nunes

2017

LISBOA

UNIVERSIDADE DE LISBOA

Faculdade de Medicina Veterinária

# THE NETWORK OF ASSOCIATIONS BETWEEN ANTIMICROBIAL CONSUMPTION, BIOSECURITY, VACCINATION AND PRODUCTION: AN ANALYSIS OF DANISH SOW HERDS USING ADDITIVE BAYESIAN NETWORKS

RICARDO ALEXANDRE PINTO LOPES

DISSERTAÇÃO DE MESTRADO INTEGRADO EM MEDICINA VETERINÁRIA

CONSTITUIÇÃO DO JURI

Doutor Virgílio da Silva Almeida

Doutor Fernando Jorge Silvano Boinas

Mestre Telmo Renato Landeiro Raposo Pina Nunes

ORIENTADORA

Doutora Lis Alban

CO-ORIENTADOR

Mestre Telmo Pina Nunes

2017

LISBOA

Dedicado a Martinho Simeão Pinto e José Lopes Espanhol,
dois grandes homens que, por crueldade da vida,
não puderam ver o neto a conquistar o primeiro grau académico.
Espero ter-vos deixado orgulhosos.

*“If I have seen further*
*it is by standing on the shoulders*
*of giants”*
- Isaac Newton

No journey is travelled alone, and this one was not an exception. For all the wonderful people who have walked alongside me in this path, you have my deepest and most sincere gratitude.

First and foremost, I would like to thank Lis Alban, for the wonderful guidance during this internship and for the privilege of working and learning with her, to Liza Rosenbaum Nielsen, for all the help and advices, and for making a young student 2500 km from home feel so welcome at work, and to Telmo Pina Nunes, for the herculean dedication, and for being one of most devoted people to the noble art of teaching I ever knew. How much I learned from you, and how much you inspired me, are variables too big for me to measure, and I will always be grateful for the privilege of working with you.

I would also like to express my gratitude to all the wonderful people from the Section for Animal Welfare and Disease Control, at the University of Copenhagen, for making me feel so much at home and happy to go to work every day. I would like to thank in particular Amanda Brinch Kruse, for sharing her data with me, and without who this project would not be possible, Matt Denwood, for all the help whenever I was getting confused with the analysis or when JAGS was driving me crazy, and, of course, to Leonardo Víctor de Knegt, for all the nerdy conversations, all the after-lunch walks (o ginásio estava sempre amarelo mas lutámos como heróis!), the great time at SVEPM, but most of all, for the true friendship. My stay in Copenhagen would not have been even half as enjoyable without this friendship, and it alone was worth this internship.

Many thanks to all the amazing people at the Danish Agriculture & Food Council, DTU-Food and DTU-Vet, the DVFA and the pharmaceutical industry, with who I had the opportunity to learn over the course of my internship, and who contributed greatly not only to this project but also to broaden my professional horizons, to the SVEPM, for granting me the bursary award which allowed me to attend such an amazing conference, and to the Erasmus+ programme, which helped to fund this internship. I would also like to thank Arata Hidano, for his help with the bootstrapping process and the good times at the SVEPM conference.

Lastly, I thank all the people who have walked this journey alongside me.

Many thanks to all my academic family, especially my godmother Susana Mendes, for being the inspiring person she is and all the good times we spent together, where I learned more than she would think, to Rita Telhada, my granny, for being my battle comrade in our artistic adventures and for never stopping to believe in my success, both as a scientist and as an

artist, to my goddaughter, Ana Catarina Rato, who I think of as my little sister, for the giving me the privilege of working with her for the Portuguese Academic Tradition, and most of all for the true friendship, and to my wonderful girlfriend, Ana Ferreira, who I am more grateful to have found than I could express in words. Meeting you alone would make these six years more than worth it.

Many thanks to my dear friends from the faculty, Manuel Veloso, João Goulão, Inês Andrade, Eva Soares and Ana Catarina Alves, who have walked this path with me since day one, and to João Figueiredo, Dinis Silva, Mariana Lopes, Carlota Gil, João Poças, Patrícia Fialho, Tiago Trovão and Ana Rodrigues, who I came to know later on. Our adventures during these years were the stuff of legends, and I am extremely grateful to be part of this little family. I look forward to see what the future has in store for us.

Finally, to my dear loving family, I thank with all my heart for the love and support I was given all my life. Especially to my parents, Nuno and Elda Lopes, I am grateful beyond words. Thank you for being who you are, and for raising me to be who I am today. This work has a piece of you in it.

# RESUMO

## A rede de associações entre consumo de antimicrobianos, biosegurança, vacinação e produção:
## Uma análise de efetivos de porcas na Dinamarca, usando Redes Aditivas Bayesianas

A necessidade de reduzir o consumo de antimicrobianos faz-se sentir cada vez mais, na agropecuária atual. Contudo, é necessário garantir a manutenção dos níveis produtivos aquando desta redução, pelo que se afigura importante a identificação de alternativas viáveis aos antimicrobianos.

Utilizando dados relativos à prescrição de antimicrobianos, parâmetros produtivos, biosegurança e vacinação, entre outros, o presente projeto conduziu uma análise à rede de interdependências entre as variáveis, com o objetivo de aprofundar conhecimento sobre a dinâmica das interações entre os vários fatores que determinam a produção porcina.

Para o efeito, foi usada a metodologia de redes aditivas bayesianas, uma técnica multivariada destinada a bases de dados de tamanho e complexidade consideráveis, cujo resultado final se materializa num gráfico acíclico direcionado – uma representação visual da rede de associações entre todos os parâmetros, discriminando as diretas das indiretas.

Como tal, foi descoberta uma ausência de associações relacionadas com consumo de antimicrobianos, indicando um maior impacto de outros fatores, como a biosegurança, na produção de porcas. O sistema SPF dinamarquês, e todos os padrões que engloba, também se revelaram influentes.

Apesar destes resultados deverem ser interpretados à luz daquilo que é o contexto dinamarquês, constata-se ser possível a redução da dependência em antimicrobianos.

Palavras chave: Redes aditivas bayesianas, redução, consumo, prescição, antimicrobianos, biosegurança, vacinação, produção, produtividade, gráfico acíclico direcionado, varas de porcas, dinamarquesas, Dinamarca

# ABSTRACT

## The network of associations between antimicrobial consumption, biosecurity, vaccination and production:
## An analysis of Danish sow herds using Additive Bayesian Networks

There is an increasing understanding that antimicrobial consumption must kept low, while not compromising productivity, in modern animal farming. However, viable alternatives to antimicrobials are needed.

Using data regarding antimicrobial prescription, production, biosecurity and vaccination, amongst others, this project included an analysis of the network of interdependencies between these variables, to deepen the understanding of the interactions between the factors which make up pig farming. The dataset included data from 160 Danish sow herds, with information regarding the sow and the weaner sector of the production.

With this in mind, additive bayesian networks methodologies were applied - a multivariate technique suited to process datasets of considerable size and complexity, which returns as final output a directed acyclic graph, representing of the network of associations between the parameters, and differencing direct from indirect ones.

As such, it was discovered no association regarding antimicrobial consumption-related variables, indicating other factores, such as biosecurity, as more impactful in sow farming. The Danish SPF system, and all the standards it implies, also appeared as relevant.

Although these results must be interpreted having the context of Danish pig farming in mind, it nonetheless reveals it is possible a reduction of the dependence in antimicrobials.

Keywords: additive bayesian networks, antimicrobial, consumption, reduction, biosecurity, vaccination, production, productivity, directed acyclic graph, Denmark, Danish, sow herds

# Index

I - Internship Report ........ 1

II - Bibliographic revision ........ 2

  II.I - The dangers of excessive antimicrobial usage ........ 2

  II.II - The effort for lower antimicrobial consumption in Denmark ........ 9

  II.III - Alternatives to AMs ........ 18

  II.IV - Additive Bayesian Networks ........ 22

  II.V - Complementary analysis ........ 31

III – Objectives ........ 33

IV – Materials and Methods ........ 34

  IV.I - The dataset ........ 34

  IV.II - Data Handling ........ 45

  IV.III - Additive Bayesian Network modeling ........ 49

V - Results ........ 52

VI – Discussion ........ 69

VII – Conclusion ........ 75

VIII – Bibliography ........ 76

Appendix I – Variables' summary ........ 82

Appendix II – Significant linear associations between the variables ........ 86

Appendix III – Frequency of arc retrieval for the sow model boostraping ........ 90

Appendix IV – Frequency of arc retrieval for the weaner model boostraping ........ 92

# List of figures

Figure 1- Number of Antibacterial New Drug Applications by year interval ............................ 5
Figure 2 - Proportion between animal species according to live biomass and AM consumption in kg of active compound ........................................................................ 7
Figure 3 - Yellow Card System chart of possible transitions ............................................... 13
Figure 4 - Prescribed AM agents for humans, and for all animal species, including the number of pigs produced, in Denmark acrosse the years.................................................. 14
Figure 5 - PCU and AM sales during 2014, by country, for 29 European countries ............. 16
Figure 6 - Example of a DAG, of a hypothetical ABN analysis on book sales. ...................... 23
Figure 7 - Example of a hypothetical discriminant analysis represented in a scatterplot. ...... 31
Figure 8 - Example of a hypothetical multi-group discriminant analysis. .............................. 32
Figure 9 - Regions of Denmark ........................................................................................... 38
Figure 10 - Exact search network scores obtained for each defined maximum parent limit .. 51
Figure 11- Number of farms by herd type. ........................................................................... 53
Figure 12 - Number of farms by SPF class. .......................................................................... 53
Figure 13 - Comparison between the number of sows and weaners ................................... 54
Figure 14 - Number of farms by AM prescription level, for the sow sector ........................... 55
Figure 15 - Number of farms by AM prescription level, for the weaner sector ...................... 56
Figure 16 - Comparison between AM prescription in sows and in weaners ......................... 57
Figure 17 - Comparison between the distribution of biosecurity scores ............................... 58
Figure 18 - Optimal Bayesian Network returned by the exact search for the sow model ...... 60
Figure 19 - Optimal Bayesian Network returned by the exact search for the weaner model. 60
Figure 20 - Scatterplot of the farms' values according to discriminant functions 1 and 2. ..... 61
Figure 21 - Partition plot presenting the result of the discriminant analysis ......................... 62
Figure 22 - Histograms presenting the separation of the observations according to the 4 groups attained by the first, second and third discriminant functions ................................... 63
Figure 23 - Example of histograms depicting data separation by successful discriminant functions ............................................................................................................................ 63
Figure 24 - Best performing FFT, with the internal biosecurity scores, the vaccination status for *Mycoplasma* and the ADDs/day for sows as cues for classifying farms as Top 20% or Bottom 80% ......................................................................................................................... 65
Figure 25 - Number of arcs recovered in the boostrapping of the sow model, after the removal of the arcs not included in the original DAG ......................................................... 66
Figure 26 - Number of arcs recovered in the boostrapping of the weaner model, after the removal of the arcs not included in the original DAG ......................................................... 67
Figure 27 - Final DAG for the sow model. ........................................................................... 68
Figure 28 - Final DAG for the weaner model ....................................................................... 69

# List of tables

Table 1 - Evolution of the Yellow Card Initiative thresholds across the years ………………..12
Table 2 - AM sales of veterinary antimicrobial agents marketed mainly for food-producing animals, population correction unit (PCU) and sales in mg/PCU, by country, for 2014, in ascending order of mg/PCU ……………………………………………………………………17
Table 3 – List of variables in the dataset …………………………………………………………35
Table 4 - External and internal biosecurity subtotals ……………………………………………42
Table 5 - Support percentage for every arc in the DAG for the sow model …………………..67
Table 6 - Support percentage for every arc in the DAG for the weaner model ………………68

## List of acronyms and abbreviations

ABN – Additive Bayesian Networks

ADD(s) – Animal Daily Dose(s)

ADDs/day – Animal Daily Doses per 100 animals per day

AM(s) – Antimicrobial(s)

AMP – Antimicrobial Prescription

CDC – Center for Disease Control and prevention

CHR – Central Husbandry Register

DADD - Defined Animal Daily Dose(s)

DAG – Directed Acyclic Graph

DANMAP - Danish Antimicrobial Resistance Monitoring and Research Programme

DDD(s) – Defined Daily Dose(s)

DNA – Deoxyribonucleic Acid

DVFA – Danish Veterinary and Food Administration

ESVAC - European Surveillance of Veterinary Antimicrobial Consumption

FFT – Fast and Frugal Trees

MCMC – Markov Chain Monte Carlo

PCU – Population Correction Unit

PCV2 – Porcine *Circovirus* type 2

PRRS – Porcine Reproductive and Respiratory Syndrome

SVEPM – Society for Veterinary Epidemiology and Preventive Medicine

VASC - Veterinary Advisory Service Contracts

WHO – World Health Organization

# I - Internship Report

As part of this project, an internship in Lisbon was undertaken, from January 2nd to January 28th 2017, which was continued in Copenhagen, Denmark, from January 29th to April 29th 2017. The part of the internship which took place in Denmark had support from the Danish Agriculture & Food Council and the University of Copenhagen. It was supervised by Lis Alban, and co-supervised by Liza Rosenbaum Nielsen and (in Lisbon) Telmo Pina Nunes. Financial aid was provided by the Erasmus+ programme.

To develop the competences required for this project, a training period took place in the Faculty of Veterinary Medicine of the University of Lisbon, from September 1st to December 16th, 2016. During this time, data handling techniques were learned, with supervision by Telmo Pina Nunes, and a course in data science regarding the use of the R programme (R Programming, by Johns Hopkins University, available in the Coursera platform at https://www.coursera.org/learn/r-programming) was completed. Further proficiency with this tool was achieved by completing a certain number of tasks, which required basic dominion over data analysis techniques and the use of Geographic Information Systems (which lead to learning how to use these tools, namely QGIS).

From the previous tasks, which included an analysis of the Portuguese situation regarding Aujeszky's disease in pig farming, with fellow colleagues Manuel Veloso and João Figueiredo, resulted a report, which was presented at the XII Simposium of the Portuguese Pig Farming Scientific Society (XII Simpósio da Sociedade Científica de Suinicultura), which took place November 18th, 2016. In this presentation, the difference between the estimated date at which 80% of the Portuguese farms would be free from the disease and the date this would be achieved if the national Plan for Control and Erradication of Aujeszky's Disease was strictly followed was highlighted, as models for the eradication of the disease over time, according to each of these two situations, were presented.

In January, while continuing to learn about data handling and analysis, a research regarding ABN (Additive Bayesian Networks) was also conducted, as this was decided to be the method of choice for processing the dataset in this project. To provide a basic overview of the dataset, a search for linear associations between all the continuous variables was performed, as a preliminary analysis.

Over the duration of the internship in Copenhagen, knowledge about ABN was deepened, as the analysis of the database using that same modeling methodology was conducted. In parallel, many meetings took place with colleagues from various different fields and areas of expertise (namely colleagues from the University of Copenhagen, the Danish Agriculture & Food Council, the Danish Technical University, the Danish Veterinary and Food Administration and the pharmaceutical industry), who provided valuable input regarding the project, contributing with their own perspective and personal experience. Simultaneously, through the various meetings, a whole plethora of workplaces and career options regarding veterinary epidemiology and risk assessment were presented, which contributed to enrich my personal understanding of this profession's possibilities, and to broaden my professional horizons.

Halfway throughout the internship, after an application being sent, a bursary award for the SVEPM (*Society for Veterinary Epidemiology and Preventive Medicine*) Conference and Annual General Meeting, taking place in Inverness, Scotland, was received. This made it possible for me to attend the conference, as well as presenting a poster describing this project and presenting its preliminary results. The poster was presented in a poster pitch, and was later awarded with the SVEPM poster prize. The attendance to the conference, and to the Young Scientists and Researchers event, held the day before the beginning of the conference, further contributed to my general awareness of the possibilities for a veterinarian in the field of epidemiology, as well as providing a chance to meet many colleagues in the same area.

The internship culminated in a seminar held at the University of Copenhagen, where the latest results were presented to an audience composed of many colleagues from the area of epidemiology, animal welfare, pharmaceutics and the food and veterinary services. The seminar was further enriched with presentations regarding the theme of antimicrobial usage from two other colleagues, Amanda Brinch Kruse and Leonardo Victor de Knegt, from the University of Copenhagen.

# II - Bibliographic revision

## II.I - The dangers of excessive antimicrobial usage

The development of antimicrobial (AM) resistance by microorganisms through mutational resistance is a natural process, easily understandable by recalling Darwin's theories on natural selection and basic concepts of population genetics: a selective pressure (in this case, the AM) over a number of organisms, which hinders or kills a certain subset of the total population, and which is resisted by a different subset (due to a mutation in the genetic code, for instance), will enhance the second group's chance of survival and reproduction. Said resistance is normally conferred through reorganization of the AM receptors, resulting in lower affinity, decrease of the drug absorption, more efficient excretion of the AM or changes to the general metabolism, as reviewed by Munita & Arias (2016).

Improved chances of passing down genes which offer resistance to said selective factor will result in a bigger percentage of the total offspring possessing the genes that confer resistance, when compared to the previous generation. As such, should the selective pressure persist (and ignoring other factors), over the course of time the whole population will drift towards a higher and higher number of resistant individuals, or even suffer altogether a dramatic shift into a completely resistant population, in the cases where the susceptible organisms are eradicated outright, leaving only the resistant ones.

In addition, bacteria can also obtain resistance against AMs through Horizontal Gene Transfer, one of the main boosters of bacterial AM resistance (and one of the most dangerous), through which the microorganism receives genetic material from an external source. Bacteria can either incorporate naked DNA (a histone-free chain of DNA), suffer transduction (receive genes through a phage intermediate) or acquire new genetic material through direct trades between microorganisms, involving cell-to-cell contact, which increases the microorganism’s capabilities to resist AM pressure exponentially, as resistances traits are possible to share in mass amongst populations (Munita & Arias, 2016).

The emergence of resistance against AMs in bacteria populations is a natural evolutionary response to selective pressures, much like any other Darwinian selection process, and it is not something new. Most of the antimicrobials used nowadays are natural compounds, or artificial modifications of those, produced by microorganisms as a defense mechanism to external, toxic substances, as reviewed by Holmes et al. (2016). Therefore, AM resistant bacteria have existed for centuries, having been selected over time to respond to those same substances in a way which would improve their survivability. However, the number of AM resistance cases has grown over the years, at an unusually fast pace for natural selection. This is deeply connected to the human discovery of AMs, and in great part, to their overuse, the plummeting number of new AMs available and the extensive use of AMs in agriculture, as reviewed by Ventola (2015).

The modern age of AMs began with the discovery of the penicillin in 1928. The mere concept of AMs would revolutionize modern medicine and allow the treatment of many previously untreatable diseases. However, just a dozen years after the discovery of this antibiotic, two members of the same discovery team identified a penicillinase of bacterial origin. To safeguard the medical advances provided by this AM, this discovery forced the artificial manipulation of the penicillin in an attempt to make it immune to this enzyme, hence surpassing this AM resistance. Eventually, new AMs compounds were discovered and deployed in medicine, and they as well faced a similar problem, with resistant bacteria arising. Nowadays, despite the large amount of AMs available and all the knowledge obtained since the penicillin was first used, bacterial infections are once again a threat to human health, sometimes untreatable and deadly, since they can bypass or even ignore our AMs. (Davies & Davies, 2010; Ventola, 2015)

Looking back, the very "father" of the penicillin, Sir Alexander Fleming, had foreseen the arousal of this danger, just 17 years after his discovery of the penicillin. In an article from The *New York Times* dating back to 1945, 5 years after the discovery of the penicillinase by two of his coworkers, he predicted: "The public will demand [the drug and]...then will begin an era... of abuses...". In the same statement, Fleming implied the irresponsibility of AM misuse, claiming "[…] the thoughtless person playing with penicillin treatment is morally responsible for the death of the man who finally succumbs to infection with penicillin-resistant organism." ("Penicillin's finder assays its future," 1945; Spellberg & Gilbert, 2014).

One of the factors which makes AM resistance such a serious danger is the lack of availability of new AMs.

As previously stated, AM resistant bacteria emerge as part of a natural phenomenon, and that emergence cannot stopped, only slowed and contained. For 'tackling AM resistant infections with AMs' to be considered a viable long-term solution, new AMs had to be developed and deployed at a steady rhythm, so that the obstacle of new resistance emergence against older AMs would be easily surpassed with the newest AMs, to which bacteria had little time to adapt to, and develop resistances. That, unfortunately, seems not to be the case, considering the trends of AM production.

The rate of development and approval of new AMs is slowing down, as represented in Figure 1. Data from the CDC (Center for Disease Control and prevention), as reviewed by Ventola (2015), shows a clear decline in the number of antibacterial new drug applications per year interval. Between 2005 and 2014, only 9 new AMs were developed and approved. Between 1980 and 1989, the same time span - only 25 years earlier - the number of approved AMs was 29.

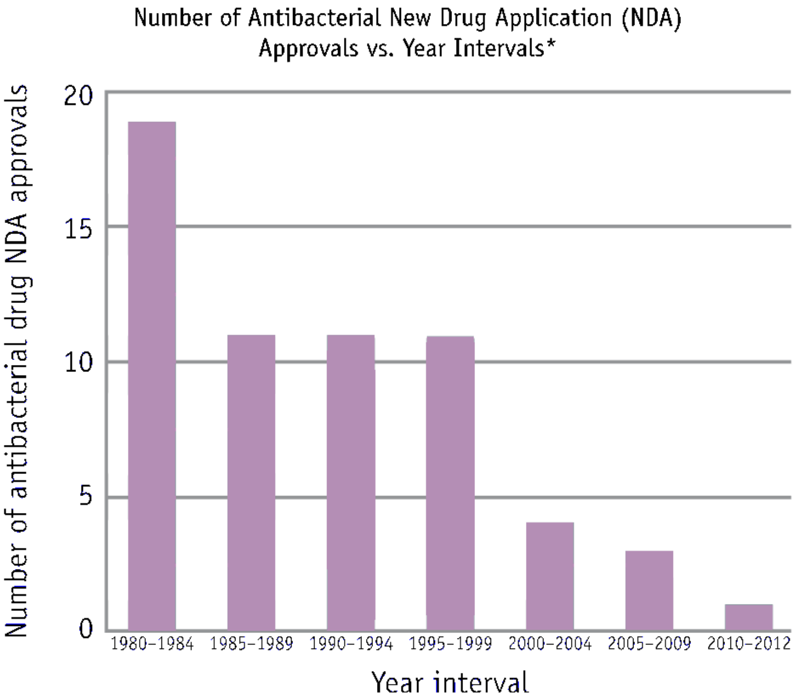


**Figure 1- Number of Antibacterial New Drug Applications by year interval, as reported by the CDC, according to data from the United States Food and Drug Administration's Center for Drug Evaluation and Research. Graph adapted from *Antibiotic Resistance Threats in the United States, 2013* (CDC, 2013)**

There are many reasons behind the decrease in number of new AMs, as it is a complex multifactorial issue. Overall, the AM development field is a risky one for a pharmaceutical company to invest on, and a difficult one to thrive in.

Currently, treatment for chronic conditions is much more profitable for pharmaceutical companies. According to the Office of Health Economics in London, the net present value of a drug used to treat a neuromuscular disease is 20 times superior to an AM's (as reviewed by Ventola, 2015). That difference can drive companies away from new AMs research. Not only AMs are normally prescribed for short periods of time (while treatment for chronic diseases requires commitment over a longer time period), they are relatively cheap and usually used as a last resort, in order to prevent the emergence of resistance. There is also the possibility of postaproval adverse effects - something that occurred, for example, with trovafloxacin, which only after its release revealed to be causing liver toxicity. Furthermore, the most recent AMs tend to be kept for infections which have resisted older agents, and are therefore used as the very last treatment option (which brings the pharmaceutical companies no substantial short-term profit), and even with such precautions against AM resistance, there is no form of prediction as to when will a resistance appear, making it impossible to ascertain a life expectancy for the AM, which could be a decade or a mere year. These circumstances offer the pharmaceutical companies few assurances and big risks for less profit, when compared to other types of drugs. Over these issues, many regulatory barriers also provide extra problems for the industry to overcome, such as different clinical trial requirements between countries or changes in regulatory and licensing rules (Conly & Johnston, 2005; Ventola, 2015)

As far as resistance emergence is concerned, the misuse and overexploitation of AMs is the leading cause (CDC, 2013), but even though many reports and articles on the subject focus on the use of AM in human medicine, their use in agriculture is also one main contributor to the magnification of AM resistance-related problems, and the threat they pose to human health.

The use of AMs in livestock is accepted as beneficial, increasing the quality of the product, the quantity and the rate at which it is produced (Boeckel, Brower, Gilbert, Grenfell, & Levin, 2015; Michael, Dominey-howes, Labbate, Maria, & Elisabeth, 2014). This can be interpreted as a direct consequence of the animal health improvement, as disease and infection-free, healthy animals have a higher productivity than diseased animals. As such, farmers quickly understood the benefits of AMs as growth promoters, and they were used as such. Since 2006, the use of AMs in animal feed as growth promoters has been banned, as part of the EU food safety strategy (European Comission, 2005). However, these substances are still used widespread, without close supervision in some cases (which is a serious matter as resistances occur even when the proper guidelines are respected), as this ban does not apply to countries outside the European Union.

The extent at which AMs are used for agriculture is also alarming: it is estimated that, in the United States of America, the use of AMs in food animals makes up about 80% of the total annual consumption in the country, and the estimated total global consumption as of 2010 was 63.151 (±1.560) tons (Boeckel et al., 2015). In Denmark, considering the year of 2015, about 105237 kg of AMs were sold to be used on production animals, and that already represents a 4698 kg reduction from the previous year. From that amount, about 75% are AMs destined to pig farming, as represented in Figure 2. Pigs and cattle are, by far, the two most important species in Danish animal farming, which in terms of live biomass are roughly equivalent: amounting to 43% and 47%, respectively, surpassing all the other species in this regard. However, since a big part of the cattle biomass is due to the dairy industry, and dairy cows require lower levels of AMs when compared to growing animals, it becomes clear why pig farming represents such a big slice of the AM consumption when compared to cattle, which represents only 12% (Statens Serum Institut, DTU VET - National Veterinary Institute, & DTU Food - National Food Institute, 2015).

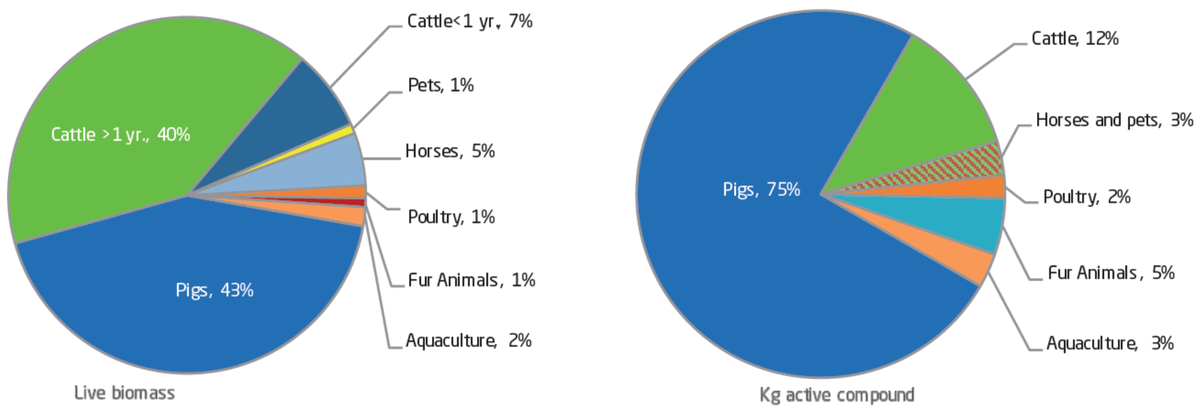


**Figure 2 - Proportion between animal species according to live biomass (left) and AM consumption in kg of active compound (right). Figure taken from DANMAP 2015 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark, ISSN 1600-2032, available at danmap.org** (Statens Serum Institut et al., 2015)

Even ignoring the possible intake of AMs when people ingest food produced from animals treated with AMs, the use of these drugs in farm animals enables the emergence of resistant bacteria, which can be transmitted to humans through the food supply. Considering that about 50% of the AMs used in human medicine are also prescribed as treatment in veterinary medicine, this is a serious health risk (CDC, 2013; Ventola, 2015). The potential for animal-to-human transmission of AM-resistant microorganisms has been recognized since the 1960s: Anderson and Lewis (1965) confirmed the spread of an AM-resistant strain of *Salmonella typhimurium* across England trough calf movements, and while they affirm no evidence was found that the drug-resistant bacteria were, in any way, more virulent than their non-resistant counterparts, the survivability advantage they possess against non-resistant strains increase their chances of affecting human health, either through direct infection, causing disease, or indirectly (and arguably more worrisome), transmitting their resistance to commensal microbiota, perpetuating the resistance gene (Anderson & Lewis, 1965; Holmes et al., 2016).

The magnitude in which AMs are used in farm animals can also be problematic, since many treatments are administrated as group treatments. As healthy animals are exposed to AMs, an unnecessary selective pressure is exerted over the animal's microbiota that favors resistant bacteria, therefore increasing the chances of resistance transmission to other (possibly pathogenic) bacteria and to humans. Furthermore, the environmental microbiota can also be affected by the use of AMs in farm animals, as up to 90% of the compounds are excreted through urine and feces, which can then spread across large distances through water (either superficial or underground) or in ground fertilization. This, while not posing immediate danger to human health, contributes to increasing the pool of microorganisms resistant to AMs, and resistant genes, with all the public health implications discussed above that would bring (Ventola, 2015).

Both in human medicine and in veterinary medicine, another promoter of AM resistances is the drug misuse - that is, the inadequate prescribing of AMs in situations where their prescription would be incorrect. Older studies show that, at some point, the prescription of antibiotics in human medicine for upper respiratory tract infections in the United States amounted for 75% of the yearly AM prescriptions in ambulatory setting, which are mostly caused by viruses (Arnold & Straus, 2009; Gonzales, Steiner, Lum, & Barrett, 1999). There is also evidence that the duration and/or choice of agent in 30 to 50% of cases where antibiotic treatment is prescribed in the United States is incorrect (CDC, 2013). Overall, the unnecessary, inappropriate or suboptimal use of AMs in AM therapy is beneficial for resistance emergence, through the mechanisms already discussed (Munita & Arias, 2016; Ventola, 2015).

All this considered, it is clear that the menace which AM-resistant microorganisms pose is very serious, and, unless dealt with soon, has the potential of reversing decades of advances in medicine, rendering us unable to treat bacterial infections – ones easily treatable mere years prior.

Recalling the mechanisms through which microorganisms acquire resistance to AMs, we know that AM resistances appear easily, but are very hard to eliminate (Mateu & Martin, 2001) and one of their main promoters is the overuse (and misuse) of AMs (Ventola, 2015). Logically, the fight against AM resistances – one we must employ if we want to keep using AMs in the future – must focus on more rational and sustainable AM usage policies. Those should contemplate many key factors to minimize the development of resistances, neither sacrificing the treatment effectiveness nor the public health. One of those factors would be better criteria as to when to use AMs (based on more accurate diagnosis, on knowledge about the used AM, AM susceptibility tests, to name a few) combined with better compliance and respect for the correct dosages and treatment duration. One other factor, perhaps the most obvious one, would be to reduce the usage of AMs (Mateu & Martin, 2001).

By reducing the AM consumption as a whole, the decrease in selective pressure on the microorganisms would, at the very least, slow the emergence of resistances. Considering the level of AM consumption at farm level, and its importance on the promotion of AM resistance, this seems like an area where it is relevant to intervene.

Mind that, while there is a general idea that AM resistant bacteria are somehow less fit than its non-resistant counterparts, maybe based on the belief that if should be more costly for the organism's metabolism, that is not what usually happens. The case of resistance against fluoroquinolones is a good example of this. Despite the worldwide dissemination of flouroquinolone-resistant microorganisms, studies reveal that combined mutations on resistant bacteria can improve their *in vitro* and *in vivo* fitness, even in the absence of the

AM. Resistant bacteria can be as fit, or even fitter than non-resistant ones (Holmes et al., 2016; Marcusson, Frimodt-møller, & Hughes, 2009).

What this means is that reducing the prevalence of AM resistant strains of bacteria is not as simple as cutting down the AM usage, and waiting for the population to shift back into a non-resistant majority (or totality). The reversibility of the AM resistance problem is very complex, and not linear, at all. Some studies report that, in countries such as Denmark and Norway, some measures limiting the use of certain AMs came to reveal a drop in the amount of resistant bacteria isolated from the animals in the years after the ban, but even several years after the selective pressure was removed, the resistance did not disappear, possibly because of co-selection. It is unpredictable the time it would take for a AM resistance to decrease until undetectable levels (Aarestrup et al., 2001; Borgen et al., 2000).

However, that is not to say that the reduction of the amount of AMs used will have no effect on this problem at all. Eradication of AM resistance may not be as linear as we would wish, since resistant microorganisms can persist in the animal or in the environment for long periods of time, even after the removal of the selective pressure of the AM which gave birth to them in the first place, but that would definitely contribute to slow down the spreading of AM resistance and the development of new ones, and that, by itself, is a crucial measure we must employ if we want to combat this problem.

## II.II - The effort for lower antimicrobial consumption in Denmark

There are already many initiatives aiming to reduce the antimicrobial usage in agriculture (and their usage overall as well). One of the first countries who adhered to this paradigm of lowering AM consumption was Denmark. There, the use of AMs in food animals is under close surveillance, and a multitude of factors has been enabling a reduction in the total amount of AMs used, such as the referred monitoring (and legal infrastructure supporting it), many interventions in the animal sector (such as the Yellow Card initiative), voluntary actions taken by the industry or the DANMAP (Danish Veterinary and Food Administration, 2015a; Statens Serum Institut et al., 2015).

DANMAP – the Danish Antimicrobial Resistance Monitoring and Research Programme – was created from an initiative of the Danish Ministry of Health and the Danish Ministry of Food, Agriculture and Fisheries. This program participants include the National Food Institute and Danish Veterinary Institute, at the Technical University of Denmark (DTU) - which gather data from veterinary practices, slaughter plants and food control laboratories, as well as data from the Danish Veterinary and Food Administration (DVFA) - and the Statens Serum Institut, which provides data from human medicine. DANMAP further receives data directly from the DVFA and Vetstat, a database further described in IV – Materials and Methods. Its objectives include monitoring the consumption of AMs, the occurrence of AM resistance and to study associations between the previous two, with the core principles of One Health in mind.

Since its implementation in 1995, DANMAP has been monitoring AM use in humans and animals. DANMAP's first interim report was published in 1996 and since then it has been providing yearly reports on the AM and AM resistance universe, which often make up the foundations for measures and interventions put in place not only by the veterinary and food authorities but also by the animal industry associations.

As the DANMAP 2015, the commemorative report celebrating the 20 years of the program, enlightens us, Denmark has been employing restrictive politics regarding AMs for some time. Many measures and incentives are in place to encourage a more responsible and intelligent use of AMs, as well as some others to discourage their abuse.

There is, for instance, a high level of transparency in medicine pricing by the pharmaceutical industry, thanks to regulations, and the veterinary practitioners make no profit through the sales of antibiotics, therefore eliminating the incentive to overprescribe. Instead, the main income for a veterinarian in Denmark comes from the specialized diagnostic services and treatment they offer, rather than medicine sales. This can be a reflection of legislation measures which first appeared in 1995, implementing VASC (Veterinary Advisory Service Contracts), which are mandatory for large pig herds since 2010 and assures frequent veterinary advisory visits to animal farms, therefore contributing to a better herd health, an extended treatment arsenal at the farmer's disposal and, perhaps more importantly, a shift from a treatment-oriented mentality to a preventive mindset. This allows the reduction of AM consumption, through disease prevention and treatment optimization (Statens Serum Institut et al., 2015).

In Denmark, all prescription-only medicines can only be obtained by farmers through a veterinarian, who must have obtained it from a pharmacy, or alternatively directly from the latter. Some vitamins, ectoparasitic drugs and mineral complexes are except from this, as they can be obtained from dealers authorized by the Danish Medicines Agency, but apart from those, strict regulations are in place to closely monitor the distribution of veterinary drugs (Danish Veterinary and Food Administration, 2017).

Any prescription of a drug for veterinary purposes since 2000, including (but not limited to) medicated feed, vaccines and antimicrobials, must be reported by the veterinarian or the pharmacies to an official state-owned database, called VetStat – the Danish Veterinary Medicines Statistics Program. The purpose of this register consists in monitoring the veterinary drug usage and allow the DVFA to act and regulate certain farms' drug usage (of AMs for example), should it be necessary, while providing transparency as a basis to ensure compliance with rules and legislation, but also to provide data for research and control purposes and to act as an advisory tool to veterinarian practitioners.

VetStat effectively covers 100% of animal herds receiving treatment with prescription only drugs, informing, for each prescription, the date of purchase, the product purchased and what was the amount acquired, the identification of the prescribing veterinarians, as well as the entity providing the drug, the herd identification and age/disease groups targeted by the prescription (Danish Veterinary and Food Administration, 2015a, 2017; Dupont et al., 2017).

Antibiotics face specific regulations, as they are only allowed to be used in production animals in cases when the attending veterinarian diagnosed an infection that demands AM treatment. The treatment should only be administrated by the veterinarian, and only on certain conditions can he prescribe AM treatment for the farmer to use – cases in which precise instructions must be given in writing, including many specifications such as the identification of the animals the drug and dosage to administrate, the administration route, clinical signs to be observed before treatment and withdrawal period. Even so, AM treatment in these conditions cannot exceed 5 days in duration.

Considering the specific case of fluoroquinolones, AMs belonging to that group may only be used when facing infections which have been confirmed, by laboratory examination, as being non sensitive to any other AM, except in acute diseases where the agent shows no sensitivity to other AMs, and never for a time period longer than 5 days. Also, treatment with fluoroquinolones in production animals must be notified to the Regional Veterinary Officer within two weeks of its conclusion (Danish Veterinary and Food Administration, 2017).

One of the measures which had the biggest impact on the decrease of AM consumption dates to 2010. Named the “Yellow Card Initiative”, it was a response from the DVFA to the rise in AM consumption between 2001 and 2009, which reached 45%, four fifths of which were due to pig farming. Its initial goal was to achieve a reduction in AM consumption of 10% from 2009 to 2013, goal which was met and which has grown, as the current objective is to achieve a 15% reduction by 2018, against the AM consumption levels in 2015 (Danish Veterinary and Food Administration, n.d., 2017b; Statens Serum Institut et al., 2015)

The initiative is based on a warning system, in which a card is issued every time a farm exceeds the AM threshold defined by the DVFA. The thresholds, presented in Table 1, which are specified for three different age groups and defined in Animal Daily Doses per 100 animals per day (unit of measure further discussed in IV – Materials and Methods), are updated regularly by the authorities, having been lowered across the years. Furthermore, as of March 31st, 2017, the thresholds refer to a weighted AM consumption level, in response to different AMs possessing different degrees of importance and/or danger regarding resistance emergence.

| Yellow card thresholds for pigs (in ADD/100 animals/day) | Piglets, sows, gilts and boars | Weaners (up to 30 kg) | Finishers |
|---|---|---|---|
| September 2010 – May 2013 | 5,2 | 28 | 8 |
| June 2013 – October 2014 | 5 | 25 | 7 |
| November 2014 – March 2017 | 4,3 | 22,9 | 5,9 |
| April 2017 – December 2017 | 4,1 | 21,8 | 5,6 |
| January 2018 onwards | 3,8 | 20,2 | 5,2 |

**Table 1 - Evolution of the Yellow Card Initiative thresholds across the years. Values taken from the DVFA's report "Special provisions for the reduction of the consumption of antibiotics in pig holdings (the yellow card initiative)", available at https://www.foedevarestyrelsen.dk/english/SiteCollectionDocuments/Dyrevelfaerd%20og%20veterinaermedicin/Veterin%C3%A6rmedicin/Yellow%20Card,%20English%20version,%20180517.pdf**

If a farm's average AM consumption exceeds at least one of the thresholds, over a nine month period, a yellow card is issued – an order, or injunction, which compels the farmer to reduce the AM consumption back within the thresholds within nine months of the yellow card. During this period, the authorities can also forbid storing AMs meant to be used in the feed or drinking water in the farm, and conduct one or more unannounced inspections to the farm.

After the nine month period, should the AM consumption levels remain over the limits (or if the AM levels have lowered below the limits, but then risen above them once more within twelve months after the initial nine month period expiry date), the holding enters a stage of Increased Supervision. A second injunction is then issued, which compels the farm owner to require additional aid. This comes under the form of expert advice from an impartial veterinarian (one other than the regular veterinarian assisting the farm), and contemplates devising a concrete plan, containing specific interventions to decrease AM consumption. The holding then has a five month time limit to return the consumption levels below the thresholds. As with the first injunction, the DVFA can forbid AM storage and carry inspection visits to the farm.

These cards, or injunctions, are harmful to the farmers, as they are required to pay not only a fee for each injunction and the prohibitions issued but also for the inspection visits and the veterinary expert services required for the second opinion. Receiving cards translate into increased financial costs for the holdings. It is, therefore, highly undesirable to be issued a yellow card. This way, there is a considerable financial cost for engaging in bad AM management policies, over the "moral weight" of such practices, thus encouraging a sustainable use of AMs within a farm.

In the case five months go by after the second injunction, without the consumption levels returning to the allowed levels, the DVFA issues a third injunction – the Red Card – which compels the owner to implement one of more of the measures included in the devised plan of action, mentioned previously, or, as a last resort, forces the owner to reduce the number of animals enough to assure the AM consumption falls below the established limit. Unlike the previous injunctions, the Red Card has no time limit, and remains active until the AM levels are within the allowed threshold. Unscheduled inspection visits to the farm and AM storage prohibitions are also possible, just as before (Danish Veterinary and Food Administration, n.d.).

Figure 3 displays the possible transitions within the Yellow Card Initiative, for easier reading and interpretation.

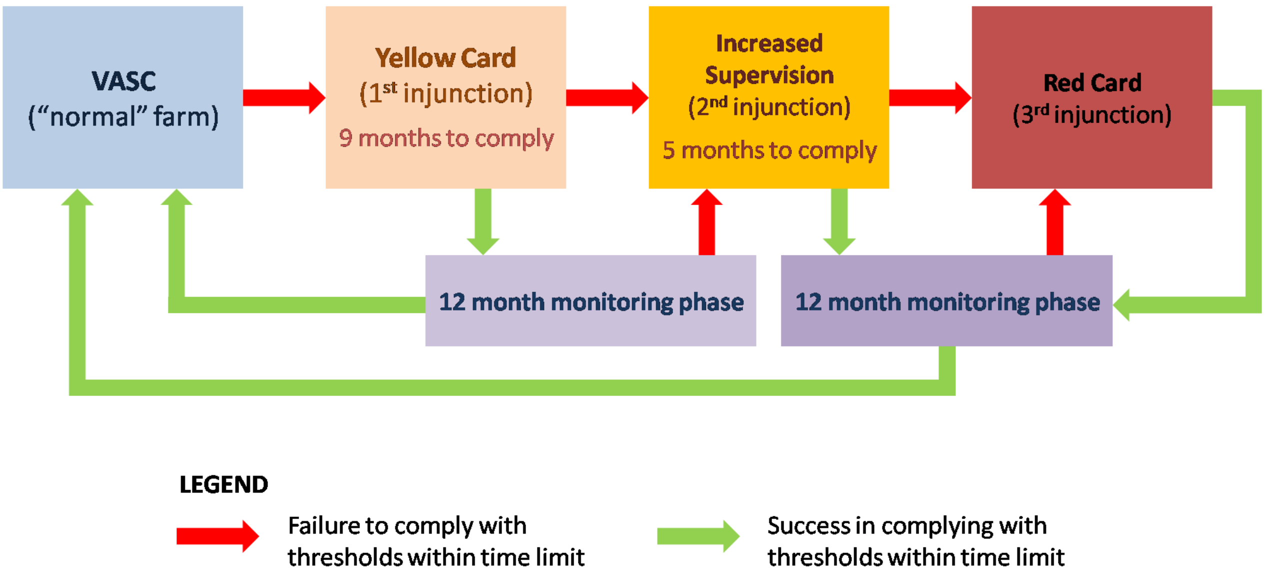


Figure 3 - Yellow Card System chart of possible transitions

The Yellow Card Initiative has helped greatly in the reduction of AM consumption in the Danish animal sector. From 2010 to 2011, once the Yellow Card was set in motion, the AM consumption has been gradually decreasing. As of 2015, AM consumption by production animals had decreased by 19% when compared to 2009, when it had hit a peak in consumption (Statens Serum Institut et al., 2015).

In spite of this remarkable decrease in consumption, this was not the only measure which contributed to it. In fact, in Denmark we can notice a great effort from the industry itself, when it comes to tackling the AM resistance issue. After the DANMAP reports raised concerns regarding resistance emergence against 3rd and 4th generation cephalosporins, the pig industry has voluntarily decided, in 2009, to stop using those AMs in pigs, except as a last resort. The decision was later mimicked by the Danish dairy cattle association, in 2014.

Resorting to VetStat, it is possible to confirm the clear, dramatic effect this had on cephalosporin consumption, which plummeted after this decision was taken, their use reaching close to zero in 2011. According to DANMAP 2015, the amount of cephalosporins sold for use in pigs, in 2015, was around 1 kg. In contrast, the amount sold for cattle amounted to 70 kg, and for pets it was estimated to be around 98 kg (Statens Serum Institut et al., 2015).

Across the years, despite a massive increase in the Danish animal production, namely in the pig sector, which witnessed a rise in production of almost 90% in animal heads from 1990 to 2015, the AM consumption has been decreasing steadily, while keeping up with the industry expansion, as one can see in Figure 4.

The treatment proportion – that is, the amount of AMs prescribed per number of animals and per time unit – in the pig sector (which accounts for around 75% of the total consumption, in kg of active compound, for veterinary purposes) has dropped about 22% from 2009 to 2015. This reduction is all the more impressive when we consider that the pig production in 2009 was 27603 thousand heads (1898 million kg) while in 2015 it was of 30874 thousand heads – 1954 million kg. That translates into a 12% increase in the number of animal heads produced, keeping pace with the 19% AM reduction (Statens Serum Institut et al., 2015).

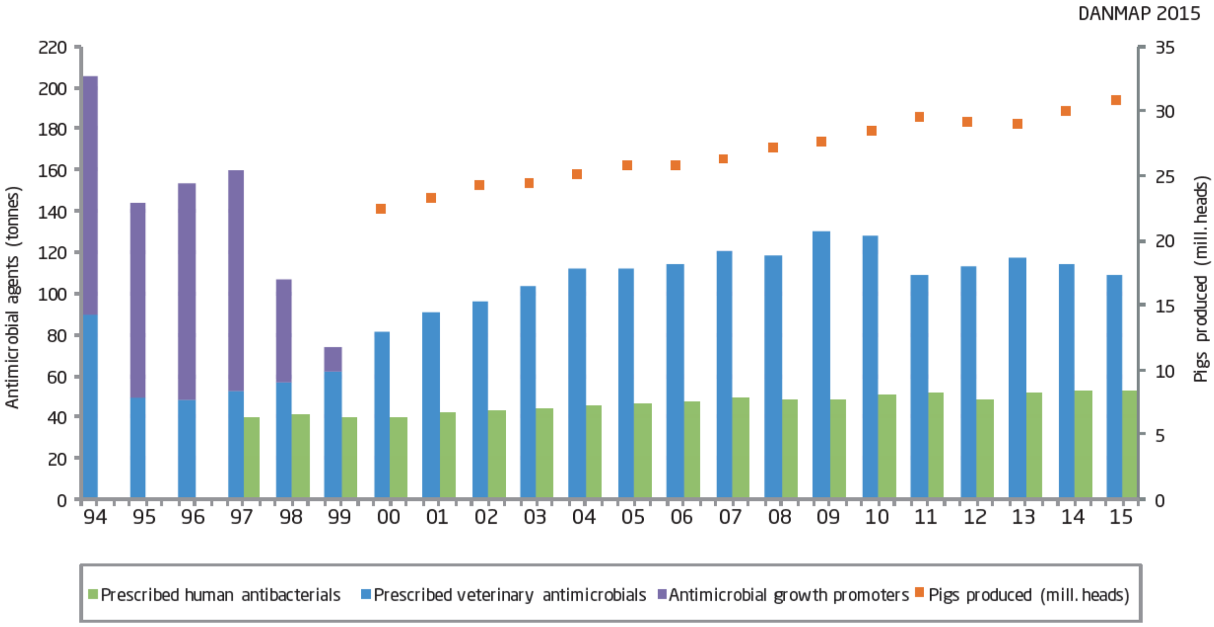


**Figure 4 - Prescribed AM agents for humans, and for all animal species, including the number of pigs produced, in Denmark acrosse the years.**

**Figure taken from DANMAP 2015 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark, ISSN 1600-2032, available at danmap.org**

(Statens Serum Institut et al., 2015)

When we compare these numbers with other European countries, some differences become clear. According to the European Medicines Agency's latest ESVAC report (European Surveillance of Veterinary Antimicrobial Consumption): Sales of veterinary antimicrobial agents in 29 European countries in 2014, Denmark is one of the ten countries with a highest animal production for food. This group also includes, in decreasing order of production, Germany, France, Spain, United Kingdom, Poland, Italy, Netherlands, Romania and Ireland, of which only Ireland has a lower food-producing animal population (European Medicines Agency, 2016).

However, in that same group, Denmark occupies the second place for less AM sales per animal population unit, and third for lower total AM sales.

As it can be seen in Figure 5, in 2014 Denmark is said to have produced 2415.3 tonnes of food animals, in PCU (Population Correction Unit – a unit calculated by multiplying the numbers of livestock and slaughtered animals by the estimated weight at the most probable time for treatment). With a total of 106.8 tonnes in AM sales during that year, that translates into an average 44.2 mg of AM per PCU. Other countries such as Italy or Spain (which produce, respectively, 1.6 and 2.9 times more than Denmark, in PCU) present AM sales values which are, respectively once more, 13.4 and 27.8 superior to Denmark's (European Medicines Agency, 2016).

Considering the 29 countries included in the EVASC report, which are listed in table 2, Denmark ranks tenth amongst the one with lowest mg/PCU, Norway being the number one. Comparing the data with Portugal, which occupies the 26th place out of 29 in the countries with lowest mg/PCU, with 190 tonnes in AM sales and 942.2 thousand tonnes of PCU, Danish production regarding food animals in 2014 was 2.6 times bigger while the total mass of sold AM was almost a half of the Portuguese counterpart. This implies an AM use 4,56 times bigger in Portugal when comparing it to Denmark, in mg/PCU (European Medicines Agency, 2016).

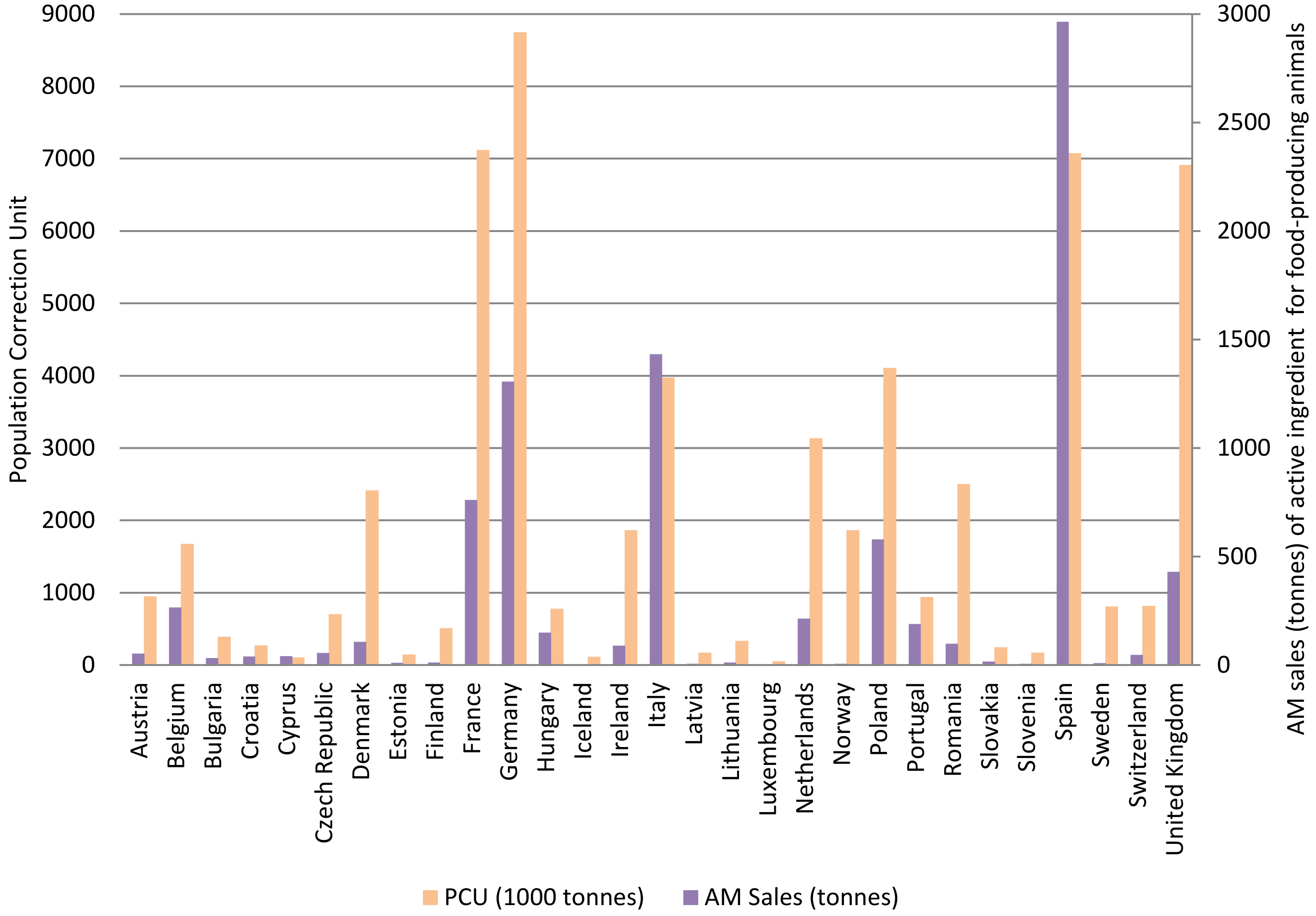


**Figure 5 - PCU and AM sales during 2014, by country, for 29 European countries. Data taken from EMA's sixth ESVAC report: Sales of veterinary antimicrobial agents in 29 European countries in 2014**

All of these statistics point towards one thing: the policies in place and the system employed in Denmark are successfully reducing the need for AMs, while still maintaining a very high production, bigger than most European countries.

This not only proves it possible to reduce the AM load on food-producing animals without this meaning an inevitable submission to worst production, but also point towards a good example.

It is undeniable that the Danish system works, and there are lessons to be taken from it, regarding alternatives to the overexploitation of AMs which we have been witnessing in the last few decades. It is, without a doubt, a matter worth investigating, for a treatment mentality based on AM usage is unsustainable, considering all the problems associated which were previously mentioned, such as AM resistance or the slow pace at which new AMs are developed and deployed.

| | AM Sales (tonnes) | PCU (1000 tonnes) | mg/PCU |
|---|---|---|---|
| Norway | 5.8 | 1866.1 | 3.1 |
| Iceland | 0.6 | 115.8 | 5.2 |
| Sweden | 9.3 | 810.8 | 11.5 |
| Finland | 11.4 | 509.4 | 22.3 |
| Slovenia | 5.7 | 171.2 | 33.4 |
| Lithuania | 11.9 | 335 | 35.5 |
| Latvia | 6.3 | 173 | 36.7 |
| Romania | 98.1 | 2501.8 | 39.2 |
| Luxembourg | 2.1 | 52 | 40.9 |
| Denmark | 106.8 | 2415.3 | 44.2 |
| Ireland | 89.6 | 1866.4 | 48 |
| Austria | 53.4 | 948.3 | 56.3 |
| Switzerland | 46.4 | 816.3 | 56.9 |
| United Kingdom | 429.6 | 6914.7 | 62.1 |
| Slovakia | 16.3 | 248.1 | 65.9 |
| Estonia | 9.8 | 144.4 | 68 |
| Netherlands | 214.5 | 3135.2 | 68.4 |
| Czech Republic | 55.9 | 702.6 | 79.5 |
| Bulgaria | 32.6 | 393.5 | 82.9 |
| France | 761.5 | 7119.8 | 107 |
| Poland | 578.5 | 4108.8 | 140.8 |
| Croatia | 40.2 | 273.4 | 147.2 |
| Germany | 1305.8 | 8748.6 | 149.3 |
| Belgium | 265.7 | 1678 | 158.3 |
| Hungary | 150.4 | 779.1 | 193.1 |
| Portugal | 190 | 942.2 | 201.6 |
| Italy | 1431.6 | 3977.4 | 359.9 |
| Cyprus | 41.7 | 106.6 | 391.5 |
| Spain | 2963.9 | 7077.1 | 418.8 |

**Table 2 - AM sales (in tonnes of active ingredient) of veterinary antimicrobial agents marketed mainly for food-producing animals, population correction unit (PCU) and sales in mg/PCU, by country, for 2014, in ascending order of mg/PCU. Table taken from EMA's sixth ESCAV report: Sales of veterinary antimicrobial agents in 29 European countries in 2014**

## II.III - Alternatives to AMs

Despite all problems associated with AM usage (even if most of them are not related to using AMs strictly speaking, but rather their overexploitation and misuse), the benefits of AMs in animal farming and food production are undeniable.

There is an increased demand for animal source foods and animal protein, particularly in developing countries, and for farmers to be able to satisfy this demand, and answer the food needs of a 7-billion people planet, farmers need to work efficiently and intensively. This implies producing the maximum amount of product with the highest quality possible in the least amount of time, while avoiding any harmful problems to production (Boeckel et al., 2015; Hao et al., 2014; Michael et al., 2014), and indeed, countries such as Brazil, Russia or China have experienced a shift towards vertically integrated intensive production systems, with a high cost-efficiency, in order to cope with this demand (Boeckel et al., 2015).

In this context, by helping to maintain the animal's good health, AMs are a powerful weapon in the farmer's arsenal, either they're used for therapy or as growth promoters. One could even argue they're essential in order to sustain this kind of production system in most contemporary farms. Studies show that, should the general trend continue, AM consumption by animal production is estimated to rise by 67% between 2010 and 2030, 34% of which will be due to shifts in production systems towards intensified animal farming requiring a more intensive AM usage (the rest being due to the increase in the number of animals raised for food production) (Boeckel et al., 2015).

As such, a tighter control regarding the use of AMs can be accompanied with worse production values. At the very least, that will be the first though in a farmer's mind if proposed with regulations or policy changes that limit the allowed amount of AMs to be used in the farm.

It is easy to recommend a reduction in the use of AMs at farm level while contemplating the higher goal of reducing AM resistance and improving human health in the future, but we must not forget the implications of such measures at smaller scales, in the present. Each farm needs to be economically viable and measures that compromise production (and therefore the potential profit or even the ultimate economical sustainability of the farm) without any short-term returns, will be undesirable to most farmers. This does not deny the need for the reduction of the AM prescription, of course, but the solution cannot ignore the current production system's dependence on AMs. Therefore, it can be concluded that there is a need to find an alternative to AMs, which can replace these agent's role supporting the modern animal farming practices.

This need, along with other factors such as the increasing public concern about consuming AM residues in animal meat, the threat of AM resistance or the bans regarding AM use in some countries, has been motivating substantial research for alternatives to AMs in the last two decades, as reviewed by Thacker (2013). These practices, substances, supplements and changes to the production systems make up a whole plethora of alternatives, which differ from each other both in effectiveness and, one could say, popularity.

From these, some are intended to replace AMs as an administration-based substances, such as probiotics, prebiotics, enzymes and AM peptides, acidifiers, plant extracts and nutraceuticals. AM peptides, for example, are able to take effect while avoiding resistance mechanisms, and offer a broader spectrum when compared to AMs, being effective against fungi and viruses too (Thacker, 2013; Wang, Zeng, Yang, & Qiao, 2016). Dietary supplementation with flavonoid-containing plant extracts and yeasts also showed benefits in neonatal diarrhea in calves, controlling the intestinal microbiota (Bi, Yang, Diao, & Tu, 2017). And though there has been some concerns about its safety, spray dried plasma also presents itself as a promising alternative in piglets, particularly in post-weaning phase, offering a perhaps more natural form of protection (Pérez-bosque, Polo, & Torrallardona, 2016). And this is not to mention compounds such as clay minerals, egg yolk antibodies, essential oils, recombinant enzymes and others, which show inconsistent results, offering few certainties when it comes to their viability as AM replacements (Thacker, 2013).

The list of potential alternatives to AMs spreads on, and its presentation would consume hundreds of text pages, which is outside the scope of this dissertation and the goal of this project. Essentially, the compounds destined to be used as a replacement of AMs are numerous, and many of them seem to show benefic effects on animal health. However, while scientific reports often present said beneficial effects, actual comparison against AMs is rarer, and when it is presented, if often reveals that, despite improving animal health and helping to prevent/treat diseases, the alternatives fall short to AMs in terms of power. Though they exhibit a positive effect, often their costs, the necessary amount for treating the animals or even the effectiveness and treatment success rate make them more undesirable than the simple, cheap and proven solution which is resorting to AMs. Furthermore, the unknown risks associated with the implementation of new practices must also be taken into consideration, as they can compromise an AM alternative’s potential. For example, Ontario experienced a porcine epidemic diarrhea virus outbreak in January 2014, in which contaminated porcine spray dried plasma is suspected have been a risk factor regarding the disease introduction and spread. The investigation to assess if the spray dried plasma was, indeed, a risk factor was inconclusive, but the danger regarding the usage of these new alternative methods remains a matter requiring more research, as there is less information available about them (Pasma, Furness, Alves, & Aubry, 2016).

However, not every proposed alternative to AMs is based on administration, and there seems to be some degree of focus on good farming practices and disease prevention that does not rely on the administration of compounds. One of these passive measures, which is often described as a sustainable, intelligent form of preventing diseases and minimize the damages they sow, and a good method to compensate the drop in productivity a cut in the AM usage would bring, is biosecurity.

Biosecurity relies on measures which prevent diseases from entering and spreading within a farm, and it is pointed as a relevant factor towards sustainable agriculture. For practical reasons it is often subdivided in two synergic parts: External Biosecurity, which normally defines all the preventive measures put in place to prevent the breach of pathogens into the farm or the infection of the farm animals, and Internal Biosecurity, which refers to the barriers that prevent the spreading of pathogens within the farm animals, once infection has taken place. The fundamentals of biosecurity as an AM alternative are based on the fact that better defenses against the settling of diseases will lower the frequency of infections, thus reducing the amount of occasions at which treatment is needed, namely with AMs, therefore reducing the overall consumption, as reviewed by Postma et al. (2015).

Biosecurity has already been associated with improvements in pig farming, for example with higher number of weaned piglets per sow per year, and with lower treatment incidences (Postma et al., 2015, 2016), and therefore recognized as a positive (one would say “right”) way to manage a pig holding. For farmers concerned with the implications that an abandonment of AMs would have on the farm profit, investing in biosecurity presents a significant advantage over other alternatives, since not only does it contribute to a good herd health, it also seems to have a positive effect in the farming efficiency, and therefore the business. The implementation of biosecurity measures, and the farm make-over necessary to accommodate these changes, could pose a threatening initial cost, but the long-term benefits would theoretically compensate the initial investment. A “biosecure” farm would have several advantages, not only in health improvements, but also on lower maintenance costs after the first investment, better production and improved animal welfare, which is becoming a growing concern of both farm owners and the consumers.

Another similar method which also tackles the resistance problem through reducing the amount of treatment necessary is vaccination.

Just like good levels of biosecurity, vaccines have been pointed as a way to minimize the prevalence of infections, thus reducing the overall need to recur to AMs and preventing possible dissemination of resistant agents. At the same time, vaccinated animals against a certain pathogen, having some degree of resistance against colonization from that agent, decrease the population of bacteria capable of interchanging resistance elements. Also, a herd vaccinated against a certain set of pathogens allows the use of narrower-spectrum AMs, as the pool of possible agents responsible for the disease, from which the ones the herd is vaccinated against is not a part of, is smaller. More recently, it has also been proposed the use of vaccines that specifically target resistant strains of bacteria, or even resistance determinants, such as resistance conferring proteins, which have been showing some promising results (Lipsitch & Siber, 2016; Postma et al., 2016).

All this considered, although we know alternatives to antimicrobials exist, and some of them are associated with better production parameters, there is a need for further research in order to know if they're viable: that is, to understand if a shift from AMs to any given alternative will not damage the production, thus making it less appealing for farmers to make the transition.

Examples such as Denmark are evidence that it is possible to have very high production while still maintaining a low AM usage, but conditions vary from country to country, and a complete mimicry of the Danish production system, or any other country with low AM usage per population unit such as Norway, Sweden, Finland or Romania, is not possible. The specific characteristics of every nation, regarding the edaphoclimatic context, the geopolitical setting, the financial situation of the country and so many other factors which can influence the production panorama are different in every case. The consequence of this fact is that there is no universally correct system, but rather the most appropriate one in each one. What we can strive to identify are factors which, within said "successful systems", are key to compensate their lower AM consumptions when it regards the production levels.

This problem, however, is not an issue that can be faced using linear reasoning. The way AM usage, AM alternatives, the different production parameters and other pertinent factors are associated with each other weaves a complex web of interdependencies, one which can only be correctly interpreted in a multivariate fashion. Though some linear associations can be found recurring to univariate analysis, assessing relationships of causality based only on that can lead to incorrect results.

If we wanted to understand the physical dynamics of our Solar system, and how the gravity of every celestial body took part in this very clockwork-like dance around the Sun, we would never model it considering every planet, moon or asteroid in pairs. There is a much greater mechanism in motion, which cannot be truly understood unless we consider every nut and bolt. Similarly, a farm is a living environment, where the influence two given factors have on each other cannot, likewise, be truly understood without considering the interactions of every other factor in that relationship, and the effect the latter has on the rest of the system. Only by gazing upon the bigger picture can we identify confounding factors, discriminate direct from indirect associations, admire the synergism and antagonism of certain pairs of variables, and truly understand the dynamic of how all these factor work together and influence each other in a more holistic way.

That being said, modeling in such a holistic way the dynamics between AM consumption, AM alternatives in motion, production and others can provide that necessary insight about how this system works in a farm setting with low AM usage, and which factors might be behind those levels of AMs and production.

## II.IV - Additive Bayesian Networks

As mentioned previously, due to the complex web of interdependencies that exists between AM consumption, biosecurity, production and the other factors in a farm context, resorting to univariate analysis and linear regressions to analyze these associations would be a poor choice. A multivariate technique would be necessary to our objective: to truly understand how all our variables work together, as opposed to understanding how they are associated in pairs.

Even so, there are several multivariate methodologies which could replace Additive Bayesian Networks (ABN) as our technique of choice, such as factor analysis or principal component analysis. Several reasons justify the use of ABN over other possibilities. However, before indulging in them, a brief introduction to ABN will be given.

Bayesian network modeling is a multivariate machine learning technique, particularly indicated to process big, complex (one would even say “messy”) datasets, composed of many inter-dependent variables. It is nowadays used widely in biology and epidemiology, more specifically in fields such as systems biology, HIV and influenza research, complex disease systems analysis, and others. Unlike many other techniques, it attempts to identify the most optimal model, returned as a graphical outcome, indicating the dependency structure between all the possible variables, effectively eliminating any researcher’s biases when it comes to variable selection (Lewis & Mccormick, 2012; Pittavino, Lewis, & Furrer, n.d.).

This technique is, in its true nature, exploratory, and rather than modeling the studied variables as a causal model, potentially coercing the dataset in undesired ways, it focuses on the distinction between direct and indirect associations, commonly referred to as structure discovery, and determining the network graphical model which best describes the web of interdependencies between all of the dataset’s variables. This allows an easy, though ample, understanding of the dataset’s association structure.

ABN effectively manages to analyze the dataset in a much broader perspective than other methodologies, considering the database as a whole, without focusing on any particular variable over others (Firestone et al., 2014; Lewis, Brülisauer, & Gunn, 2011; Pittavino et al., n.d.).

The graphical model produced comes under the form of a Directed Acyclic Graph (DAG), a visual representation of the joint probability distribution of the associations’ network, where every variable is presented as a node, and every association found is represented as a directed arc between the two related nodes. An example of these graphs is given in Figure 6. Being a directed graph, no cyclical paths are allowed. However, in many real-world problems, it is reasonable to assume the data cannot distinguish between equivalent network structures. This means there is likelihood equivalence, and as such, two equivalent network structures have the same probability of resulting from the same database, in Bayesian network modeling, and, because of that, we cannot ascertain causal relationships based on the direction of the arcs in a DAG. Ultimately, this means that in Bayesian network modeling it is possible to ignore arc direction. What is noteworthy is the presence or absence of arcs between variables, indicating direct association, and not the possible direction of the arc (Heckerman, Geiger, & Chickering, 1995; Lewis & Mccormick, 2012).

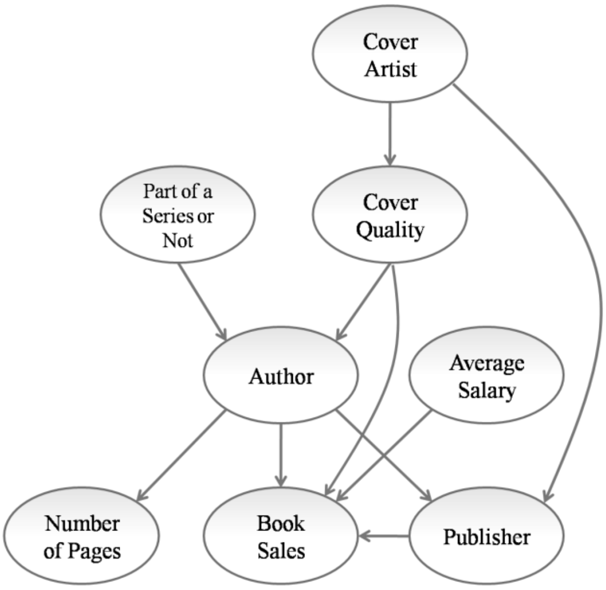


**Figure 6 - Example of a DAG, of a hypothetical ABN analysis on book sales. Every arc depicts an association. Note there are no cyclical paths**

The choice of ABN over other multivariate analysis techniques can be justified by a few reasons.

First, ABN is a methodology well suited for a dataset as complex as the one used in this study, with many possible interdependencies existing hidden in the midst of over 50 variables, each with 160 data entries (Pittavino et al., n.d.). Even more than discovering possible associations between variables, ABN allows us to not only know, but actually represent, graphically, the probabilistic structure in which the different factors we're analyzing organize themselves into. This way, a very good foundation is built which we can use when advising farmers. In fact, the output of the analysis is actually able to be interpreted directly by the farmers and farm owners, breaking down a barrier of communication between them and the scientists.

Second, the result produced by Bayesian network modeling assumes every single variable in the dataset as potential dependent and explanatory variables, as they are submitted in the analysis computer script. Other methodologies require the specification of one or more explanatory variables, against which the dependent ones will be matched, searching for a significant association between them. This can introduce a small researcher bias, as many possible, and potentially relevant, associations between the dependent variables are never investigated, and by doing that, the research field of vision narrows, and the big picture is lost. (Firestone et al., 2014; Lewis et al., 2011)

That is not the case with ABN. While other techniques seek to model according to expert opinion, so inference can be performed from that (including this way some degree of subjectivity), Bayesian Networks perform structure discovery through empirical analysis, in order to deliver an exploratory analysis and identify the optimal DAG for any dataset (Lewis et al., 2011). This way, there is not any focus on any particular variable of interest, since every possible association is presented, and the result is a much more holistic approach.

Finally, one of the main features of structure discovery processes is that they distinguish between direct and indirect associations, allowing us to differentiate, for example, factors that affect production directly (and are, therefore, the most suitable ones to focus in, to maintain the production levels if AM usage was to be reduced) and factors that only affect it indirectly, and can only have an significant effect through other variables (Lewis et al., 2011). It would be easy resorting to linear models and map all the possible associations of pairs of variables, but that would result into an unrealistic, potentially overestimated network of dependencies, as variables only indirectly associated would still reveal some degree of interdependency. Moreover, it would not be a multivariate approach.

To illustrate this, imagine two individuals holding a long wooden board, with a marble in the middle, each person representing a variable. In a simple linear system like this, depending on how high each of them hold their end of the board, it is easy to predict which direction the marble will roll towards. But what if instead of two individuals, there were three? And instead of a wooden plank, they were holding a triangular sheet of linen, not only different in shape, but malleable, able to form creases, each end being held by a different person? In a system like this, considering three variables, the complexity increases substantially, and it is more difficult to make predictions. It is then clear how chaotic it would be if instead of three, there were over 50 individuals holding the edges of the linen sheet.

In the end, this is a very complex and interdependent system, one that is not linear. The variables weave a delicate equilibrium, very intricate and complex in nature, which cannot be truly understood unless we use an approach that is more holistic in nature, and accounts for the indirect pathways in which the variables organize themselves and influence each other. In that regard, ABN satisfy those requirements.

That is not to say Bayesian Networks have no disadvantages.

The biggest problem with ABN, something that derives from its structure discovery, is the computational requirements, which are very demanding. The search method employed to determine the optimal DAG, the one which best describes the dataset, has a great influence on this matter, but to analyze every single possible DAG from a certain dataset in order to find the one best fitting, unless the variable number is very small, is impossible, and more than that, impractical to do. Even if there are available computational resources, a dataset containing only 10 variables would have about $10^{18}$ possible DAGs to investigate. Other technical obstacles are present, such as the need to estimate model parameters and goodness of fit measures by simulation, which is computationally very demanding (Lewis et al., 2011).

Also, when dealing with datasets comprising a large number of variables, there is the problem of over-fitting. To correct this, we need to apply some corrective procedures, such as majority consensus, or pruning, even though they are used originally in other techniques. However, these are, in the end, corrective measures, *ad-hoc* approaches to compensate a flaw in the technique. Quoting Lewis, Brülisauer and Gunn (2011), “The unresolved challenge is to summarize the results from many different networks in a statistically robust and interpretionally cohesive way”.

One could also consider, for absence of a better term, the “lack of numbers” a disadvantage. Unlike many other methodologies, ABN does not provide p values, or any other statistical strength measures which can be used to compare results or evaluate the strength of the associations.  Values translating the model goodness of fit exist, but they are of no use in most contexts other than determining the best DAG out of all the possible ones. It should be said, though, that providing the mentioned values is not the goal of ABN.

This is a technique which focuses on structure discovery and determining optimal graphical models out of empirical data, with a high level of objectivity, rather than one that aims to prove hypothesis and give evidence of their strength (Lewis et al., 2011). Classical, or frequentist, approaches use probability to ascertain the frequency of certain data happening over a great number of trials, relying in confidence intervals and hypothesis testing as its main ways to draw inference. Bayesian approaches are part of a different school of thought, making use of what is described as subjective probability. Instead of determining values to translate the probability of a future event happening, such as a certain lottery ticket being drawn, Bayesian Methods express a degree of belief that the same event will happen. That is, they express the uncertainty of assertions regarding unknown parameters in terms of probability, using that probability to express a degree of belief over a certain statement, weighting prior knowledge about the model parameters, before the data is observed, and the insight gained with the analysis of the data. This is achieved through the posterior distribution (Glickman & Dyk, 2007).

Being a mostly exploratory technique also has its downsides. The outputs from ABN are, indeed, not causal models, as they only provide evidence of the statistical dependencies present in the database. In this regard, they are comparable to traditional modeling techniques, such as linear regression, although it should be said that, unlike said approaches, Bayesian Networks are much more holistic, and the produced models are multidimensional, considering all the available variables at once and never determining variables of interest, and all the biases that can come from that (Firestone et al., 2014).

There are three inter-related steps to structure discovery in ABN: parameter learning, network scoring and structure learning.

Parameter learning refers to the calculation of all the model parameters based on cumulative knowledge gained, as every new observation in the dataset is analyzed. Given a network structure of a certain Bayesian network, we can fit this early network to the dataset based on certain priors, which allow for a posterior distribution to be calculated.

Consider a very basic model with two associated binary variables, X and Y, where both can be either T (true) or F (false). A graphical representation of this network would be simply two nodes (X and Y) connected by an arc, signifying their association. To this Bayesian network, there would be three free parameters

1. The probability of X being true: P(X=T)
2. The probability of Y being true if X is true: P(Y=T|X=T)
3. The probability of Y being true if X is false: P(Y=T|X=F)

After determining priors for each of the probabilities, the next step would be to calculate posterior distributions based on observed data, thus updating the prior knowledge and building upon it. For example, should a single datum be observed, where X was true and Y true as well (X=T,Y=T), both the mean probabilities of observing X=T and Y=T|X=T would be updated. It is this continuous procedure of updating previous knowledge with every new observation that is known as parameter learning (Lewis et al., 2011).

The network scoring is deeply tied with the model selection, a step crucial in structure discovery and which was assumed in the previous example, where X and Y were said to be associated from the beginning.

The network score of a model translates its goodness of fit, and it is given by the marginal likelihood of the data, considering the said model, or rather the log marginal likelihood, to avoid numerical accuracy issues. This value is equivalent to the Bayes factor if the model's priors are equal, and it allows the comparison between models to select the most fitting one (Lewis et al., 2011).

The most commonly used network scoring metrics are the Bayesian Dirichlet equivalence metric (BDe) and K2 metric, from which the first was the one used in this analysis. Unlike the K2 metric, which assumes all parameter priors are uninformative, the BDe metric defines the priors for each variable in terms of an equal sample size, meaning the number of assumed prior observations at each node is taken into consideration, so that the priors are distributed between the individual parameter priors at each node. This way, the sum of all the prior observations across all parent combinations for each node, considering that each pair of associated variables defines one as a parent and the other as a child – will result in the same sample size. The BDe metric is also likelihood equivalent, which, as mentioned before, turns the direction of the arcs in the DAG irrelevant, not translating relationships of causality. The associations revealed with the ABN can later be biologically interpreted, and the model can be later complemented with linear approaches, as to investigate causality (Heckerman et al., 1995; Lewis et al., 2011).

Finally, structure learning refers to the search for the optimal model, making use of the DAG's network score.

It is impossible to analyze all possible DAGs generated from a certain dataset, in order to identify the one with best network score. Moreover, it could happen that a good number of different DAGs have the same network score. From this issue arises the need to define a search method do single out the best fitting model.

One option is to run a heuristic search, as proposed by Heckerman et al. (1995). The fundamentals of this search method are simple. From a randomly constructed DAG, its network score is determined. After that, a local search is executed, considering all the possible DAGs exactly equal to the first one except for a single "perturbation", which can be an added arc, a removed arc or an arc direction reversal. If one of the slightly altered DAGs (compared to the original one) has a higher network score that the original one, that DAG is then defined as the current working network, and the search across all "perturbed" networks is repeated for that better DAG. The process continues until no network with higher score is found, signifying it is not possible to further improve the score, as the optimal local network was found (Lewis et al., 2011).

To build a robust method, several local searches should be performed and then combined into a majority consensus network, which would contain the arcs featured in the majority of all the local searches (Lewis et al., 2011).

However, in datasets containing less than 20 variables, as it was our case, the most robust search method to identify the globally best DAG is generally an exact search. To adjust for over-fitting, this network would later need to undergo a parametric bootstrapping process, which is a very demanding procedure, computationally speaking. Heuristic searches deal with the impossibility of examining all possible DAGs through the process previously explained. However, a different approach was suggested by Friedman & Koller (2003), where, instead of sampling across all the possible DAGs, the search was instead made through sampling across node orderings, where a node ordering consists of a set of integers from 1 to n, where n represents the number of variables in the dataset. This way, DAGs are considered consistent with orderings where for each node in the order, they follow their parents, and not the opposite. Each order is, effectively, a collection of DAGs, as they can be imagined as a general restriction for DAGs which to not respect certain parent-child associations, and a DAG can also be consistent with several different orders. Because of this, a certain bias is introduced when it comes to the estimation of posterior probabilities, as stricter structural features, being consistent with a higher number of orders, are favored. Despite this, searching across orders in a dataset with n variables has the advantage of having n! orders to search, instead of $2^{\binom{n}{2}}$ different DAGs (Pittavino, Lewis, & Furrer, 2016).

To search across orders, aside from the Friedman & Koller (2003) method, one could also adopt the approach suggested by Koivisto & Sood (2004), where despite being a order based method, an exact, exhaustive search is performed, instead of sampling across orders. This method has the advantage of being able to find the optimal DAG for the dataset as well as the posterior probabilities for structural features, such as individual arcs. It is, however, only executable with a smaller number of variables, up to 20 when using a typical desktop computer (Pittavino et al., 2016).

While searching for the network structure with the maximum goodness of fit, the ideal scenario would be to do so without any kind of limitation, specifically when it comes to the number of parents.

In a Bayesian Network, variables linked through an arc are associated in parent-child relationships, where the variable from which the arc originates is named the parent and the one which receives the arc is named the child. To conduct a search without any kind of limitation regarding how many parents a variable can have can be computationally impossible, and the run time for the function which identifies the optimal DAG increases greatly with the number of parents. To deal with this, there is a need to establish a maximum number of parents when looking for the best fitting network (Pittavino et al., 2016).

One currently suggested approach, which was the one used in this project, is to start off with an *a priori* limit of one parent per node (variable), determine the optimal structure and ascertain its network score. Then, repeating the process while increasing one by one the maximum number of parents per node allowed, it will come to a point where, despite the increase in the parent limit, the optimal DAG identified suffers no further change, and so does the network score. At that point, the DAG with the maximum goodness of fit has been found, and that maximum parent limit can be set (Pittavino et al., 2016).

Finally, once the best fitting model, with the largest, least negative log marginal likelihood has been identified, an adjustment to compensate for the overfitting this metric can easily induce is in order. Indeed, the log marginal likelihood can many times include associations in the model which, should the study be repeated many times, would only be retrieved a few times. One common method to adjust this overfitting is to use parametric bootstrapping, one of the most defensible solutions to solve this issue, though very computationally demanding (Pittavino et al., 2016).

The principle behind parametric bootstrapping is repeating the model to identify the most supported structures. Put simply, once the optimal model has been selected, several data sets are simulated from it, with the same size as the original database. Then, the analysis is repeated for every simulated dataset, and the retrieved structures (the arcs between variables) are accounted for. The reason behind why this technique is so demanding is because the model search needs to be repeated for every simulated dataset, and several hundreds of simulations are required to build robust results (the suggested number of repetitions is often between 5000 and 10000 simulations), hence the need to maximize the search process (Pittavino et al., 2016).

To run the bootstrapping, a Markov chain Monte Carlo (MCMC) software must be used to make the simulations, which must be fed with the model definition (under the form of a BUG file) containing the structure of all the dependencies in the model, a computer script defining the MCMC simulation, a file setting random number seeds for every simulation and the probability distributions for every parameter in the model, namely the marginal parameter distributions (Pittavino et al., 2016).

Monte Carlo methods are simulation-based methods, which, given a specific probability distribution, generate several plausible values from the same distribution, which in the context of Bayesian data analysis is often the posterior distribution. Monte Carlo samples carry the same information about the posterior distributions as analytically computed plots, allowing us to take the same inferences regarding the mean and the standard deviation from either of them.

To acquire samples from posterior distributions, one of the most important techniques is MCMC, which makes use of a Gibbs sampler to cycle through a set of conditional posterior distributions (which have less unknown parameters, making them simples to simulate), sampling from each one conditional on the most recent draw of the remaining parameters (Glickman & Dyk, 2007).

The marginal parameter distributions, which need to be estimated in order to perform the bootstrapping, can pose some problems, as this assumes the data has sufficient information to accurately estimate the density, which can be roughly visualized as the shape of the distribution, for every parameter in the DAG. For that reason, the marginal densities estimated should be examined visually, and it should be investigated if the probability densities – the approximated area under the curve of the marginal densities - are roughly equal to one. Should they be so, we can assume they are reliable (Pittavino et al., 2016).

Once the bootstrapping is completed, the number of arcs recovered during the simulations can be examined, as to assess if the model did overfit the data. The optimal model can then be pruned to cross out the overfitted arcs.

Similarly to the approaches in majority consensus networks, one option to do this is the removal of every arc which was not recovered in more than 50% of the simulations, the usual cutoff for structural support of arcs in ABN. This number is not, at all, comparable to the 95% usually used in statistics, and is a reasonable threshold above which arcs can be considered real. This can be understood considering an example given by Fraser Lewis in the abn package vignette: imagining a basic system with two covariates and only one node, the “response” variable, where the later is almost completely dependent on the covariates, which in their turn are almost collinear. In this case, only one arc would be needed in the model, and running several optimal model searches would return, in 50% of the cases, an arc from the first covariate to the response variable, and in the other 50% cases, an arc from the second covariate to the same variable. Even though these variables are associated almost 100% with the response variable, their arcs would never surpass 50% of support, which tells us why 50% is, in fact, a very reasonable threshold.

Regarding the pruning of the arcs, there are two options, which are mostly case-specific and are completely up to the researcher. One of them is considering the arc direction, and maintaining only arcs with more than 50% recovery rate in each direction. The other (which recalling the likelihood equivalence discussed earlier, invokes the fact that the data cannot, most often, discriminate between opposing arc directions) consists in summing the number of times any arc, between the same variables, was recovered in the simulations, regardless of direction, and eliminating the ones with less than 50% overall support.

After pruning the arcs due to over-fitting, the final model can be obtained, and the parameters can be examined.

## II.V - Complementary analysis

As it will be explained further into this dissertation, two complementary analysis were executed, as to confirm the results obtained with the Additive Bayesian Networks. Those were Discriminant Analysis and Fast and Frugal Trees.

Discriminant Analysis is a technique which, by using variables which discriminate the dataset between two (or more) groups - the discriminatory variables - develops a function, or equation, able to specify the difference between the two divisions (and classify any future additional data accordingly). If we visualize the data represented in a scatterplot, with two variables of interest as the orthogonal axes and the discriminatory variable represented as, for example, differently colored dots, this separation would be achieved with the determination of a function which would divide the bidimentional plane in two areas, corresponding to the two groups, in which every observation in the plot inside any of the groups would be classified (or “discriminated”) as belonging to that group. Technically, this function would be determined by the projection of the observations, resulting of the linear combination of the original variables, onto a new axis, Z: essentially an oblique axis intersecting the X axis at the angle θ that would allow to maximize the ratio between the between-group sum of squares and the within-group sum of squares, thus allowing the best separation between groups (Sharma, 1996).

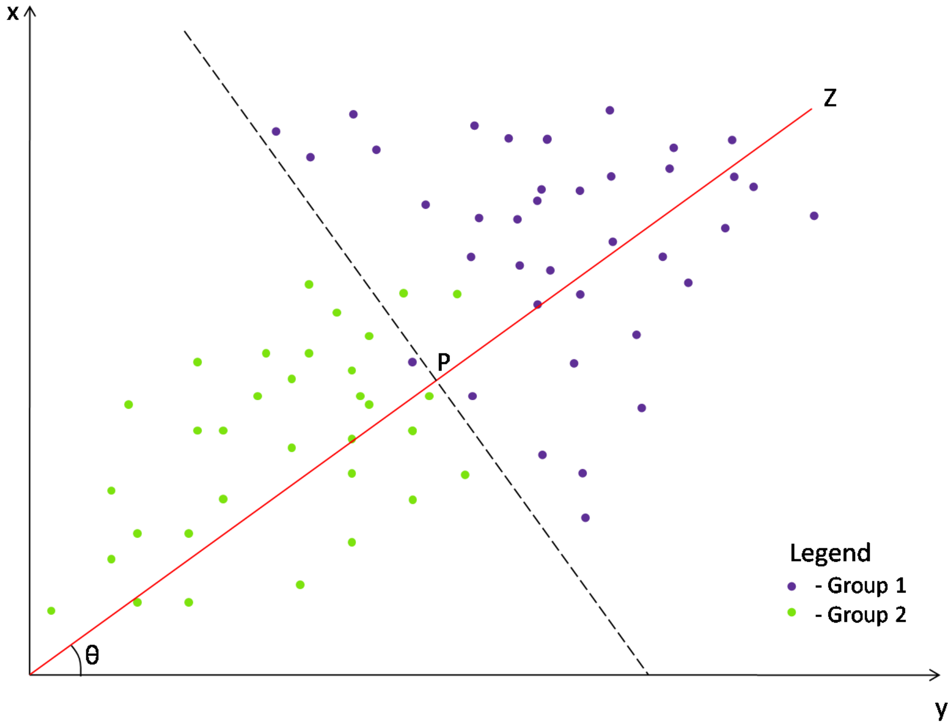


**Figure 7 - Example of a hypothetical discriminant analysis represented in a scatterplot. Z represents the new axis in which the variables are projected and separated according to the discriminant variable (here represented by the colour of the dots) as they fall before or after the cuttoff value (P). The challenge here consists in determining the angle θ which allows the best and most truthful separation**

Multiple-group discriminant analysis works in a similar way to two-group discriminant analysis, except it may require more than one discriminant function, depending on the dataset, working together to classify the dataset into the desired number of groups. Having multiple discriminant functions, even if the axes represented by them (which need not to be orthogonal to each other) do not discriminate the dataset effectively individually, they do so simultaneously (Sharma, 1996).

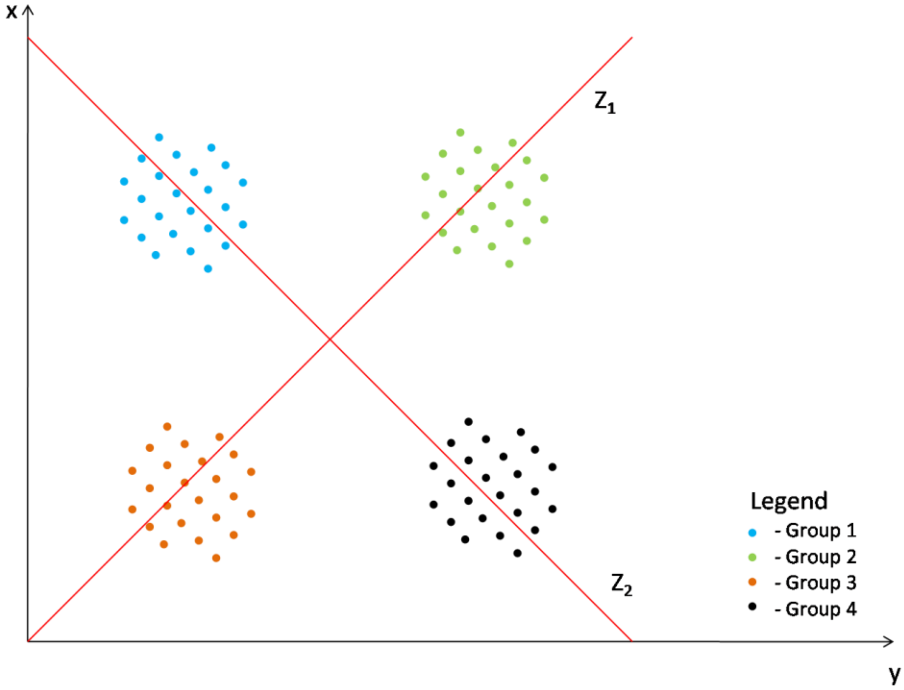


**Figure 8 - Example of a hypothetical multi-group discriminant analysis.**
**Note how the discriminant function $Z_1$ is able to properly discriminate groups 2 and 3, but is unable to do so with groups 1 and 4, and $Z_2$ effectively separates groups 1 and 4 along its axis, but cannot do it for groups 2 and 3. However, between the two of them, they achieve a perfect separation of the dataset.**

Once the discriminant functions in a discriminant analysis have been identified, the next step is to determine criteria for the classification of eventual new observations, amongst the groups. This is achieved by dividing the discriminant space in mutually exclusive regions, and using the discriminant scores (the projection of the new variables onto the discriminant functions) to analyze in which division do the new observations fall into. To do so, the discriminant scores of the new observations are compared to the discriminant scores, which act as cutoff values for the different zones of the discriminant space, results which are combined amongst all the discriminant functions to determine the effective "group" that observation belongs to (Sharma, 1996).

It is, then, possible to program an algorithm which is able to classify a certain observation amongst the defined groups, given the parameters of all relevant variables.

This can be of interest for this project, as if a certain variable, such as biosecurity, for example, acted as a good discriminant variable, effectively dividing the dataset in terms of AM consumption, or even production parameters, that would prove there was truly an effect of the discriminant variable (in this case, the biosecurity) in terms of production, or AM consumption. Similarly, other variables could be used as discriminant functions, as to verify if they had an effect on the other variables.

Fast and Frugal Trees (FFT) are binary decision trees where, at every decision level, there is a least one exit leaf, meaning that for every checked cue (every binary criterion which composes the tree), one of its possible outcomes leads to a decision, while the other continues the tree (or not, in case of the final binary cue) (Martignon, Vitouch, Takezawa, & Forster, 2003).

These trees are widely used in situations where a fast but accurate decision making is required, such as triage at hospitals, HIV testing, amongst others, and can be used a prescriptive and descriptive model (Luan, Schooler, & Gigerenzer, 2011).

FFT are composed of sequentially ordered cues, normally composed of if-then clauses. Once a decision must be taken with the tree, all which is required is to match the necessary parameters with the criteria in every cue, eventually reaching an exit leaf which dictates the decision. When matched against other decision making trees, FFT have the minimal number of exits, as opposed to full trees. They can effectively be considered trees pruned to the maximum extent, which give them advantage over other decision trees in terms of frugality, simplicity of decision-making, and also robustness, since FFT' simplicity make them less susceptible to overfitting, unlike very complex models which tend to fit known data very well but can then have less accuracy when processing new samples (Luan et al., 2011).

To the investigation of associations between production, AM consumption and AM alternatives such as biosecurity or vaccination, FFT can be of use much like Discriminant Analysis: should a tree be determined where factors such as biosecurity, vaccination or AM prescription as cues could determine the level of production in a farm, that would, just as in the Discriminant Analysis example, prove an association between the variables, and the effect they have on each other.

## III – Objectives

This project aimed to analyze data from 160 Danish sow herds and model the interaction between the variables of interest, in order to identify which factors are associated with production in this specific setting, and how they relate to AM levels and other pertinent variables.

The chosen methodology to attain that answer was to use Additive Bayesian Networks, a machine learning multivariate approach which, having as a final output a directed acyclic graph describing the interactions between all the variables in the dataset, will provide that insight in a clear and easy to interpret way.

Potentially, variables (such as biosecurity or vaccination), could be identified as positively influencing production, while not relying on AMs, which then could be explored as sustainable AM alternatives. The visualization of the web of interactions between the variables would also allow for a better insight regarding the direct and indirect effects of AMs and their alternatives. This way, plans could be devised to mitigate the negative impacts in productivity due to a reduction in AM consumption, exploring alternative pathways in the network which would benefit the production.

In addition, this project also aimed to look independently into the sow and the weaner-related part of the production, as the two of them have substantially different approaches when it comes to AM usage. For that reason, a sow model and a weaner model will be built separately, including only variables of the respective sector of the production, with a few exceptions, as described below.

# IV – Materials and Methods

## IV.I - The dataset

For this study, 364 Danish herds were selected according to three criteria:

1. The herd was a sow herd (it had at least sows and weaners)
2. The number of sows present in the farm was, at least, 100 animals
3. The farm in question had productivity data available in the SEGES database (further information regarding SEGES will provided in page 38)

Part of the necessary information was supposed to be obtained through a telephone interview with the farm owner. From the selected 364 herds, 88 could not be contacted when the interview was attempted. A total of 36 farmers refused to participate when contacted, claiming they did not have time. 15 had sold their farms since the input of the data, 28 were not interested in participating in the study and 21 gave other reasons for why not to participate.

After those removals, 176 farms remained available for the study, from which 16 were excluded due to the farmer not completing the interview. In the end, a total of 160 farmers participated fully in the interview.

Hence, the dataset comprised data regarding the use of AMs and vaccines, as well as productivity for 160 Danish sow herds for the year of 2014, and with biosecurity data from 2015. This was the result of merging several databases and acquiring information directly from the farmers. Amanda Brinch Kruse was responsible for the collection and compilation of the dataset.

The study variables are listed and summarized in table 3, and described more thoroughly afterwards.

**Table 3 - List of variables in the dataset**

| **Farm data** | |
|---|---|
| • CHR Number | • Region |
| • Adherence (or not) to the SPF system | • SPF class |
| • Number of employees | • Employment of foreign employees |
| • Exporting (taking place or not) | • Years of experience of the farmer |
| • Age of the youngest building | • Age of the oldest building |
| **Herd data** | |
| • Herd type | • Pig total |
| • Number of sows | • Number of weaners |
| • Number of finishers | |
| **Vaccination status** | |
| • Vaccination against PCV2 | • Vaccination against PRRS |
| • Vaccination against *Actinobacillus Pleuropneumoniae* | • Vaccination against *Mycoplasma hyopneumoniae* |
| • Vaccination against *Lawsonia* | |
| **Antimicrobial prescription (AMP) data** | |
| • AMP for sows and piglets (total) | |
| • AMP for sows and piglets for gastrointestinal indications | • AMP for sows and piglets for respiratory indications |
| • AMP for sows and piglets for individual treatment | • AMP for sows and piglets for group treatment |
| • AMP for weaners (total) | |
| • AMP for weaners for gastrointestinal | • AMP for weaners for respiratory |

| | |
|---|---|
| indications | indications |
| • AMP for weaners for individual treatment | • AMP for weaners for group treatment |
| • AMP for finishers (total) | |
| • AMP for finishers for gastrointestinal indications | • AMP for finishers for respiratory indications |
| • AMP for finishers for individual treatment | • AMP for finishers for group treatment |
| **Production parameters** | |
| • Mortality in sows | • Feed used for sows |
| • Farrowing percentage | • Percentage of sows unsuccessfully inseminated |
| • Litters/sow/year | • Piglets per litter |
| • Stillborn piglets per litter | • Days suckling pigs stay with the sow |
| • Mortality until weaning | • Weaned pigs per litter |
| • Mortality in weaners | • Weight gain for weaners |
| • Feed conversion ratio | |
| **Biosecurity scores** | |
| • Total biosecurity score | • Internal biosecurity score |
| • External biosecurity score | • 6 internal biosecurity subscores |
| • 6 external biosecurity subscores | |

Table 3 - List of variables in the dataset (continuation)

The data from the different datasets was merged according to the CHR number.

CHR, which stands for Central Husbandry Register, is the Danish Ministry of Environment and Food central database, operating since 1992, which plays a very important role in animal registration and identification in Denmark, and which has been approved by the European Commission as fully operational cattle registration database in 1999. All herds with cattle, pigs, sheep, goats and commercial herds with deer, poultry, foxes, minks, fish, mussels and crayfish have their data stored in this database. Within this database, each holding is assigned a holding number, the CHR number, which acts as a farm identification number and to which are connected other data, such as the address of the holding and geographical co-ordinates, the keeper and the owner’s name, address, contact numbers and VAT (Values Added Tax) or CPR (Danish Personal Identification Number) numbers, as well as the number of animals and veterinary events, cattle movements (of pigs, sheep and goats), screening results for transmissible spongiform encephalopathies and Salmonella, medicines used, and others. Since 1998, data regarding each individual head of cattle has been stored in the system, including the animal identification, species, breed, age, offspring information, allocation of ear tags every movement and veterinary event (Danish Veterinary and Food Administration, 2015b).

The CHR number is a very important tool in epidemiology, allowing easy traceback of diseases, investigation of infection dissemination, to name a few, while also preventing the spreading of diseases. This happens because it is possible to impose movement restrictions on certain holdings. Furthermore, it is also not allowed for Danish farmers to move animals out of the holding when animal identification and registration has not been respected, as the necessary documents will not be issued. Abattoirs must also verify if the animals are correctly recorded in the CHR before slaughtering them.

The information recorded into CHR must be updated yearly, as every farmer is asked to confirm all the registered information, including the number of animals, and correct it should there be any need.

Many of these data are of public access, including the name and address of the owner and the farmer, if the farm has any veterinary problems, what is the level of danger regarding *Salmonella*, the species being farmed and even the geographical location of the farm, available by satellite image. All this information, and more, are accessible to any consumer with the CHR number of a farm, using the website chr.fvst.dk, effectively allowing consumers to truly know what they eat (Danish Veterinary and Food Administration, 2015b, 2016).

Accessing the CHR number, it was possible to locate each farm and identify the Region it belonged to, of 4 different possibilities, all of which represented in Figure 9: North, Central and Southern Region of Denmark (respectively Nordylland, Midtylland and Syddanmark) and also Zealand (Sjaelland) – which included as well the territory corresponding to the fifth official region of Denmark, the Capital Region of Denmark (Hovedstaden). The herd size variables were also obtained through the CHR number, namely the Number of Sows, Number of Weaners, and Number of Finishers. The first refers to the number of sows present in each farm per year, the number of weaners consists in the number registered pigs between 7 and 30 kg, and the number of finishers refers to the registered pigs with weights superior to 30 kg. The variable Pig total translates the sum of the three previous variables.

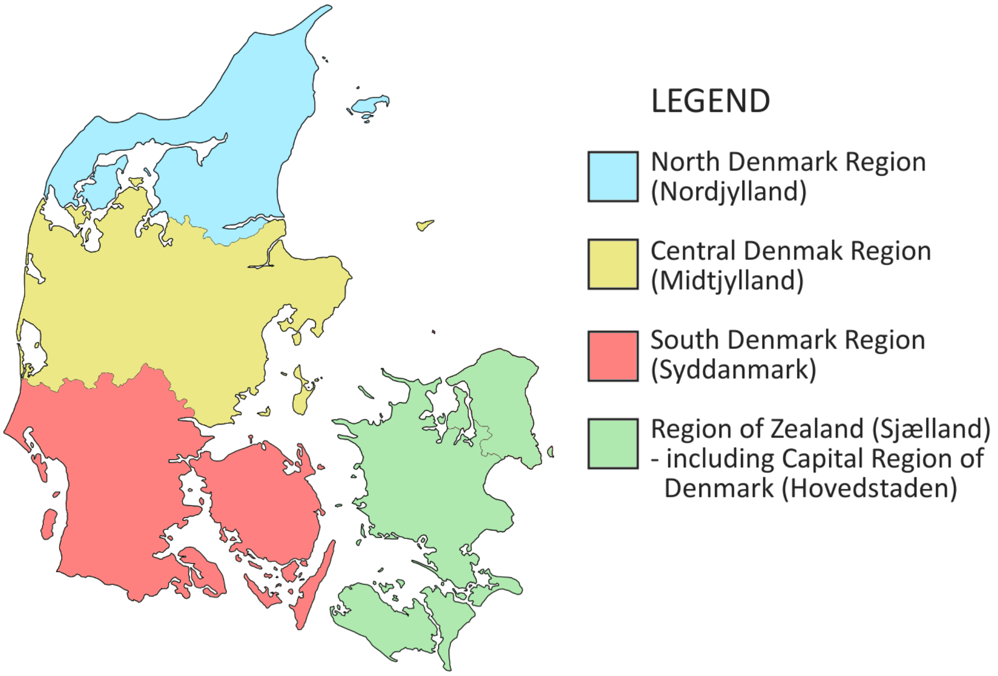


**Figure 9 - Regions of Denmark**

**Adapted from an original picture, by Jarke, licensed under CC BY-SA 3.0**

From these parameters, the database farms were classified with a Herd type referring to the number of finishers present in the farm, which informs us the kind of production it undertakes (whether it is a farrow-to-finisher full-line production farm, or a farm that sells the weaners to be fattened in other farms). As such, each farm was classified with one out of three herd types:

- Full-line: Farms with a number of registered finishers superior to 1,5 times the number of registered sows
- Remainder: Farms which, as opposed to ones where finishers are the main focus of the production, have them as a remainder of the main production (for example, pigs that could not be sold between 7 and 30 kg because of a hernia), and though there are registered finishers in the farm, they are less than 1,5 times the number of sows.
- No finishers: Farms with no registered finishers, meaning all of them are either sold or moved to another CHR number owned by the same farmer.

To explain the origin of some other data, a small introduction about SEGES is first needed.
The Danish Agriculture & Food Council is an organization which came to exist from the merging of five former ones, namely the Danish Agriculture, the Danish Bacon and Meat Council, the Danish Agricultural Council, the Danish Dairy Board and the Danish Pig Production. It represents the Danish agricultural and food industry, promoting its political influence, handling the professional interests of its members and implementing research

programmes regarding pertinent subjects, such as veterinary issues and animal health and productivity (Danish Agliculture & Food Council, n.d.).

The Danish Agriculture and Food Council is the owner of SEGES, a company which is the research and innovation centre for the Danish Agriculture Advisory Service. The purpose of SEGES is to offer farmers the tools, the knowledge and the advisory, to help them maximize their productivity and improve their business, while taking the animal welfare and the state of the environment in account to allow a healthy and sustainable farming. This ensures a bridge of communication between the latest scientific findings and the farmers, who then have access to recent and trustworthy information, to help them improve their practice. To build its knowledge, SEGES participates in many projects, both national and international, and maintains partnerships with research institutes, public authorities and private companies from all around the globe. It also develops and maintains IT tools and accounting programs used by farmers to manage their businesses, and are responsible for many research and development projects specifically regarding pig farming (Danish Agliculture & Food Council, n.d.; SPF-Danmark, n.d.-b).

Data regarding the farm's production parameters, as well as the SPF class, for this project was taken from the SEGES's database. The productivity data is obligatorily collected from the farms, but farmers can choose whether to allow SEGES to have access to their data or not, so they can write their report. This also played a part in what farms were selected for this study. Nevertheless, the farm's productivity values where crossed with the previous data according to the CHR number, adding the following variables to the dataset:

- Mortality in Sows, given in percentage
- Feed used for sows
- Farrowing percentage
- Percentage of sows with non-successful inseminations
- Number of litters per sow per year, in average
- Number of piglets per litter, alive at birth, in average
- Number of days the suckling pigs stay with the sow, in average
- Mortality until weaning, given in percentage
- Number of weaned pigs per litter, in average
- Mortality for weaners, given in percentage
- Weight gain for weaners, in average
- Feed conversion rate, in average
- Whether the farm was part of the SPF system or not
- SPF class, more specifically the color code regarding the safety level in the health status

The Danish SPF system is a health programme for pigs that has been developed over the last forty years by SPF-Danmark, the world's largest transporter of pigs. The system is based on a certain set of standards with the herd health in mind, regarding protection against infectious diseases, health inspections to prevent dissemination of diseases through trade and transportation, emission of health status declarations to clarify health statuses for pig buyers intending to acquire animals, live pig transactions being based on written agreements agreed upon by both parts and animal transportation executed in specifically designed vehicles with high biosecurity levels.

For a herd to be declared as being part of the SPF system, the farmer must sign a written agreement, accepting to follow the SPF health rules regarding protection against infection, health inspection, purchase of pigs with a known health status and transport of pigs to the destination herd in approved SPF vehicles (SPF-Danmark, n.d.-b).

The declaration for a herd which is part of the SPF system contains a health status, supplementary information and conditional status.

The supplementary status information contains data regarding any health condition or specification relevant for the sale that has no direct influence on the health status, such as certain aspects of the production (outdoor herds, pigs received via approved vehicles, safe farms of origin, for example), salmonella records or edema. The conditional status refers to temporary restrictions on sales, such as sales bans, or anomalies in the herd health, such as suspicion of an infectious disease or clear deviation from the SPF health rules.

As for the first part, and the one more relevant for this project, the health status consists in a code, composed of three elements, a color, a SPF designation and suffixes, which translate, respectively, the safety level, the collective designation indicating the farm is free from the SPF diseases, and eventual corrections stating particular SPF diseases the farm is currently not free from. The seven diseases known as "SPF diseases", which are expected to be absent from the farm, except when mentioned by the suffixes, are:

- Enzootic pneumonia by *Mycoplasma hyopneumoniae*
- Pleuropneumonia by *Actinobacillus pleuropneumoniae*
- Atrophic rhinitis
- Dysentery
- Mange
- Lice
- PRRS

The color code which indicates the safety level has two levels: red and blue. Red herds are herds possessing the highest level of safety, which are either nucleus belonging to DanAvl (a brand owned by the Danish Pig Research Centre) or multiplier herds. The Blue level corresponds to herds ranked with the second-highest level of safety, namely the production herds that focus on weaners.

Using the SPF system ensure the farmers a higher level of safety when buying pigs from other sellers, offering those producers all the relevant information regarding the herds to be bought from, dissipating any doubts or mistrust regarding significant undesirable health problems to be found later on, after the purchase is made (DanAvl, n.d.; SPF-Danmark, n.d.-a).

As such, and in conclusion, herds from farms that are a part of the SPF system are considered better herds, in terms of animal health. For the purpose of this project, the data corresponding to the SPF health status was retrieved, from SEGES as well, which resulted in two variables: Adherence to the SPF system, a Boolean variable reflecting whether or not the farm belonged to the SPF system of not, and SPF class, more specifically the SPF Health Status' safety level, corresponding to the color code described previously. For this second variable, three values were possible. Farms with the highest and second-highest level of safety were marked as "Red" and "Blue", respectively, and farms not part of the SPF system were marked as "Conventional", a level therefore below the blue class.

Contacting the farmers individually, some additional information regarding the farm was retrieved, namely the Number of employees, if the Employment of foreign employees happened or not, if the farm was Exporting pigs or not (at the weight of 7 and/or 30 kg), the number of Years of experience of the farmer and the general age of the facilities, measured in two separate variables – the Age of the Youngest Building in the farm, and the Age of the Oldest Building, both in years.

The biosecurity data was also gathered through interview with the farmers, in which they were asked if they were willing to have the biosecurity of their farm tested, or not. For those who responded affirmatively, the biosecurity of their farm was assessed using Biocheck.ugent®, a scientific, risk-based, weighted scoring system, to evaluate a herd's biosecurity. The system was developed by the Veterinary Epidemiology Unit of Ghent University, in collaboration with the Flemish Animal Health Organization, and it is constantly updated. The scoring is attributed after answering to a questionnaire available for free at www.biocheck.ugent.be, containing 127 questions (for pigs, as for poultry they amount to 90) referring to all the pertinent components of biosecurity, subdivided in external and internal biosecurity.

The result comes under the form of a percentage score for the total biosecurity, which results from the mean between the external and the internal biosecurity scores, where a 0 reflects a farm with almost negligible levels of biosecurity, and a 100 indicates a farm with best-scoring biosecurity. The report returned after the completion of the questionnaire also presents values for 6 subtotals for the external and for the internal scores (discriminated in Table 4). In the same report, average values are also provided, so farmers checking the biosecurity of their herds can compare the results obtained.

As such, in our dataset, values were included for the Total Biosecurity score, the Internal Biosecurity score and the External Biosecurity score, as well as the 12 subtotals for each of the total biosecurity components (Internal biosecurity score subtotal A to F and External biosecurity score subtotal A to F). The areas each subtotal evaluates are described in Table 4. The Biocheck.ugent® website also provides deeper insight on each of these subtotal, as on biosecurity in general and the questions employed in the questionnaire (Ghent University, 2017).

| Total biosecurity component | Subtotal score | Meaning |
|---|---|---|
| **External biosecurity** | A | Purchase of animals and semen |
| | B | Transport of animals, removal of manure and dead animals |
| | C | Feed, water and equipament supply |
| | D | Personnel and visitors |
| | E | Vermin and bird control |
| | F | Environment and region |
| **Internal biosecurity** | A | Disease management |
| | B | Farrowing and suckling period |
| | C | Nursery unit |
| | D | Fattening unit |
| | E | Measures between compartments and the use of equipment |
| | F | Cleaning and disinfection |

Table 4 - External and internal biosecurity subtotals

Finally, the AM prescription data and the vaccination data were retrieved from the VetStat database, introduced previously in the dissertation’s introduction. The vaccination status was translated into 5 Boolean variables, referring to whether or not each farm had acquired vaccines against each agent, as recorded in VetStat:

- Vaccination against PCV2 (Porcine Circovirus type 2)
- Vaccination against PRRS (Porcine Reproductive and Respiratory Syndrome)
- Vaccination against *Actinobacillus Pleuropnemoniae*
- Vaccination against *Mycoplasma hyopneumoniae*
- Vaccination against *Lawsonia intrecellularis*

Attending to the fact that prescription entries in VetStat have an age group specification, and a disease group targeted by the prescription, it was possible to retrieve more specific information regarding the AM prescription data, which we used as a representative of AM consumption.

The age group division is made considering the expected weight at treatment, into three different categories: suckling pigs, boars, sows and bred gilts (as a group), with a standard weight of 200 kg, weaners, with a standard weight of 15 kg, and finishers, with a standard weight of 50 kg. This means any treatment targeted for one of these animal groups is registered as, and contributes to, that age group total AM prescribed amount (Dupont et al., 2017).

As for the targeted disease groups, there are six possible disease indication groups: reproduction and urogenital systems; udder; gastro-intestinal system; respiratory system; joints, limbs, hooves, skin and central nervous system; and metabolism, digestion and circulation (Dupont et al., 2017).

All this considered, 15 AM-related variables were extracted from VetStat, and added to the dataset. Three of them regarded the total amount of AMs prescribed for each age group: sows, weaners and finishers. Then, for each of those values, subtotals were obtained regarding the disease group targeted (gastro-intestinal system or respiratory system), but also translating if the treatment was a group treatment or an individual one, which was possible to assess considering the administration route reported and also the drug's specific administration instructions.

In detail, the antimicrobial prescription (AMP) variables include:

- AMP for sows and piglets (total)
    - AMP for sows and piglets for gastrointestinal indications
    - AMP for sows and piglets for respiratory indications
    - AMP for sows and piglets for individual treatment
    - AMP for sows and piglets for group treatment
- AMP for weaners (total)
    - AMP for weaners for gastrointestinal indications
    - AMP for weaners for respiratory indications

    - AMP for weaners for individual treatment
    - AMP for weaners for group treatment
- AMP for finishers (total)
    - AMP for finishers for gastrointestinal indications
    - AMP for finishers for respiratory indications
    - AMP for finishers for individual treatment
    - AMP for finishers for group treatment

Each of these parameters is measured in Animal Daily Doses (ADDs) per 100 animals per day. ADDs are a unit of measure of drug consumption, which comes from the DDD – Defined Daily Dose – with necessary adjustments for veterinary medicine.

DDDs, as they're used in human medicine, are described by the World Health Organization as the assumed average maintenance dose, per day, for a certain drug used for its main indications, in adults (WHO Collabotaring Centre for Drug Statistics Methodology, 2016). Generally speaking, it is the amount of a given medicine that is needed to administrate daily, in order to treat a certain affliction. The animal equivalent, DADD (Defined Animal Daily Dose), translates as a similar measure, only adapted according to an animal species, and an age group, when necessary (Statens Serum Institut et al., 2015). As such, DADDs are a metric which are not only specific to the agent, but also to other factors, such as animal species, age group and administration route.

Consulting VetStat, and assessing the drugs which were prescribed to a certain farm, as well as other data, such as the amount prescribed or targeted age group, and considering the specifications of the product, for example regarding administration routes, it is possible to calculate how many DADDs were prescribed to that same farm. That is, how many times the actual amount of doses prescribed exceeded the defined value for a daily administration in one animal. That value would be the total amount of drug doses prescribed to that farm, across all its animals in a give time period (the whole year of 2014, in this project's case) – the farm's number of prescribed ADDs. The AM prescription data obtained from VetStat was already under this format.

To allow an easier comparison and understanding of the AM prescription values, each total was divided by the number of animals in each farm (or in each age group, for the age-considering AM prescription values), multiplied by 100, and then divided by the number of days which the considered time period comprised (365). As such, the values used in this project reflected the amount of ADDs used in each farm, every day (in average), for each 100 animals (from now on abbreviated to ADDs/day). In a different interpretation, those values show us how many animals could be treated with the average amount of AMs used throughout a day, in each of the farms.

Other units of drug usage/prescription exist, such as the Treatment Incidence, which measures how many animals, out of a hypothetical group of 1000, receive daily AM treatment, for a given period of time (Postma et al., 2016). However, the Treatment Incidence is calculated over a standardized time period, normally corresponding to the period the animal is at risk – its lifespan – which need to be averaged or estimated. Using ADDs/day does not face this problem, as the temporal factor imbued in the metric was not estimated nor standardized, resulting instead in a gross division over the known period of time in which the AM doses were prescribed. The fact that the Treatment Incidence depends on this standardized time period also comes from the fact that it is an animal focused unit, which informs the proportion of animals treated, or according to different interpretations, the proportion of an animal's life spent receiving AM treatment. ADDs/day is a drug focused metric, translating the consumption of the whole farm in actual ADDs, instead of animals treated. This is a more appropriate approach for this project, as we are interested in the gross levels of consumption, and in the actual amount of AMs prescribed, rather than in the proportion of treated animals, since that allows us not only to compare what is the level of AM consumption daily between farms, and even between countries, but also to judge how high is that consumption when compared to the Danish normal thresholds, such as, for example, the limits the Yellow Card Initiative imposes on farms, which are also defined in ADDs/day.

It is noteworthy to say, the time of collection for all the data enunciated so far was the year of 2014, except for the biosecurity, which was from 2015. At the time the data was collected (2015), the most recent records available were from 2014. As such, since the biosecurity scores were assessed through interviews with the farmers, they regard one year forward in comparison with the rest of the variables. However, the farmers were asked if significant changes were made at the farm, which could justify a meaningful difference between the assessed biosecurity in 2015, and the score which would be achieved one year prior, to which they responded negatively.

## IV.II - Data Handling

As mentioned previously, this project aimed to look independently into the sow and weaner part of pig production, which lead to the splitting of the dataset. This was decided for several reasons.

First, ABN is a methodology which is very demanding in terms of computational resources, especially if the number of variables is high. Since our initial database was composed of over 50 different variables, this would pose a serious obstacle, since a single best-fitting model search would take more than a day to complete, which would make it impossible to perform the parametric bootstrapping later on, with the available computers.

Reducing the size of the database by a single variable has a dramatic effect of the running time of the analysis. For example, a single exact search of a database of 17 variables took 1 minute and 22 seconds to run on the computer on which most of the analyses were conducted. Enlarging the dataset to include just two more variables, and the total running time necessary to execute the search would be of 42 minutes and 18 seconds. Therefore, reducing the dataset to include only the variables necessary to specific parts of the production would allow overcoming that technical issue.

Apart from technical limitations, the two sectors (sows and weaners) are very different, also when it comes to AM policies. As such, it makes sense to analyze each of them individually, at least in a first investigation.

Finally, it is required for the Bayesian Network analysis that the database has complete cases, that is, that every value in every observation for every variable is not missing, which happened with many of the farms in the dataset. A particularly high number of missing values was found in production parameters for the weaning part of the production (mortality, weight gain and feed conversion ratio). Splitting the dataset would also ensure a higher number of observations for the analysis of the sow model, increasing its accuracy.

As such, two analyses were executed, one for the sows and another one for the weaners, and for the purposes of this the original dataset gave origin to two smaller subsets. To decrease the number of variables until it was computationally feasible to execute the analysis, only the variables pertinent to the analysis were kept, and, therefore, the two resulting datasets included the following variables:

Sow model:

- SPF Red (True/False variable)
- SPF Blue (True/False variable)
- SPF Conventional (True/False variable)
- Number of Sows
- AMP for sows and piglets (total)
- AMP for weaners (total)
- Mortality in sows
- Feed used for sows
- Farrowing percentage
- Percentage of sows unsuccessfully inseminated
- Litters/sow/year
- Piglets/litter

- Stillborn piglets per litter
- Mortality until weaning
- Weaned pigs per litter
- External biosecurity score
- Internal biosecurity score

Weaner model:

- SPF Conventional (True/False variable)
- Number of weaners
- Vaccination against PCV2
- Vaccination against PRRS
- Vaccination against *Actinobacillus Pleuropnemoniae*
- Vaccination against *Mycoplasma hyopneumoniae*
- Vaccination against *Lawsonia*
- AMP for sows and piglets (total)
- AMP for weaners (total)
- Weaned pigs per litter
- Mortality in weaners
- Weight gain for weaners
- Feed conversion ratio
- External biosecurity score
- Internal biosecurity score

Note that the SPF class variables were divided into binary variables. This was made to overcome a limitation in the R package used to execute the analysis: *abn*, by Fraser Lewis (available at https://CRAN.R-project.org/package=abn). This package allows only numerical or binary variables to be submitted for modeling. The ‘adherence to the SPF system’ variable was already binary, but not the ‘SPF class’, which was a categorical variable with three possible values: red, blue or conventional. As such, the second variable was converted into three binary variables, SPF Red, SPF Blue and SPF Conventional, with only two values, true of false, which code, respectively: SPF Red/ Not SPF Red, SPF Blue/Not SPF Blue and SPF Conventional/Not SPF Conventional. The last, being essentially equal to the Adherence to the SPF System variable, substituted the latter, which was removed from the database.

Even though the original dataset was split into a sow and a weaner model, the total AM prescription for sows was included in the weaner model and the AM prescription for weaners was included in the sow model. It was decided to include these variables in both models because, in the preliminary analysis which was conducted (and which will be described more thoroughly, further into this dissertation), a linear association was found between them. Including both variables in each of the models would allow to investigate if said associations were also significant in a multivariate level, which would reveal that AM prescription in different sectors of production in the same farm could be influencing each other.

One other variable was included in the weaner model which was related to the sow sector: the number of weaned pigs per litter. This variable was also included in the weaner model to assess if there was some degree of influence of the productivity in the previous stage of the cycle (the piglet production, here represented by the number of weaned pigs ready to enter the next stage of production) in the weaner network. An association here could confirm that better productivity in one stage would benefit the productivity in the next one, establishing a bridge of connection between the two models.

Unfortunately, due to the limited computational resources, the vaccination variables had to be left out of the sow model, as it would be impossible to model the network if they were included. However, since the weaner model had much less production variables to work with, it was possible to include them there.

It is noteworthy to say that three farms were removed from the dataset, as they presented big outliers to the analysis.

The first of the three was removed for having an abnormally big number of sows in the holding, more specifically 3000, which was the double of the farm with the second highest number of sows. For reference, the median of that variable is 600. This incredibly high number is, most likely, a typo.

Secondly, most likely an error in the input of the number as well, a farm with 10000 weaners was also removed. For this variable, the second highest number after this entry is 5500, and the median is 1750. Most likely, an extra zero was typed.

The third removed entry regards a suspicious entry reporting a farm with 414 sows reporting a 51.3% mortality in sows. Again, for reference, the median for this parameter is 9.5% and the second highest value after this farm is 19.9. The farm in question shows no other abnormal data. One could suspect of an error in the input of this value, once more, perhaps a misplaced comma, but ultimately this entry is unreliable and as such it was removed from the dataset.

After these removals, one would hope the dataset would be reduced to 157 entries. However, since complete cases are required for the Bayesian network modeling, that number was further reduced.

There were a large number of missing values all across the database. Only 4 farms in the whole dataset had complete cases, which are required to conduct the exact search for the most probable network structure. However, splitting the database into the sow and weaner model and eliminating variables which were judged as of minor importance (which were also the sources of most missing values) reduced the impact of this problem by increasing the number of usable farms.

In the end, removing incomplete cases from the two models with only the variables mentioned before left a total number of 133 valid entries for the sow model and 70 entries for the weaner model. The reason behind the considerably lower number of weaner farms with complete cases lies with the production parameters. For unknown reasons, many entries in the original database missed values in the weaner mortality, feed conversion ratio and weight gain for weaners, despite having values for all the other relevant variables, such as vaccination, AM prescription or production parameters related to the sow part of the production. As such, the number of eligible farms was lower.

Still regarding the data preparation and selection, it should be said why the finishing pig production was not investigated as well.

First, it should be taken into consideration that the number of complete cases for the finishers was even lower than in the weaners' case, amounting to only 47. This would significantly damage the accuracy of the model. But more than that, all the holdings selected for the study were sow farms, most of them focusing almost exclusively in that part of pig production and, if possessing finishers in the farm, it could have been as some kind of remainder of the main production. During a preliminary analysis of the data, it was even revealed some odd linear associations regarding finisher-related variables, that did not make much sense, such as between the total AM prescription for finishers and the percentage of sows with unsuccessful inseminations.

Considering this, it was decided not to model the finisher part of the production, as the available data did not depict that sector in a representative way, and any conclusion could be not applicable to the reality.

## IV.III - Additive Bayesian Network modeling

As mentioned previously to be the objective of this project, an ABN model was built with the data from the Danish sow herds.

Before, however, to allow a better understanding of the database before the ABN was modeled, a preliminary analysis was conducted. To this purpose, a script was designed in R to run across all the pairs of continuous variables, looking for significant associations through the means of simple linear regressions and returning only the ones with more than 95% confidence of association. Those which exceeded 95% confidence were then plotted, every dot colored according to the SPF class as to examine visually if farms grouped themselves by those variables.

The ABN analysis was executed in R (version 3.3.3), a free software environment for statistical computing. One vital R library used for the modeling using Bayesian Networks was ‘abn’, a package designed to model multivariate data with ABN in R, created by Fraser Ian Lewis and with contributions from Marta Pittavino, Reinhard Furrer and Gilles Kratzer, the latter the current maintainer of the package.

The distribution of every variable had to be necessarily defined: Gaussian for continuous variables, Binomial for binary variables and Poisson for unbounded counts. The variables in the dataset all fit amongst the first two categories. Note that no option for categorical variables exists, hence the need to split the SPF variable into several mutually exclusive binary ones, as previously mentioned.

No arc was forced into the network, as we desired to provide no prior to the structure. Some arcs were banned, though, as the SPF binary variables, being mutually exclusive, would naturally reveal themselves associated in the network. As such, all the possible arcs amongst them were forbidden.

An initial exact search was performed for each of the models with a maximum number of parents limit of 1. The search was then repeated by increasing the limit by the unit, until the optimal identified network stabilized and the network score improved no further. That parent limit was then set, and the model accepted as the optimal network.

The exact search for the sow model with a parent limit of 1 returned a DAG with a network score of -2848.263. The procedure was repeated for parent limits of 2 through 5, and the network score (and the DAG itself) stabilized at a parent limit of 2 (network score of -2577.731).

The same procedure was repeated for the weaner model, which stabilized at a parent limit of 3, and a network score of -1251.278.

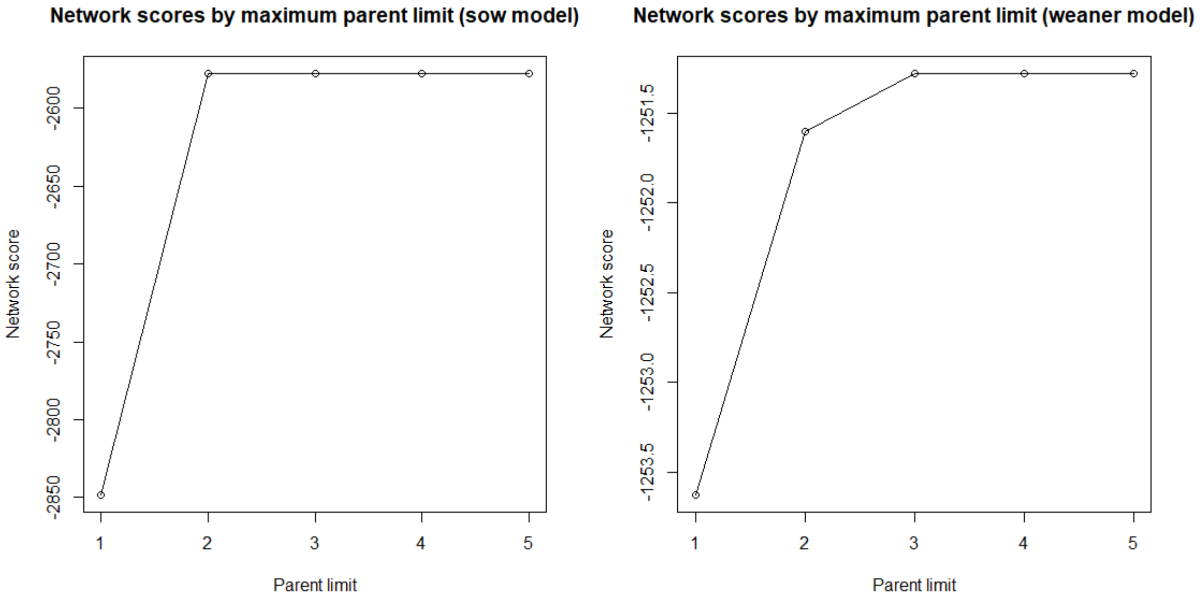


Figure 10 - Exact search network scores obtained for each defined maximum parent limit

To confirm the results obtained in the recovered Bayesian Networks for the sow model, which were surprising at first, two analyses were performed in parallel to the same data.

First, a Discriminant Analysis was conducted.

Since this method demanded a categorical variable as the discriminant variable, a new one was created from the herd size and the AM prescription for sows. As such, a four level variable came to be, coding a farm as having a big herd/small herd and a high AM prescription/low AM prescription for sows. The cutoff values used for this were the median of both original variables in the complete dataset, which were 600 sows and 2,175 ADDs/day.

The results of the linear discriminant analysis were presented in stacked histograms of the values of the discriminant function for each group, as to allow a visual interpretation of the degree of separation attained by the discriminant variable.

Furthermore, it was attempted to build a model which could classify new farms into the four groups intended (regarding the herd size and the AM consumption of the farm) based on the production parameters, namely the number of litters/sow/year and the mortality until weaning. For this, 80% of the dataset was used to train the model.

Several different variables, or combinations of those, were tested as discriminant variables (all being turned categorical using a similar method – splitting according to the median), namely the AM prescription for sows (alone), the external biosecurity, the number of litters/sow/year, the mortality until weaning, the SPF class, analysis which are more detailed in V – Results.

The dataset was also submitted to FFT, many decision trees having been generated and the best performing one being investigated, regarding whether or not it could successfully predict the level of productivity in a farm given the other variables.

That being done, the optimal ABN models obtained were subjected to the already mentioned bootstrapping process.

First, the marginal densities estimated were evaluated, as to whether or not the area under the curve summed to the unit. Then, the necessary data regarding the marginal densities was compiled into a single file, in a format which could be read by the MCMC software. The software used for these simulations was JAGS (version 4.2.0), a program originally developed by Martyn Plummer and downloadable through http://mcmc-jags.sourceforge.net. Additionally, a BUG file was written, containing the definition of the DAG identified previously. The bootstrapping was then performed in JAGS. For computational limitations, 5000 bootstrap analyses were preformed (the number often suggested in the literature was around 10000). The results were then analyzed in R.

## V - Results

Before diving directly into the ABN modeling, the dataset was first explored, and in a preliminary analysis, several linear associations between variables were retrieved and then plotted.

The complete dataset comprised information regarding 157 Danish sow herds (the three previously mentioned outliers removed): 35 from the North Denmark Region, 39 from the Central Region, 57 from the South Region and 26 from the Region of Zealand and Capital Region of Denmark.

The most common herd type was sow farms with finishers as a remainder of the main production, which composed 61% of the sample, as represented in Figure 11. The second most common type was the full line of production (farrow-to-finisher), covering 24% of the total farms, and only then farms with no finishers, making up the remaining 15% of the dataset.

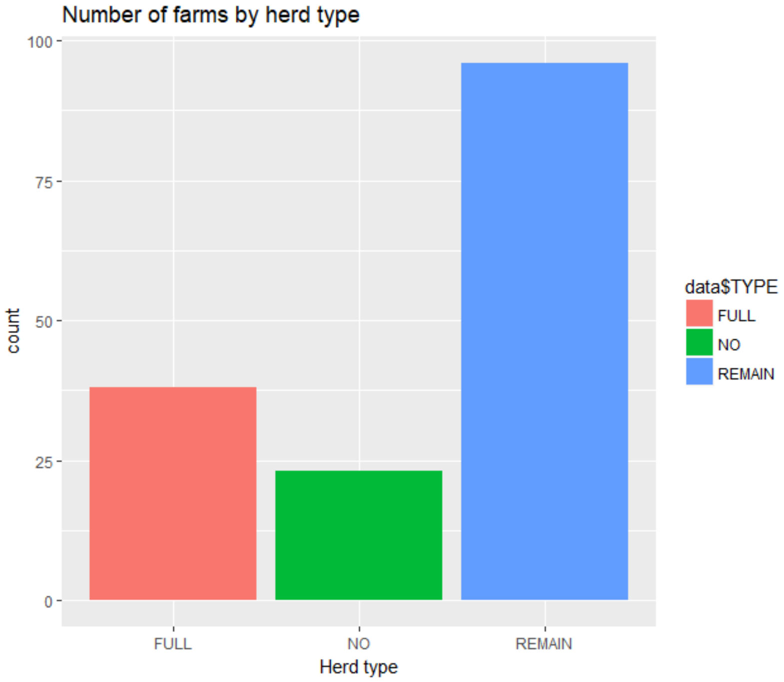


Figure 11- Number of farms by herd type. FULL indicated farms with the full line of production, NO codes farms with no finishers and REMAIN translates the number of farms with finishers as a remainder of the main production

Regarding the SPF class, the vast majority of the farms in the dataset belonged to the SPF system – 136 to be more precise. 21 farms were registered as not being part of the system, which were coded in the dataset as “conventional” farms. As for the farms that did belong to the system, only 10 ranked as “red” herds, the highest class, the remaining 126 being “blue” herds. These data are represented in Figure 12.

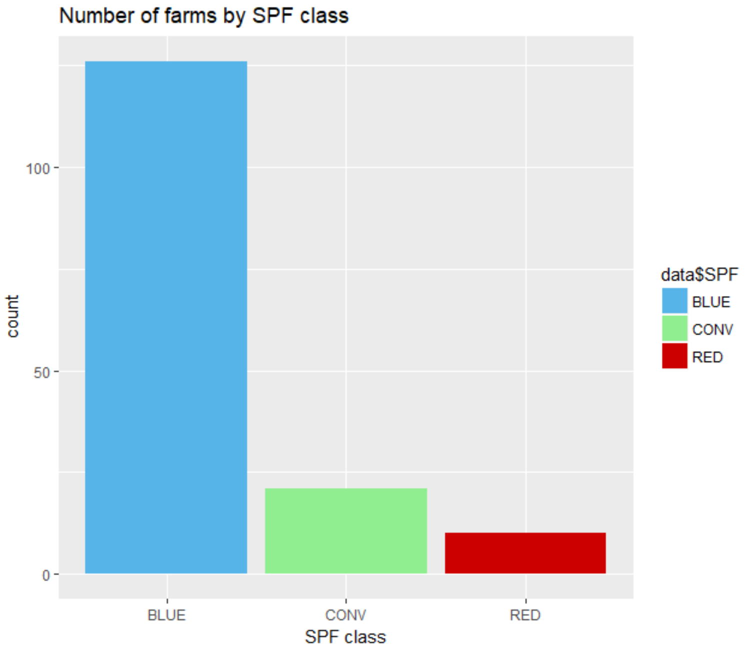


Figure 12 - Number of farms by SPF class. Red farms are superior to blue farms in the SPF system. Conventional farms do not belong to the system

The total number of animals present in each farm varied between 695 and 7480, having for 1st, 2nd and 3rd quartile, respectively, 1796, 2610 and 3670.

The number of sows was considerably smaller than the number of weaners, as represented in Figure 13, ranging between 150 and 1500 (as opposed to the weaners, which went from 40 to 5500). It should be reminded, however, that only holdings with a total number of sows higher than 100 were included in the study. The average number of sows in a farm was 615, and the average number of weaners was 1805.

Even though that sector of production was not modeled using Bayesian Networks, it can also be of interest to know the number of finishers ranged from 10 to 2688, excluding the farms which had no finishers at all. The mean for this variable was 459 pigs.

For more information regarding these variables, their maximum, minimum, mean and quartiles are displayed for consultation in appendix I.

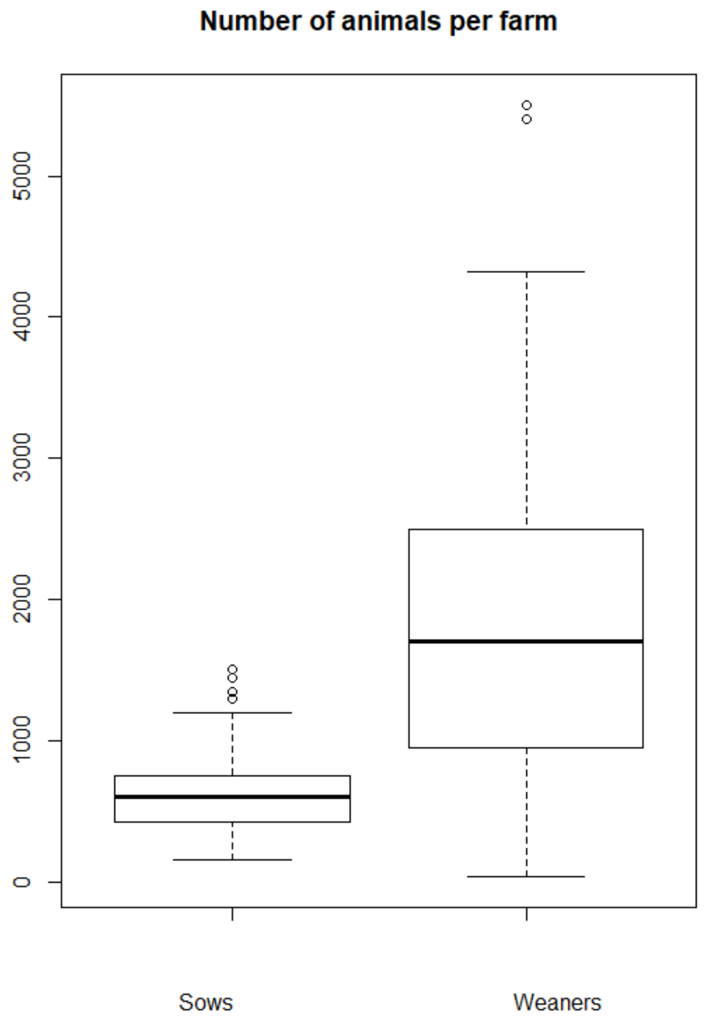


**Figure 13 - Comparison between the number of sows and weaners. Horizontal lines in each graph represent the maximum and minimum values and the quartiles.**

Regarding the AM prescription variables, the total prescription for sows and piglets ranged from 0.25 to 4.71 ADDs/day, as it can be seen in Figure 14's histogram, the mean being 2.27 ADDs/day. Looking into the subtotals of the total prescription, it was possible to notice higher amounts of AMs being prescribed for respiratory diseases than for gastrointestinal purposes. While both those subtotals presented a similar range of values, the mean and the median for AM intended for respiratory tract diseases (0.49 and 0.43) were fairly superior the equivalents in AM intended for gastrointestinal purposes (0.26 and 0.2 ADDs/day). A significant trend towards individual treatment was also noticeable, as the mean for individual treatment prescriptions – 2.01 ADDs/day – was much higher than the mean for AM prescriptions for group treatments, only 0.37.

It is noteworthy to say that the Yellow Card Initiative limit regarding AM prescription for sows and piglets by the end of 2014 was 4,3 ADDs/day. Only two farms of the total 157 (both SPF blue, in possession of finishers as remainder, one with 810 sows and the other with 1000) were exceeding the allowed limit.

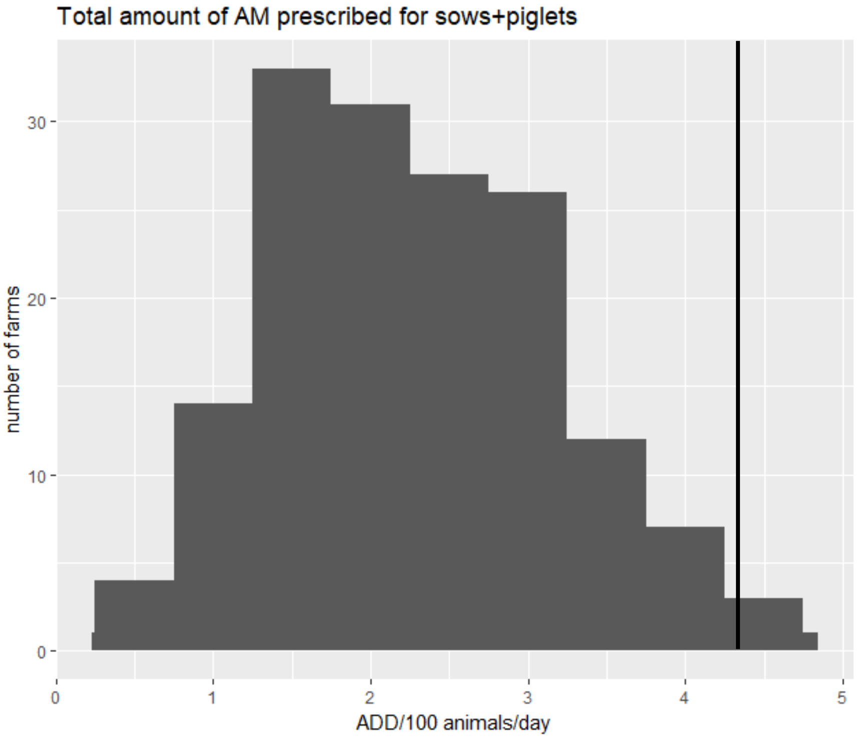


**Figure 14 - Number of farms by AM prescription level, for the sow sector. The vertical line represents the Yellow Card Initiative threshold for sows in 2014**

The AM consumption for weaners, as indicated by the prescription values, was completely different from the sows. The total consumption, which ranged from 0.56 to 22.81 ADDs/day, somewhat well distributed over all the values, was dramatically higher, as represented in Figure 15. Unlike with the sows, AM prescription for gastrointestinal purposes was higher, ranging from 0.04 to 22.74, as opposed to AM prescription for respiratory tract issues, which ranged from 0.03 to 13.05. The mean for both variables were, respectively, 7.63 and 3.35, a comparison which is evident in Figure 16.

The treatment option is also different as the mean for the group treatment is 9.25 ADDs/day, against 1.69 for individual treatment.

From the perspective of the Yellow Card, the result is also impressive. None of the farms exceeded the values established by the Danish Veterinary and Food Administration, which were of 22.9 ADDs/day by the end of 2014. Even judging by the current threshold of 21.8 ADDs/day, only 4 out of the 157 farms would indeed receive a yellow card.33

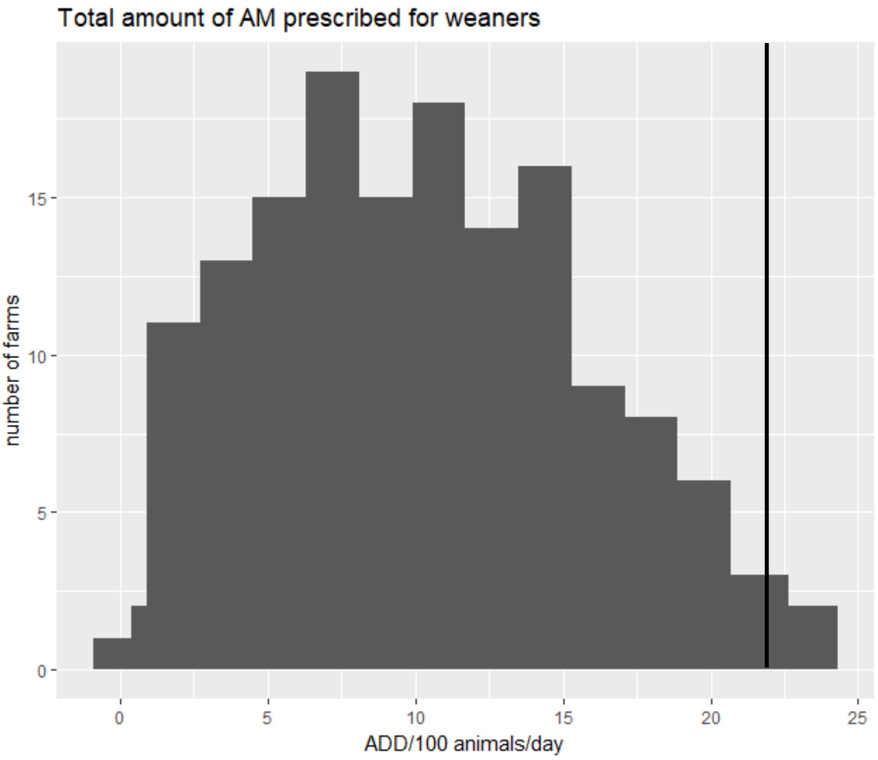


**Figure 15 - Number of farms by AM prescription level, for the weaner sector. The vertical line represents the Yellow Card Initiative threshold for weaners in 2017**

More information regarding the AM prescription variables, including the subtotals mentioned above, can be found in appendix I.

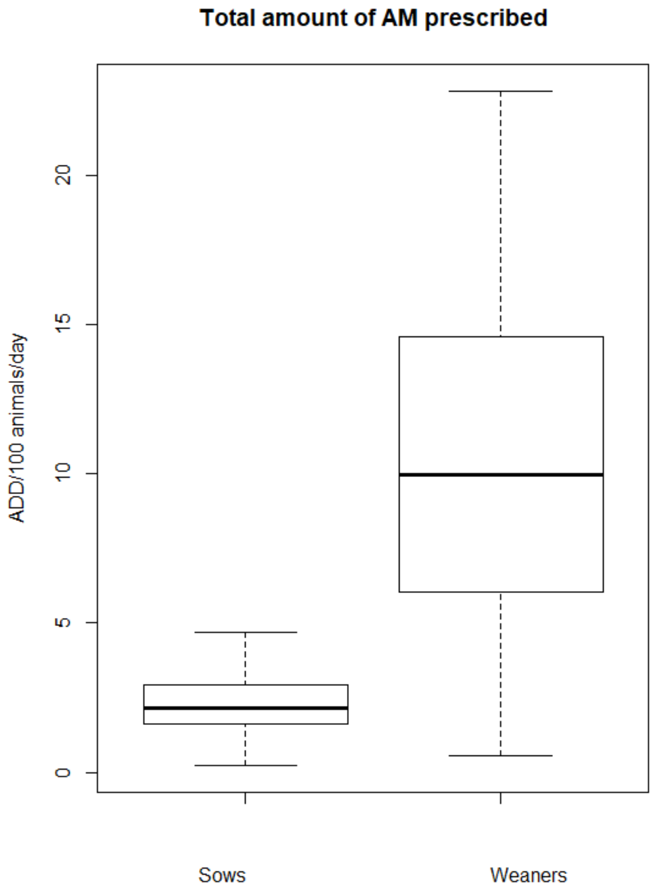


**Figure 16 - Comparison between AM prescription in sows and in weaners. Horizontal lines in each graph represent the maximum and minimum values and the quartiles. AM consumption in weaners is substantially higher.**

Regarding vaccination, a great number of farms employ vaccination against PCV2, 76% of them to be more precise. Although not as much, the percentage of farms which vaccinate the herds against *M. hyopneumoniae*, 58%, is also considerable. Vaccination against the three other agents/diseases, *A. pleuropneumoniae,* PRRS and *L. intercellularis* is not as prevalent, given the vaccinating farms percentage of 13%, 15% and 12%, respectively.

Biosecurity scores revealed themselves to be fairly high, as the total biosecurity score is, in average, 76.61, and the lowest scoring farm has 60 for the total score. The external biosecurity score, ranging from 67 to 96 with a mean of 85, appears to be fairly higher than the internal score, which ranges from 48 to 89, and has a mean of 67. 99. The comparison between these variables can be seen in Figure 16.

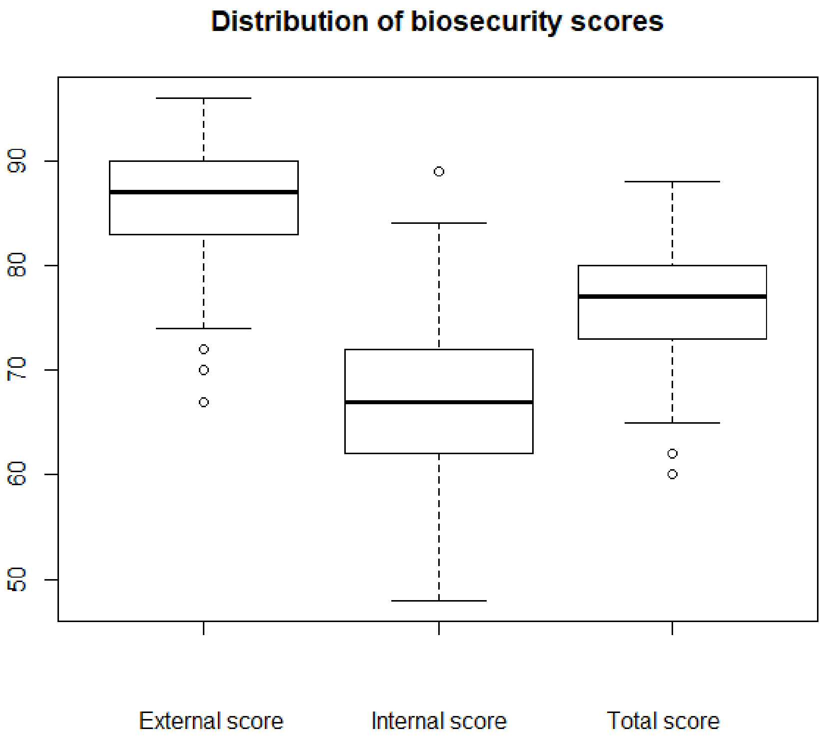


Figure 17 - Comparison between the distribution of biosecurity scores. Horizontal lines in each graph represent the maximum and minimum values and the quartiles.

For more details regarding the variables' summary, please consult appendix I.

Amongst the associations found in the linear preliminary analyses, there seems to be an effect of herd size on some variables, namely on the AM prescription for sows, which is positively related with the number of sows in a farm, and on the AM prescription for weaners, which is similarly associated with the number of weaners. Furthermore, there also seems to be a relationship between AM prescription for different age groups, as the linear regression between the AM prescription for sows and weaners was returned as significant.

One of the subtotals regarding the AM prescription for sows – the one quantifying individual treatment – appears to be positively associated with some of the biosecurity subtotals, namely the external biosecurity sub score B and the internal biosecurity sub score D. The meaning of every biosecurity subscore was previously explained in Table 4, in page 42.

The AM prescription for weaners seems to be associated with more variables, as it was returned as being positively associated with the overall external biosecurity score, and negatively with the mortality until weaning. The prescription for gastrointestinal purposes was also associated in the same way with the mortality until weaning and the external biosecurity score. AM prescription for individual treatment in weaners was negatively associated with the number of stillborn piglets per litter and the internal biosecurity sub score B, while the equivalent variable for group treatments was positively associated with the number of weaned pigs per litter and the overall external biosecurity, and negatively with the mortality until weaning.

All these associations were plotted in R, in which every observation in the graph was colored according to their SPF class, in order to identify any pattern or cluster of the same class farms across any variable, which none were found. These linear regression plots are presented in appendix II.

Once that preliminary analysis was concluded, the focus moved towards the ABN.

In the sow model, represented in Figure 18, there was evidence of a tree of associations between several production parameters. However, the only non-production variable directly associated with those parameters was the herd size, which was also directly associated with the external biosecurity, which in its turn had a direct connection with the SPF system (whether the farm was a part of it or not). No association was discovered with the AM prescription variables.

As for the weaner model, Figure 19, it was initially discovered an intricate network revolving around the SPF red variable. However, it was later discovered that, due to the need to eliminate farms with incomplete observations from the dataset, in order to perform the ABN, only two SPF red class farms remained. As such, it was decided the sample of SPF red farms was too small, and the only variable regarding the SPF system included in the analysis was whether the farm belonged or not to the system. It should be noted that this variable was equivalent to the “SPF conventional” variable, therefore, a farm with a false value for this parameter means it is not a “conventional” farm (thus, it belongs to the SPF system).

This latter variable revealed three direct associations, two with vaccination variables and a third with the external biosecurity score. There also seems to be a small line of associations between vaccination variables, namely against PRRS, PCV2 and *M. hyopneumoniae*. Once again, no direct or indirect associations were identified with the AM variables in this model.

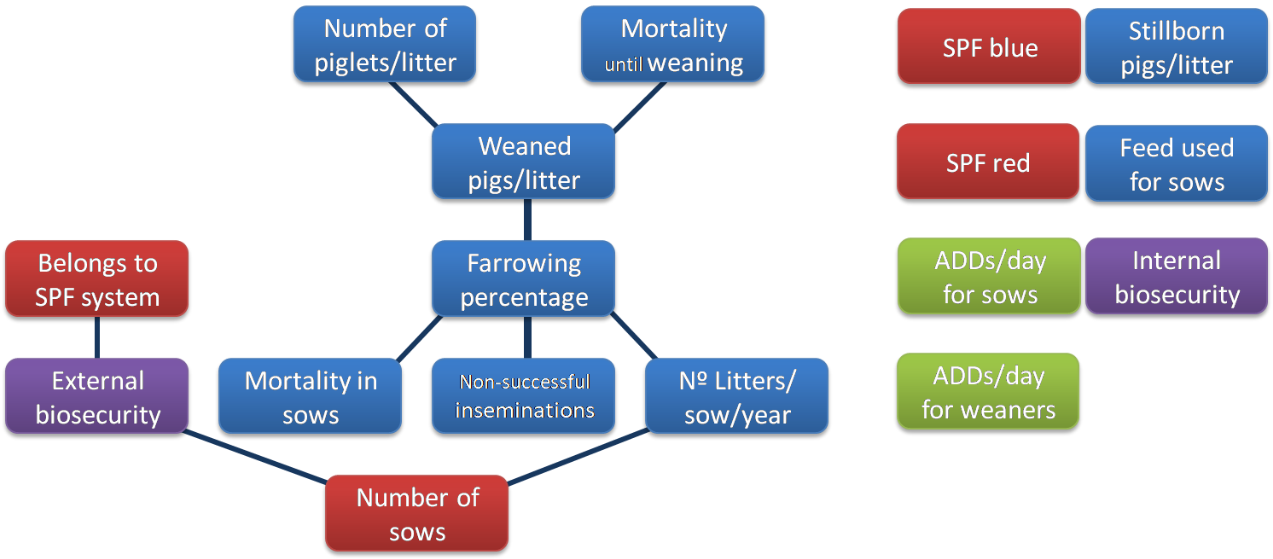


Figure 18 - Optimal Bayesian Network returned by the exact search for the sow model. Lines connecting two variables translate a direct association. The variables are color-coded: red for herd-related variables, purple for biosecurity scores, indigo for production parameters and green for AM prescription

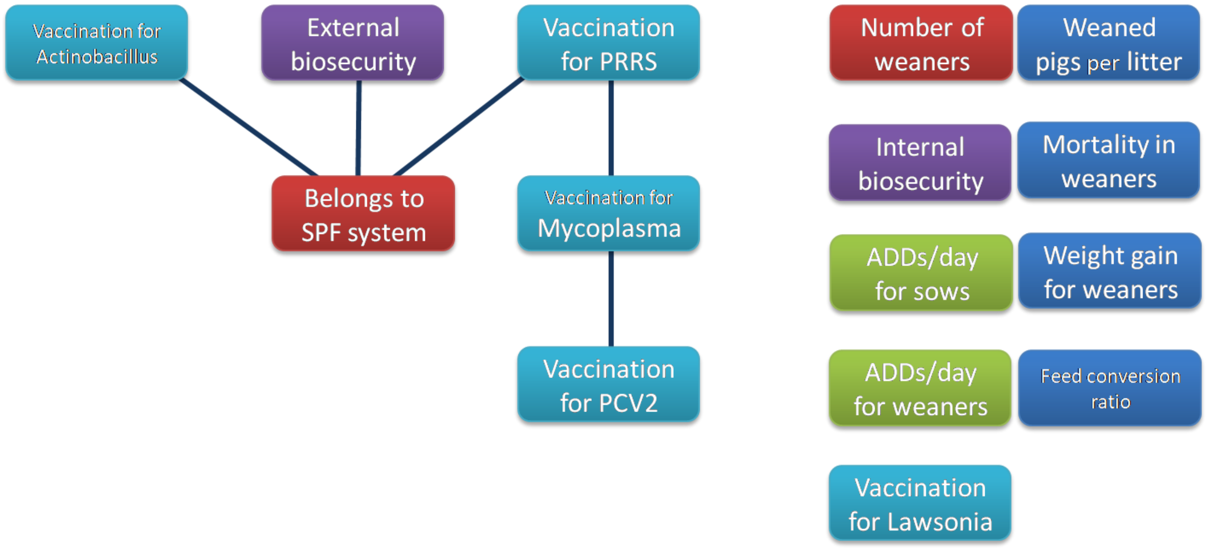


Figure 19 - Optimal Bayesian Network returned by the exact search for the weaner model. Lines connecting two variables translate a direct association. The variables are color-coded: red for herd-related variables, purple for biosecurity scores, light blue for vaccination, indigo for production parameters and green for AM prescription

At first, the lack of associations with the AM prescription variables was considered odd, and it was decided as prudent to perform a complementary analysis to confirm the results. Hence the Discriminant Analysis and the Fast and Frugal Trees mentioned in IV – Materials and Methods.

The first attempt at the discriminant analysis technique, as described previously, used a combination of herd size and AM prescription for sows as a discriminant variable. The herd size was included here because of the clear influence it had on the first Bayesian Network, and also due to a suspicion, which first arose when examining the early linear regression models, that herd size was acting as a confounding factor, translating a non-describable variable such as the farmer’s ambition, or entrepreneurism. As such, taking the number of litters/sow/year and the number of weaned pigs per litter as variables of interest, in order to gauge the level of production, it was attempted to discriminate farms through those parameters using the combined ADDs/day and herd size discriminating variable. This variable took 4 values: 1, 2, 3 and 4. Farms from groups 1 and 2 have an ADDs/day value below 2.175 for sows, as opposed to 3 and 4. Furthermore, groups 1 and 3 are farms with 600 sows or more, while 2 and 4 have less than 600.

The result was a complete failure in classifying the farms according to production levels with our discriminating variables. The projections of the farms over the discriminant functions were not grouped in any way, as evident in the scatterplot of Figure 20, which shows the values of the farms on two of said functions (as three discriminant functions were identified, three graphs like the one below were produced, and none shown any kind of grouping).

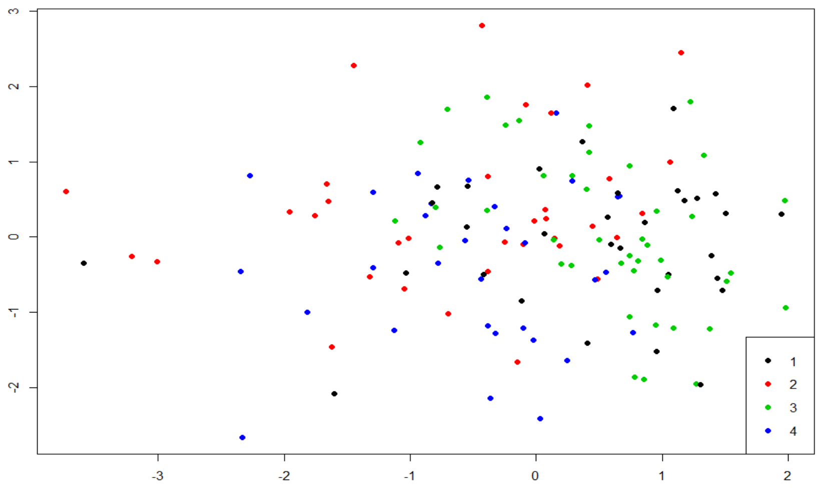


**Figure 20 - Scatterplot of the farms' values according to discriminant functions number 1 and 2. The axis indicate the values assumed by the observations on each function. The dots’ color indicates the farm’s class according to the discriminant variable. It was expected to see same-class observations group.**

The same functions were used to build a model, which should be able to classify new observations (farms) according to their AM prescription level and herd size, given values for the production, using 80% of the observations in the dataset to train the model. When it was tested, it was verified that it had a misclassification rate of 63.1%, which is rather unacceptable. The following partition plot (Figure 21) illustrates the performance of the discriminant functions, with the misclassifications marked red.

Examining the histograms exhibiting how the discriminant functions were splitting the data into the four groups (Figure 22), it becomes clear why: the degree of discrimination is mediocre. Figure 23 provides an example of a successful Discriminant Analysis.

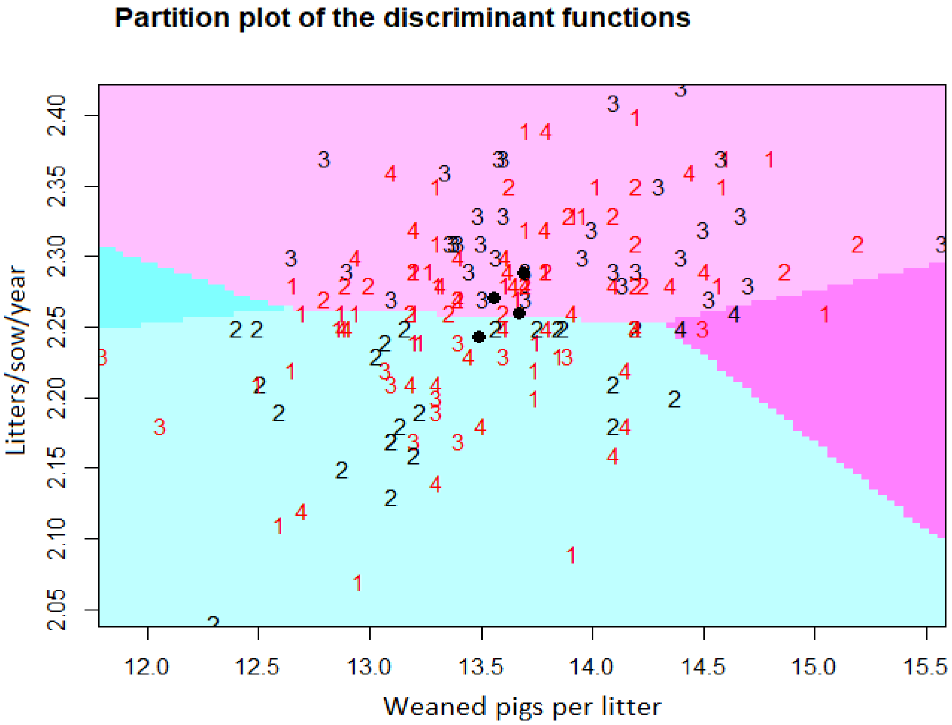


**Figure 21 - Partition plot presenting the result of the discriminant analysis. Each algorism represents an observation, the number coding the group the observation truly belongs to. Black numbers represent observations correctly classified, red numbers are misclassifications.**

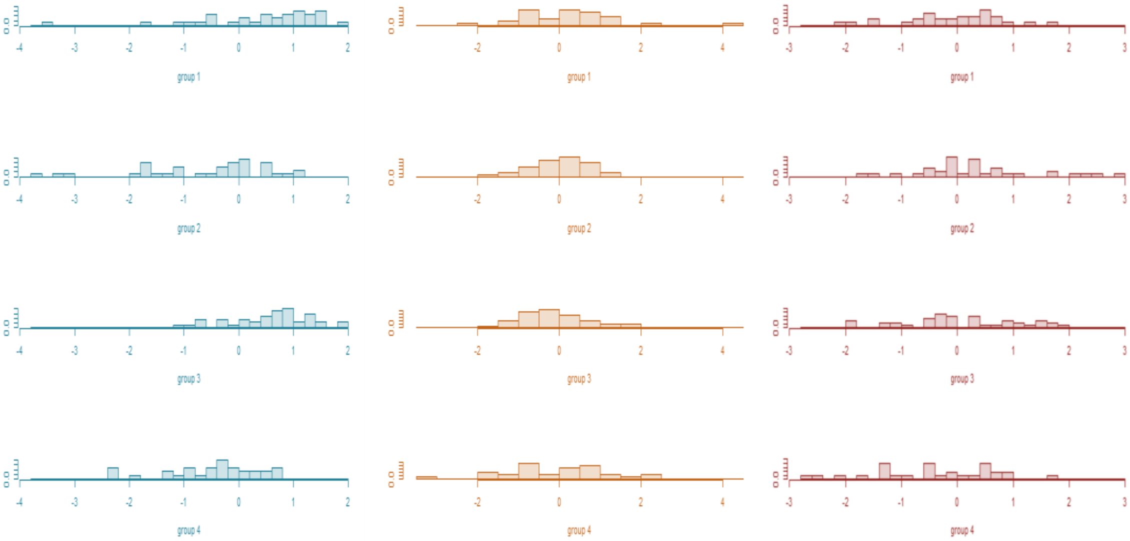


**Figure 22 - Histograms presenting the separation of the observations according to the 4 groups attained by the first, second and third discriminant functions (represented in blue, orange and red, respectively)**

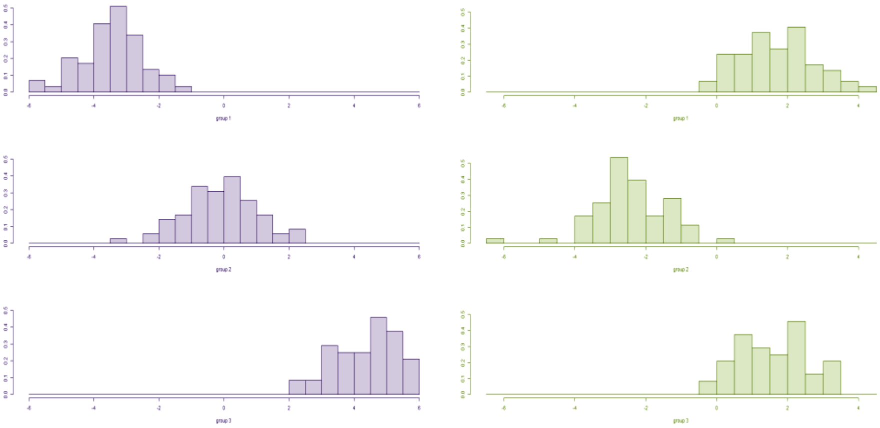


**Figure 23 - Example of histograms depicting data separation by successful discriminant functions. Notice how the observations are well separated in different zones of the scale**

Several different parameters were tried as discriminant variables, all failing to successfully separate the dataset by groups.

The ADDs/day for sows, alone, had a slightly better performance as discriminating variable than when combined with the herd size, but trying to predict the level of AM prescription given the production values still resulted in 52% misclassification. The opposite was also attempted: using production as the discriminating variable (suffering a process similar to the one used for AM prescription in order to transform it into a categorical variable) to discriminate the AM prescription and the external biosecurity, but that as well resulted in a bad misclassification ratio (40%).

In parallel, another complementary analysis was made using Fast and Frugal Trees (FFT).
For this, a binary classification key was devised to categorize the observations according to the decision tree. Four production parameters were used to create this key: the weaned pigs per litter, the number of litters per sow per year, the farrowing percentage and the mortality until weaning, where a cutoff point was used for all the factors equivalent to the median of each one (14, 2.3, 90% and 12, respectively). As such, farms were classified as being the best producing ones should they have values superior to the cutoffs for all the three first production parameters and lower than 12 for the mortality until weaning.
Using this classification key, 31 farms (20% of the sample) were classified as the best producing ones. As such, the remaining 126 farms were labeled as the bottom 80%. Cutoffs equivalent to the median were used because, when using values indicating a better production than those, only 7% of the farms were classified as the best producing ones, which would make the decision tree too demanding.

That being done, 70% of the dataset was used to train FFT models. These then attempted to classify the database, the tree with best classification ratio being retrieved as the best model.
Should a decision tree be able to classify a farm between being in the Top 20% (therefore, having a high production) or in the Bottom 80% based on variables such as the biosecurity or the AM prescription, a relationship between those variables would be proven.
However, the performance of the best performing tree, represented in Figure 24, was not satisfying enough.

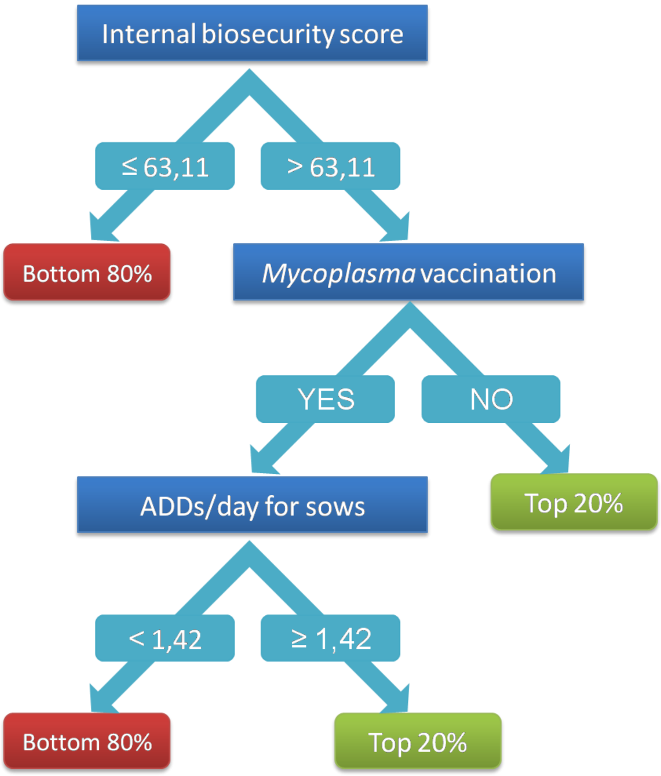


**Figure 24 - Best performing FFT, with the internal biosecurity scores, the vaccination status for *Mycoplasma* and the ADDs/day for sows as cues for classifying farms as Top 20% or Bottom 80%**

The internal biosecurity, the vaccination for *M. hyopneumoniae* and the ADDs/dat for sows were the parameters used in the cues to categorize observations as the Top 20% or the Bottom 80%, as it is displayed in the scheme above. However, only 87 of the farms were classified correctly with this decision tree. Only 62 out of the 126 farms which should have been labeled Bottom 80% were classified as such, while from the 31 Top 20% farms, 6 of them were wrongly considered as Bottom 80% by the model.

The internal biosecurity revealed to be a fairly accurate cue, as the exit leaf it lead to in case of farms with an internal biosecurity score below 63.11 marked 52 farms as Bottom 80% correctly, though it was also here that 5 of the 6 Top 20% farms were not classified as such. The vaccination against *Mycoplasma*, however, performed much worse, classifying 28 farms wrongly and only 15 correctly in the leaf exit it lead to, for non-vaccinating farms. The AM prescription for sows, in the end, was responsible for the remaining 37 misclassifications, 36 of which were farms which should be in the Bottom 80%, which were mistakenly marked as Top 20% for having a high AM prescription.

In the end, the best performing tree had a sensitivity of 81% (25 out of 31 farms) and a 49% specificity (62 out of 126 farms). Overall, there was a 45% of misclassification, which is not satisfactory enough.

As the results proven that the lack of associations between AM prescription and other variables was not due to errors in the analysis, the ABN proceeded towards the parametric bootstrapping.

The densities of the marginal posterior probabilities were estimated and visualized as to identify possible errors. Nothing unusual was discovered. Also, all the area under the curve for the densities estimated summed up to one (with an error of 0.000001 for the sows model and 0.00001 for the weaner model), indication the estimation was reliable.

The 5000 repetitions for the sow model bootstrapping took about 6 weeks. The number of arcs in the DAGs obtained from the simulations ranged from 5 to 13, but the arcs absent from the exact search model were removed (because they were not found in the initial model, we know they are due to overfitting). After that, the DAG containing the least arcs had 4, and the one with the most had, logically, 9.

A table exhibiting now many arcs were retrieved for each pair of variables can be found in appendix III. Mind the columns are the parents and the rows are the children.

As, for this analysis, the arc direction was being ignored, all arcs between the same nodes but with opposite directions were summed, and were retained should they be present in more than 50% of the bootstrapped DAGs. This meant all the arcs in the optimal DAG found through the exact search were retained.

The percentage of support for every arc, given by the number of times the arc was retrieved in the bootstrapping divided by the number of repetitions, can be found in table 5.

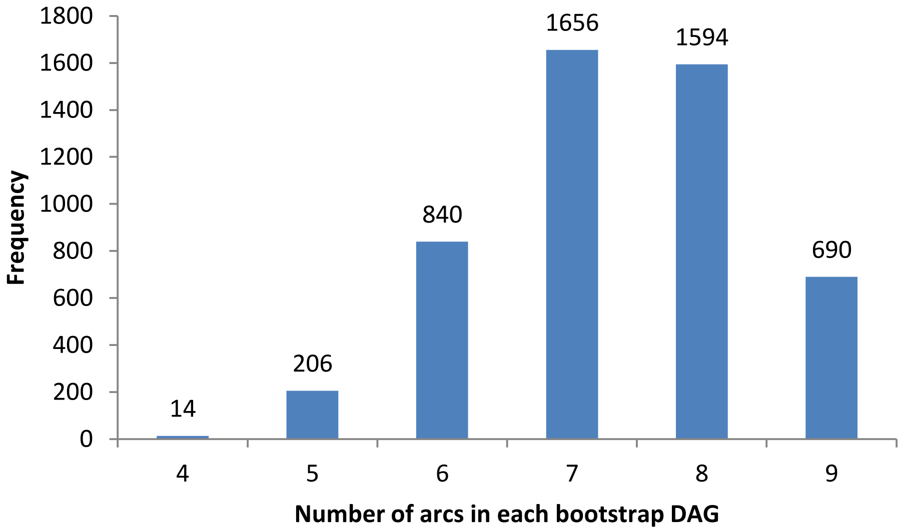


Figure 25 - Number of arcs recovered in the boostrapping of the sow model, after the removal of the arcs not included in the original DAG

| Association | Number of DAGs where the arc was present | Percentage of support |
|---|---|---|
| Belonging to the SPF system and external biosecurity | 2840 | 57% |
| Farrowing percentage and mortality in sows | 3453 | 69% |
| Farrowing percentage and percentage of sows with unsuccessful inseminations | 5000 | 100% |
| Farrowing percentage and number of litters per sow per year | 4988 | 99% |
| Litters per sow per year and number of sows | 3223 | 64% |
| Number of piglets per litter and number of weaned pigs per litter | 5000 | 100% |
| Mortality until weaning and number of weaned pigs per litter | 5000 | 100% |
| Number of weaned pigs per litter and farrowing percentage | 4391 | 88% |
| External biosecurity and number of sows | 2785 | 56% |

Table 5 - Support percentage for every arc in the DAG for the sow model

Regarding the weaner model, the 5000 analyses of the bootstrapping only took little over a week to run.

Just as with the sow model, the number of arcs retrieved in the bootstrapped DAGs varied quite greatly, this time between 1 and 15 arcs. Once again, this range was reduced once the arcs due to overfitting were removed, leaving a range of 0 to 5 arcs.

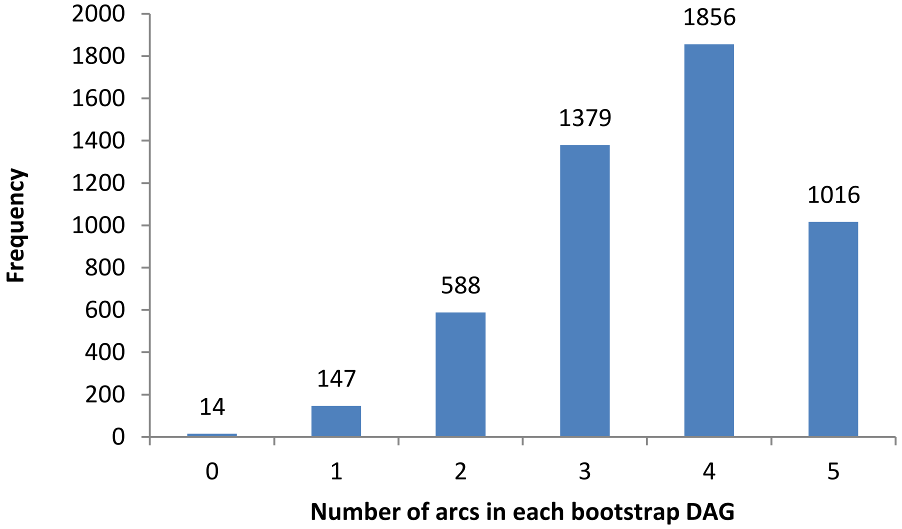


Figure 26 - Number of arcs recovered in the boostrapping of the weaner model, after the removal of the arcs not included in the original DAG

Summing, once more, the arcs between the same nodes but with opposite directions, it was possible to examine which ones were present in more than 50% of the simulated DAGs. The total amount of arcs retrieved for each association can also be found on the table in appendix IV. As for the percentage of support for each arc, that information can be found below. None of the arcs featured in the weaner model needs to be removed as well.

| Association | Number of DAGs where the arc was present | Percentage of support |
|---|---|---|
| Belonging to the SPF system and vaccination against *A. pleuropneumoniae* | 3937 | 79% |
| Belonging to the SPF system and vaccination against PRRS | 2978 | 60% |
| Belonging to the SPF system and external biosecurity | 4125 | 83% |
| Vaccination against PCV2 and against *M. hyopneumoniae* | 3179 | 64% |
| Vaccination against PRRS and against M. *hyopneumoniae* | 3745 | 75% |

Table 6 - Support percentage for every arc in the DAG for the weaner model

Given these results, we have confidence in the models previously presented, since no change to them was necessary, given the bootstrapping results. As such, the final models for the sow and the weaner part of pig production follow in Figures 27 and 28.

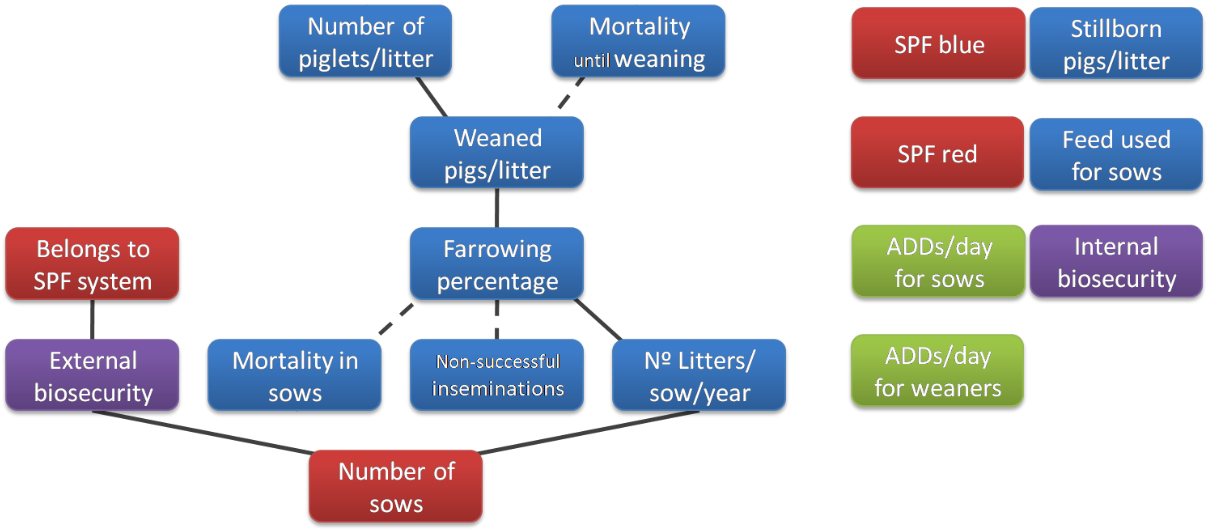


Figure 27 - Final DAG for the sow model. Full lines between variables mean a positive association. Dashed lines between two variables mean a negative association

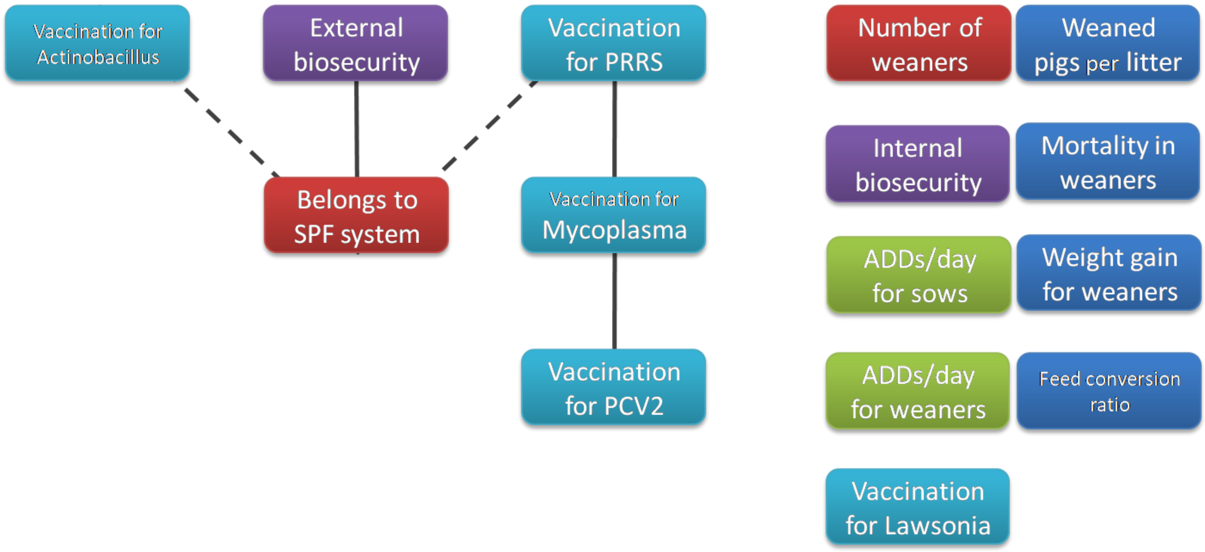


Figure 28 - Final DAG for the weaner model. Full lines between variables mean a positive association. Dashed lines between two variables mean a negative association

## VI – Discussion

At the beginning of this project, it was expected to find some degree of association between the AM prescription variables and productivity, biosecurity and vaccination. It was hypothesized that AM prescription would be related to the production variables, since there is a preconceived idea that AMs improve production, and because bivariate analyses showed linear associations between subtotals of the AM prescription for weaners and production parameters such as the mortality until weaning and the number of stillborn piglets per litter. As such, the output from the first ABN exact search was received with some surprise, as the AM-related variables were not involved with any of the remaining variables at all.
The discriminant analysis and the FFT came as a means to confirm that the initial ABN results were not as error, but rather there was no association to be found.

Analyzing closely the produced model for the sows, we can see a very strong tree of associations between several production parameters, one which not only receives a great support from the bootstrapping simulation but which also actually makes sense. Beyond the more obvious and intuitive conclusions, such as the connection between the number of piglets per litter and the number of weaned piglets per litter, we can identify an underlying pattern: one which reflects the difference in quality between farms.
Farms which present one of the production parameters at a high value are likely to have achieved said values through good farming practices, efficient production, in summary, by being high quality farms. As such, it is only natural that other production variables also rise

up to good numbers, since one flaw in the system could be enough to disrupt all the production and make its levels fall overall (a clock only works when all the gear are good and functioning), and it is only logical that farms willing to improve certain aspects of the production to achieve those high levels do not focus on singular measures, but rather improve the whole system altogether.

The positive association between the fact a farm belongs to the SPF system and the total external biosecurity score also makes sense, considering the demands and requirements such farms have to comply with to be within the system.

The association between the number of sows in the farm and both the number of litters per sow per year and the external biosecurity level is slightly more curious. Associations with the number of animals have been found even in the preliminary analysis, where the number of sows was actually positively associated in a linear regression with the AM prescription for that age group. It makes sense that these associations are not based on a biological reason. It could be, instead, that the herd size is masquerading a third variable in each association.

One possible explanation for this regards what can be considered the ambition, or entrepreneurism, of the farmers. High production levels are rewards for good and effective animal farming practices, and the revenue they produce allow, in their turn, a bigger investment on the activity. It is possible that farms with a high number of animals (in this case, sows) usually have good production levels because the farmers are more ambitious, that is, have set higher goals in terms of said production levels, and farm expansion, in terms of animal density, can be both a consequence of efficient, successful farming practices, but also a stepping stone to achieve even better productivity.

In this case, the association between high numbers of sows and high external biosecurity scores could be seen as an encouraging indicator, foretelling that top-level, high producing farms are embracing good measures of biosecurity, which from an epidemiological point of view, is at the very least good in terms of disease prevention and sustainability.

It could also be that external biosecurity is associated with herd size as a necessary measure to maintain certain standards by the owners of big farms, who need to meet certain levels of production, efficiency and other standards, such as the ones required by the SPF system, which as we can see in the Bayesian Network, is also connected with good productivity, albeit indirectly.

In fact, both external biosecurity and a status within the SPF system are associated indirectly with productivity. If we consider the SPF variable as a compliance assurance to all the specifications of the SPF system health rules (concerning protection against infection, health inspection or controlled purchase and transport of animals, for example), it becomes hard to know whether the high biosecurity is a variable which “allows” a farm to be in the SPF system, or if it is the adherence to the system what drives the farmers to elevate their biosecurity. In any case, both those variables, and all the measures they imply, are indeed indirectly associated with production. This is consistent with the results found by Postma et al. (2016), who analyzed data from farrow-to-finished herds in four European Union countries (Belgium, France, Germany and Sweden), finding a negative association between external biosecurity and treatment incidence: farms with higher biosecurity tended to have a lower percentage of the animals receiving AM treatment.

When looking into the weaner model, the produced network is quite different. However, one common feature remains: the association between the fact a farm is part of the SPF system, and its total external biosecurity score.

Once again, we can raise the question if SPF farms are associated with higher external biosecurity because the latter is a requirement for the SPF health regulations, or if it is the farms already with high biosecurity levels which tend to gather in systems such as these, in order to protect their standards, assure the costumers their product is superior, is terms of food safety, and safeguard themselves, as they become less likely to “import” diseases into the holdings.

The optimal model also identifies several associations between vaccination variables, which in the sow model had to be left out since the number of variables was already too high for the available resources. Association between vaccination against PCV2 and *M. hyopneumoniae* can be expected, since the two variables were the highest amongst the vaccination parameters. It seems intuitive that farms which vaccinate for one agent might also be interested in vaccinating against the other, as both variables had a significant number of farms administrating the vaccine. There was also a positive association between vaccination for PRRS and for *M. hyopneumoniae*. A posterior analysis revealed that every farm vaccinating against PRRS was also vaccinating against *M. hyopneumoniae*, hence the association.

The negative association between the SPF status and both the vaccination against PRRS and *A. pleuropneumoniae* is more curious. In fact, only 12% of the farms that are part of the SPF system use vaccines for PRRS, while the percentage of vaccinating farms against PRRS is 33% amongst the so-called conventional farms. The difference is bigger when regarding vaccination against *A. pleuropneumoniae*. A mere 7% out of the SPF farms employ vaccination against this agent, while this number rises to 48% in conventional farms. The fact there is a small number of non-SPF farms in the sample may be influencing this result, but the numbers do show a marked difference.

Perhaps, this is a reflection of the sanitary situation in the farms, as the holdings which vaccinate can be doing so in order to prepare against a recurring threat, one that is known to reemerge once the prophylactic treatment is interrupted, while non-vaccinating farms might be keeping the infection under control through measures that prevent the entering of the agent in the farm, rather than its dissemination once it is present. This is consistent with the conclusions from Temtem, Kruse, Nielsen, Pedersen, and Alban (2016), whose results suggest that sow herds with heavier vaccination policies for weaners have higher AM prescription values, which could be due to underlying health problems in the herd, and Postma et al (2015), who have found a positive association between the number of pathogens farms were vaccinating against and the treatment incidence with AMs. Most likely, there is also influence from the health regulations imposed by the SPF system, as the farms which belong to it are expected to be exempt from enzootic pneumonia by *Mycoplasma hyopneumoniae*, pleuropneumonia by *Actinobacillus pleuropneumoniae* and PRRS (amongst others, as previously discussed). While this would not explain the negative association the SPF status has with the vaccination, since vaccination protocols could be employed as a method to ensure, precisely, absence of said diseases, we can take advantage of the benefits ABN offers, by providing a graphical network, to make sense of this. Because the adherence to the SPF system is also associated with high external biosecurity, one good explanation for this part of the network is that betting on measures which prevent the entry of diseases in the farm reduces the need for vaccination.

AM prescription, which in the end was an indicator of AM consumption, was indeed not associated with any other variables. What was initially seen as suspicious was later confirmed by two other analysis. Thinking the problem through, it might be a fallacy to assume AM consumption equals better production, and farms with higher values of AM prescription should generally have higher levels of production because of that. The environment must be considered when making such assumptions.

Several explanations can be formulated, to justify this absence of AM influence on the network of variables interdependencies.

One of them alludes to the fact that this is a sow model. Being focused on the sow part of the production, most production parameters quantify the number of piglets produced, and the health of the sow. Since this sector has a significantly lower AM consumption than the others, considering not only the information provided by dataset but also expert opinions regarding the farming paradigm in Denmark, it would be unlikely that small variations in the already low AM consumption would benefit the production that much.

It is also possible that regardless of the amount of AMs administrated, the main problems in this phase of the production are not treatable with these drugs. Piglet problems such as starvation, energy reserves depletion or piglet being crushed by the sow, all issues frequent in this stage of the production, can only be prevented and minimized by good management and adequate animal farming practices. As such, it could be that it was simply this part of the production which was not that much dependent on AMs as the other – and the hypothesis mentioned almost certainly play a part in this dynamic.

However, a similar thing happens in the weaner model.

Just like with the sow data, the weaner model shows no association between AM variables and the remaining ones. Being an area of production in which farms have, in average, an AM consumption about 5 times higher when compared to the sow sector (in ADDs/day), it should be expected a greater effect of this variable. Furthermore, while in the sow sector it is more common the individual treatment when administrating AMs, in weaners the amount of ADDs/day a farm, in average, employs for group treatment is about 5.47 greater than the equivalent for individual treatment, the average for both this values being 9.25 ADDs/day and 1.69 ADDs/day, respectively. It is clear that not only there is a higher level of AMs being used, but the number of animals exposed to the treatment is also bigger.

As such, it is puzzling the lack of impact these numbers have on the rest of the dataset's variables. That is, until we consider the dataset itself.

The analyzed dataset contains data from 157 Danish sow herds. Considering all the restrictions and policies active in Denmark at the moment, several of them being discussed in the first chapters, the AM consumption is already substantially low. In 2014, as it was shown in table 2, the amount of AM sold per population unit, to use in food-producing animals in Denmark (44.2 mg/PCU), was almost ten times lower than the values for Spain (418.8 mg/PCU). The conditions which enable this level of AM consumption are out of the scope of this work, but it is undeniable that AM consumption in Danish farms is much lower than the average.

Either by having found ways to enhance productivity despite the lowering limits for AM usage, by compensating the protection provided by AMs with other measures, or probably both, the farms featured in the dataset are not representative of an average European farm. They have a very low average AM consumption to begin with. They also have substantially good biosecurity scores, as 95% of the sample has an internal biosecurity score between 50 and 83. In fact, over three quarters of the farms form the data set have an internal biosecurity higher than 60, according to the BioCheck.UGent® scale. The numbers are even higher when considering external biosecurity, which, as seen on the model produced with ABN, is positively associated with the production parameters: the scores for this variable range from 72 to 94, considering a 95% confidence, and the distribution of scores is much more skewed towards higher values, as shown in figure 17. 75% of the farms in the sample have a biosecurity score equal or higher than 83.

That being said, there is a need to examine the results of this analysis in the context which made it possible. A setting where high levels of biosecurity and low levels of AM consumption are the norm, and where several other measures are in place to both ensure these characteristics, and enhance them even further. For example, according to the study by Postma et al. (2015), the analyzed farms from the four countries – Belgium, France, Germany and Sweden – had, respectively, mean values of 57.8, 58.6, 63 and 63.7 for the external biosecurity, and of 51.1, 57.3, 55.4 and 58.8 for the internal biosecurity, using BioCheck.UGent® scores. The farms from the dataset used for this study have a mean external biosecurity of 85.88 and a mean internal biosecurity of 66.89.

As such, we can say there is a relationship between this environment and the associations revealed in the ABN. Although the associations regarding the AM prescription, which were expected and which motivated this work, were absent, and even though it was a goal of this project to examine how AM consumption was associated with other variables in the Danish pig production context (which was made impossible by the lack of those associations), it is not to say no conclusions can be drawn from these analyses. It can be hypothesized that this lack of associations regarding the AM related variables tell something on its own. Indeed, considering the already low amounts of AMs used in the Danish herds, it would be unlikely that the AM usage would be what's sustaining the production system, and that fluctuations in the AM consumption would result in big changes in the overall productivity. The consumption is already so low, that further reductions in its usage would not affect the other variables effectively. And in this scenario, biosecurity appears as a much stronger influence in the productivity (in the particular case of the sow sector) than the AM prescription, proving it is possible to sustain, or at the very least benefit, the production with alternatives measures to an overusage of AMs.

It should be noted that, since this sample is not representative of the average European pig farm, this analysis' results cannot be applied directly to every pig farm. The maintenance of production without the reliance on AMs works in the Danish environment because of the specific characteristics of this setting. It would be wrong to claim AMs have no effect on the average level of production, as different situations, such as in farms with less measures to assure a decent biosecurity, or different vaccination policies, or even factors which were not analyzed in this study, such as climate, geography or epidemiological context of the region, would probably manifest a different DAG when subjected to a Bayesian Network analysis.

To analyze more deeply which factors are needed to be in place to allow lower AM consumption without sacrificing the production, further research would be necessary. Repeating the ABN analysis in different contexts, such as in farms without the characteristics of the average Danish pig farm (high biosecurity, AM usage limits, sanctions for herds overusing these drugs, systems like the SPF, to name a few) would contribute to better understand the behavior of variables such as the AM prescription or the biosecurity, as well as to emphasize the differences between systems which allow such a low AM usage in Danish pig production.

However, in the characteristics of the environment where the dataset was built, the ABN model shows us that external biosecurity and the requirements for the SPF system have a positive effect on productivity, considering the sow sector, while in the weaner sector these variables are related to the kind of vaccination policies in place, being that farms with better biosecurity and which are part of the SPF system tend to rely less on vaccines. Moreover, AM consumption (here indicated by the prescription) shows no association with productivity, or any other variables for that matter, despite maintaining high levels, which is evident given the already mentioned levels of production.

## VII – Conclusion

This study used Additive Bayesian Networks to unveil the network of interdependencies between production parameters, AM prescription, biosecurity, vaccination and other variables, in Danish sow herds. Two models were produced to deal with computational limitations, incomplete registries for some observations and due to the big differences in the two sectors, regarding AM usage for example.

The sow model revealed an intricate network of associations between production parameters. The number of weaned pigs per litter was associated positively with the number of piglets per litter and the farrowing percentage, and negatively with the mortality until weaning. The farrowing percentage was also associated negatively with the mortality in sows and the percentage of sow with unsuccessful inseminations, but positively with the number of litters per sow per year, which in its turn had a positive association with the number of sows. External biosecurity revealed positive associations with both the number of sows and the fact a farm belonged to the SPF system.

AM prescription revealed no association with any variable.

From this, it can by hypothesized that biosecurity has a beneficial effect on production in the sow sector, more than AMs, and so do the requirements for the SPF system. The number of sows can be interpreted as a confounding factor, translating not only the number of animals in a holding but also the level of ambition of the farmer (one could say optimization of the farm). The lack of associations regarding AMs show these drugs do not play a determinant role in assuring the levels of production.

The weaner model also showed no effect of the AMs in production variables, which were not associated with any other parameters. There was, however, association between external biosecurity and farms belonging to the SPF system. This latter variable was negatively associated with farms vaccinating against *Actinobacillus pleuropneumoniae* and PRRS, which might indicate farms with high biosecurity and respecting SPF requirements do not need to vaccinate their animals against these, as the diseases are under control with the measures already in place. Vaccination against PRRS was positively associated with vaccination against *Mycoplasma hyopneumoniae*, which in its turn was associated with vaccination against PCVs, also positively. This reveals a pattern regarding vaccination policies, as farms which vaccinate against one agent are likely to vaccinate against others as part of their vaccination protocols.

These results should be interpreted considering the Danish pig farming context - that is, farms with high biosecurity, reduced AM consumption, several systems in place which ensure optimization of the epidemiological situation of the country, amongst others. It cannot be drawn, from the results of this analysis, that AMs do not benefit production. However, under the circumstances of the farms from which the data used in this ABN analysis was collected, AMs have no impact on the production, probably because several other measures in place minimize the need for them.

It would be interesting to repeat this analysis with a different dataset, with data regarding farms from other countries, with different productivity levels, biosecurity scores, vaccination policies and/or levels of AM consumption. Since different settings would, most likely, produce different networks, analyzing those differences would contribute to better understanding the ways in which AM consumption influences, and is influenced by, all the other relevant factors in pig farming.

## VIII – Bibliography

Aarestrup, F. M., Seyfarth, A. M., Emborg, H., Hendriksen, R., Bager, F., & Pedersen, K. (2001). Effect of Abolishment of the Use of Antimicrobial Agents for Growth Promotion on Occurrence of Antimicrobial Resistance in Fecal Enterococci from Food Animals in Denmark. *Antimicrobial Agents and Chemotherapy*, *45*(7), 2054–2059. https://doi.org/10.1128/AAC.45.7.2054

Anderson, E. S., & Lewis, M. J. (1965). Drug Resistance and its Transfer in Salmonella typhimurium. *Nature*, *206*(4984), 579–583.

Arnold, S., & Straus, S. (2009). Interventions to improve antibiotic prescribing practices in ambulatory care ( Review ). *Cochrane Database of Systematic Reviews*, (4).

https://doi.org/10.1002/14651858.CD003539.pub2.www.cochranelibrary.com

Bi, Y., Yang, C., Diao, Q., & Tu, Y. (2017). Effects of dietary supplementation with two alternatives to antibiotics on intestinal microbiota of preweaned calves challenged with Escherichia coli K99. *Scientific Reports*, *7*(5439), 1–12. https://doi.org/10.1038/s41598-017-05376-z

Boeckel, T. P. Van, Brower, C., Gilbert, M., Grenfell, B. T., & Levin, S. A. (2015). Global trends in antimicrobial use in food animals. *Proceedings of the NAtional Academy of Sciences of the United States of America*, *112*(18), 5649–5654. https://doi.org/10.1073/pnas.1503141112

Borgen, K., Simonsen, G. S., Sundsfjord, A., Wasteson, Y., Kruse, H., & Olsvik, O. (2000). Continuing high prevalence of VanA-type vancomycin-resistant enterococci on Norwegian poultry farms three years after avoparcin was banned. *Journal of Applied Microbiology*, *89*(3), 478–485.

CDC. (2013). *Antibiotic Resistance Threats in the United States, 2013*. *Centers for Disease Control and Prevention*. https://doi.org/CS239559-B

Conly, J., & Johnston, B. (2005). Where are all the new antibiotics ? The new antibiotic paradox. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *16*(3), 159–160.

DanAvl. (n.d.). DanAvl. Retrieved August 6, 2017, from http://www.danavl.com/

Danish Aglicullture & Food Council. (n.d.). The Danish Agriculture & Food Council. Retrieved August 10, 2017, from http://www.agricultureandfood.dk/about-us/fokus

Danish Veterinary and Food Administration. (n.d.). *Special provisions for the reduction of the consumption of antibiotics in pig holdings (the yellow card initiative)*. Glostrup. Retrieved from https://www.foedevarestyrelsen.dk/english/SiteCollectionDocuments/Dyrevelfaerd og veterinaermedicin/Veterinærmedicin/Yellow Card, English version, 180517.pdf

Danish Veterinary and Food Administration. (2015a). *Facts*.

Danish Veterinary and Food Administration. (2015b). The Central Husbandry Register (CHR). Retrieved August 8, 2017, from https://www.foedevarestyrelsen.dk/english/Animal/AnimalHealth/Central_Husbandry_Register/Pages/default.aspx

Danish Veterinary and Food Administration. (2016). Livestock Identification, Registration and Traceability. Retrieved August 8, 2017, from https://www.foedevarestyrelsen.dk/english/Animal/AnimalHealth/Prevention_control_animal_diseases/Livestock_identification_registration_and_traceability/Pages/default.aspx

Danish Veterinary and Food Administration. (2017). Distribution and use of veterinary drugs in Denmark. Retrieved August 12, 2017, from https://www.foedevarestyrelsen.dk/english/Animal/AnimalHealth/Veterinary_medicine/Pages/default.aspx

Danish Veterinary and Food Administration. (2017). The Yellow Card Initiative on Antibiotics. Retrieved June 29, 2017, from https://www.foedevarestyrelsen.dk/english/Animal/AnimalHealth/Pages/The-Yellow-Card-Initiative-on-Antibiotics.aspx

Davies, J., & Davies, D. (2010). Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews*, *74*(3), 417–433. https://doi.org/10.1128/MMBR.00016-10

Dupont, N., Fertner, M., Birkgaard, A. C., Andersen, V. D., Nielsen, G. B., Kruse, A. B., & Knegt, L. V. de. (2017). *Improving institutional memory on challenges and methods for estimation of pig herd antimicrobial exposure based on data from the Danish Veterinary Medicines Statistics Program (VetStat)*. Retrieved from https://arxiv.org/ftp/arxiv/papers/1705/1705.08663.pdf

European Comission. (2005). Ban on antibiotics as growth promoters in animal feed enters into effect. Retrieved April 4, 2017, from http://europa.eu/rapid/press-release_IP-05-1687_en.htm

European Medicines Agency. (2016). *Sales of veterinary antimicrobial agents in 29 European countries in 2014*. Retrieved from http://www.ema.europa.eu/docs/en_GB/document_library/Report/2016/10/WC500214217.pdf

Firestone, S. M., Lewis, F. I., Schemann, K., Ward, M. P., Toribio, J. L. M. L., 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. *Preventive Veterinary Medicine*, *116*(3), 243–251. https://doi.org/10.1016/j.prevetmed.2013.11.015

Friedman, N., & Koller, D. (2003). Being Bayesian About Network Structure . A Bayesian Approach to Structure Discovery. *Machine Learning*, *50*, 95–125.

Ghent University. (2017). Biocheck.ugent. Retrieved August 15, 2017, from http://www.biocheck.ugent.be/

Glickman, M. E., & Dyk, D. A. (2007). *Basic Bayesian Methods*. (W. T. Ambrosius, Ed.), *Methods in Molecular Biology* (Vol. 404). Totowa, New Jersey: Humana Press.

Gonzales, R., Steiner, J. F., Lum, A., & Barrett, P. H. (1999). Decreasing Antibiotic Use in Ambulatory Practice. *JAMA*, *281*(16). Retrieved from http://jamanetwork.com/pdfaccess.ashx?url=/data/journals/jama/4627/

Hao, H., Cheng, G., Iqbal, Z., Ai, X., Hussain, H. I., & Huang, L. (2014). Benefits and risks of antimicrobial use in food-producing animals. *Frontiers in Microbiology*, *5*(June), 1–11. https://doi.org/10.3389/fmicb.2014.00288

Heckerman, D., Geiger, D., & Chickering, D. (1995). Learning Bayesian networks: trhe combination of knowledge and statistical data. *Machine Learning*, *20*(3), 197–243.

https://doi.org/http://dx.doi.org/10.1023/A:1022623210503

Holmes, A. H., Moore, L. S. P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., … Piddock, L. J. V. (2016). Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet*, *387*(10014), 176–187. https://doi.org/10.1016/S0140-6736(15)00473-0

Koivisto, M., & Sood, K. (2004). Exact Bayesian structure discovery in Bayesian networks. *Journal of Machine Learning Research*, *5*, 549–573.

Lewis, F. I., Brülisauer, F., & Gunn, G. J. (2011). Structure discovery in Bayesian networks: An analytical tool for analysing complex animal health data. *Preventive Veterinary Medicine*, *100*(2), 109–115. https://doi.org/10.1016/j.prevetmed.2011.02.003

Lewis, F. I., & Mccormick, B. J. J. (2012). Revealing the Complexity of Health Determinants in Resource-poor Settings. *America Journal of Epidemiology*, *176*(11), 1051–1059. https://doi.org/10.1093/aje/kws183

Lipsitch, M., & Siber, R. (2016). How Can Vaccines Contribute to Solving the Antimicrobial Resistance Problem ? *mBio*, *7*(3), 1–8. https://doi.org/10.1128/mBio.00428-16.Invited

Luan, S., Schooler, L., & Gigerenzer, G. (2011). A Signal-Detection Analysis of Fast-and-Frugal Trees. *Psychological Review*, *118*(2), 316–338. https://doi.org/10.1037/a0022684

Marcusson, L. L., Frimodt-møller, N., & Hughes, D. (2009). Interplay in the Selection of Fluoroquinolone Resistance and Bacterial Fitness. *PLoS Pathogens*, *5*(8). https://doi.org/10.1371/journal.ppat.1000541

Martignon, L., Vitouch, O., Takezawa, M., & Forster, M. R. (2003). Naive and Yet Enlightened : From Natural Frequencies to Fast and Frugal Decision Trees. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgement and Decision Making* (pp. 189–212). Chichester: John Wiley & Sons, Ltd. https://doi.org/10.1002/047001332X.ch10

Mateu, E., & Martin, M. (2001). Why is anti-microbial resistance a veterinary problem as well? *Journal of Veterinary Medicine, Series B*, *48*(8), 569–581. https://doi.org/10.1046/j.1439-0450.2001.00475.x

Michael, C. A., Dominey-howes, D., Labbate, M., Maria, C., & Elisabeth, J. (2014). The antimicrobial resistance crisis : causes , consequences , and management. *Frontiers in Public Health*, *2*(September). https://doi.org/10.3389/fpubh.2014.00145

Munita, J., & Arias, C. (2016). Mechanisms of Antibiotic Resistance. *Microbiology Spectrum*, *4*(2). https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.Mechanisms

Pasma, T., Furness, M. C., Alves, D., & Aubry, P. (2016). Outbreak investigation of porcine epidemic diarrhea in swine in Ontario. *The Canadian Veterinary Journal*, *57*(1), 4–8.

Penicillin’s finder assays its future. (1945). *New York Times*, p. 21.

Pérez-bosque, A., Polo, J., & Torrallardona, D. (2016). Spray dried plasma as an alternative to antibiotics in piglet feeds , mode of action and biosafety. *Porcine Health*

*Management*, *2*(16), 1–10. https://doi.org/10.1186/s40813-016-0034-1

Pittavino, M., Lewis, F., & Furrer, R. (n.d.). r-bayesian-networks. Retrieved April 4, 2017, from http://www.r-bayesian-networks.org/

Pittavino, M., Lewis, F., & Furrer, R. (2016). *abn : an R package for modelling data using additive Bayesian networks*. Retrieved April 4, 2017 from https://www.researchgate.net/publication/303371852_abn_an_R_package_for_modelling_multivariate_data_using_additive_Bayesian_networks

Postma, M., Backhans, A., Collineau, L., Loesken, S., Sjölund, M., Belloc, C., … Dewulf, J. (2015). The biosecurity status and its associations with production and management characteristics in farrow-to- fi nish pig herds. *Animal*, *10*(3), 478–489. https://doi.org/10.1017/S1751731115002487

Postma, M., Backhans, A., Collineau, L., Loesken, S., Sjölund, M., Belloc, C., … Dewulf, J. (2016). Evaluation of the relationship between the biosecurity status , production parameters , herd characteristics and antimicrobial usage in farrow-to-finish pig production in four EU countries. *Porcine Health Management*, *2*(9), 1–11. https://doi.org/10.1186/s40813-016-0028-z

Sharma, S. (1996). *Applied Multivariate Techniques*. (T. Kent & P. Sellers, Eds.). New York: John Wiley & Sons, Inc.

Spellberg, B., & Gilbert, D. N. (2014). The Future of Antibiotics and Resistance : A Tribute to a Career of Leadership by John Bartlett. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *59*(Suppl 2), S71–S75. https://doi.org/10.1093/cid/ciu392

SPF-Danmark. (n.d.-a). SPF status designations. Retrieved August 11, 2017, from https://www.spf.dk/en-us/health/spf-status-designations/

SPF-Danmark. (n.d.-b). The Danish SPF system. Retrieved August 11, 2017, from https://www.spf.dk/en-us/health/the-danish-spf-system

Statens Serum Institut, DTU VET - National Veterinary Institute, & DTU Food - National Food Institute. (2015). *DANMAP 2015 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark*.

Temtem, C., Kruse, A. B., Nielsen, L. R., Pedersen, K. S., & Alban, L. (2016). Comparison of the antimicrobial consumption in weaning pigs in Danish sow herds with different vaccine purchase patterns during 2013. *Porcine Health Management*, *2*(23), 1–11. https://doi.org/10.1186/s40813-016-0042-1

Thacker, P. A. (2013). Alternatives to antibiotics as growth promoters for use in swine production : a review. *Journal of Animal Science and Biotechnology*, *4*(35), 1–12. https://doi.org/10.1186/2049-1891-4-35

Ventola, C. L. (2015). The Antibiotic Resistance Crisis. *Pharmacy & Therapeutics*, *40*(4), 277–283. https://doi.org/Article

Wang, S., Zeng, X., Yang, Q., & Qiao, S. (2016). Antimicrobial Peptides as Potential Alternatives to Antibiotics in Food Animal Industry. *International Journal of Molecular Sciences*, *17*(603). https://doi.org/10.3390/ijms17050603

WHO Collabotaring Centre for Drug Statistics Methodology. (2016). DDD - Definition and general considerations. Retrieved August 13, 2017, from https://www.whocc.no/ddd/definition_and_general_considera/

# Appendix I – Variables' summary

## Summary of the continuous variables

| Variable | Mean | Minimum | 1st quartile | Median | 3rd quartile | Maximum |
|---|---|---|---|---|---|---|
| **Pig total** | 2878 | 695 | 1796 | 2610 | 3670 | 7480 |
| **Number of sows** | 615 | 150 | 4250 | 600 | 750 | 1500 |
| **Number of weaners** | 1805 | 40 | 950 | 1700 | 2500 | 550 |
| **Number of Finishers** | 458,6 | 0 | 100 | 240 | 600 | 2688 |
| **Number of employees** | 4 | 1 | 3 | 4 | 5 | 15 |
| **Number of years of experience of the farmer** | 25,64 | 3 | 20 | 25 | 30 | 50 |
| **Age of the youngest building** | 9,17 | 0 | 6 | 9 | 12 | 25 |
| **Age of the oldest building** | 45,7 | 6 | 21 | 39 | 55 | 165 |
| | | | | | | |
| **Total AMP for sows** | 2,266 | 0,251 | 1,639 | 2,175 | 2,928 | 4,708 |
| **AMP for sows for gastrointestinal indications** | 0,262 | 0,002 | 0,065 | 0,196 | 0,393 | 2,147 |
| **AMP for sows for respiratory indications** | 0,494 | 0,018 | 0,216 | 0,428 | 0,685 | 2,438 |
| **AMP for sows for individual treatment** | 2,038 | 0,250 | 1,394 | 1,984 | 2,626 | 4,708 |
| **AMP for sows for group treatment** | 0,373 | 0,006 | 0,113 | 0,262 | 0,473 | 1,674 |
| | | | | | | |
| **Total AMP for weaners** | 10,3 | 0,559 | 6,062 | 9,981 | 14,516 | 22,807 |
| **AMP for weaners for gastrointestinal indications** | 7,634 | 0,041 | 3,238 | 6,646 | 11,249 | 22,738 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **AMP for weaners for respiratory indications** | 3,35 | 0,033 | 0,751 | 3,177 | 4,775 | 13,046 |
| **AMP for weaners for individual treatment** | 1,687 | 0,097 | 0,484 | 1,038 | 2,054 | 12,556 |
| **AMP for weaners for group treatment** | 9,253 | 0,285 | 5,367 | 8,84 | 12,175 | 22,25 |
| | | | | | | |
| **Total AMP for finishers** | 1,853 | 0,044 | 0,662 | 1,6 | 2,745 | 6,572 |
| **AMP for finishers for gastrointestinal indications** | 1,59 | 0,029 | 0,469 | 1,612 | 2,382 | 4,059 |
| **AMP for finishers for respiratory indications** | 0,96 | 0,015 | 0,234 | 0,527 | 1,279 | 3,805 |
| **AMP for finishers for individual treatment** | 0,917 | 0,044 | 0,213 | 0,697 | 1,133 | 4,125 |
| **AMP for finishers for group treatment** | 1,983 | 0,178 | 1,438 | 1,73 | 2,569 | 6,23 |
| | | | | | | |
| **Mortality in sows** | 9,5% | 0,1% | 7,75% | 9,5% | 11,9% | 19,9% |
| **Feed used for sows** | 1501 | 1171 | 1409 | 1506 | 1593 | 1981 |
| **Farrowing percentage** | 87,22% | 73% | 85% | 88,13% | 90% | 93% |
| **Percentage of sows with unsuccessful inseminations** | 5,91% | 1,28% | 4,16% | 5% | 7,83% | 13% |
| **Litters/sow/year** | 2,266 | 2,04 | 2,23 | 2,27 | 2,31 | 2,42 |
| **Piglets/litter** | 15,68 | 13,7 | 15,26 | 15,62 | 16,1 | 17,36 |
| **Stillborn piglets/litter** | 1,664 | 0,82 | 1,47 | 1,7 | 1,9 | 2,3 |
| **Days suckling pigs stay with the sow** | 30.38 | 24 | 28,16 | 30 | 32,32 | 40,19 |
| **Mortality until weaning** | 13,27% | 6,85% | 11,5% | 13% | 15% | 20% |
| **Weaned pigs/litter** | 13,6 | 11,8 | 13,19 | 13,6 | 14,1 | 15,57 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Mortality in weaners** | 2,9% | 0,9% | 2,22% | 2,8% | 3,44% | 6% |
| **Weight gain in weaners** | 441,5 | 295 | 414 | 442 | 462 | 640 |
| **Feed Conversion Ratio** | 1,92 | 1,64 | 1,83 | 1,92 | 2,03 | 2,29 |
| | | | | | | |
| **Total biosecurity score** | 76,61 | 60 | 73 | 77 | 80 | 88 |
| **External biosecurity score** | 85,88 | 67 | 83 | 87 | 90 | 96 |
| **Internal biosecurity score** | 66,89 | 48 | 62 | 67 | 72 | 89 |

## Summary of the discrete variables

Region:

- North Denmark Region: 35 farms
- Central Denmark Region: 39 farms
- South Denmark Region: 57 farms
- Region of Zealand (including Capital Region of Denmark): 26 farms

Type of farm:

- “Full” (full production cycle): 38 farms
- “Remain” (farms with finishers as remainder of the main production): 96 farms
- “No” (farms with no finishers): 23 farms

Exporting status:

- Farms exporting: 54
- Farms not exporting: 103

SPF status:

- Farms belonging to the SPF system: 136
- Farms not belonging to the SPF system: 21

SPF class:

- Red: 10 farms
- Blue: 126 farms
- Conventional (not part of the SPF system): 21

Vaccination against PCV2:

- Farms vaccinating: 119
- Farms not vaccinating: 38

Vaccination against *Mycoplasma hyopneumoniae*:

- Farms vaccinating: 91
- Farms not vaccinating: 66

Vaccination against *Actinobacillus pleuropneumoniae*:

- Farms vaccinating: 20
- Farms not vaccinating: 137

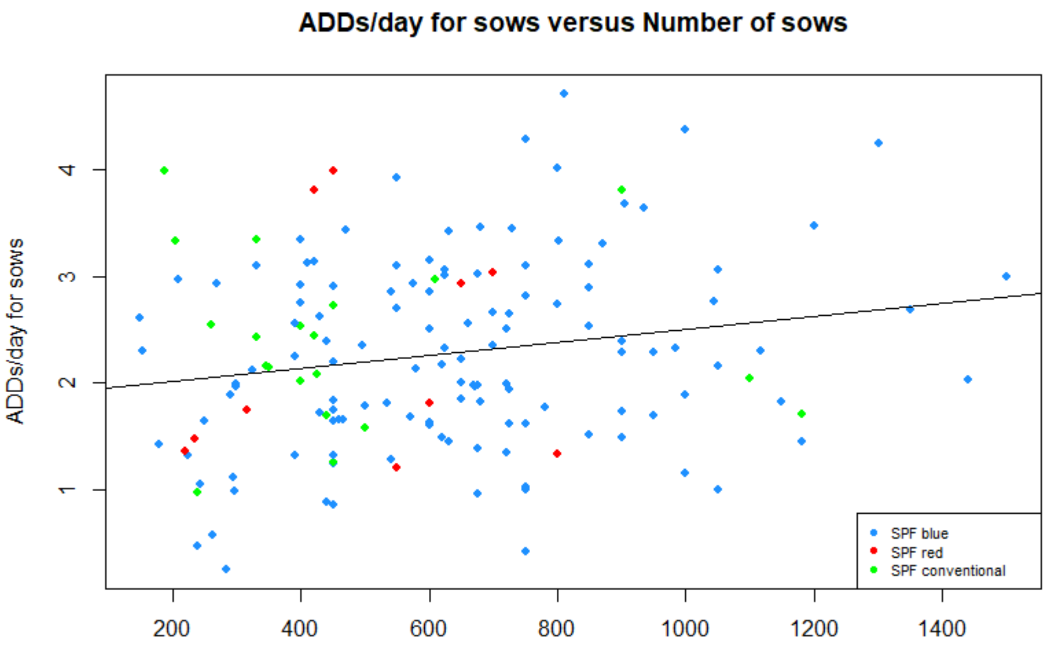
Vaccination against PRRS:

- Farms vaccinating: 23
- Farms not vaccinating: 134

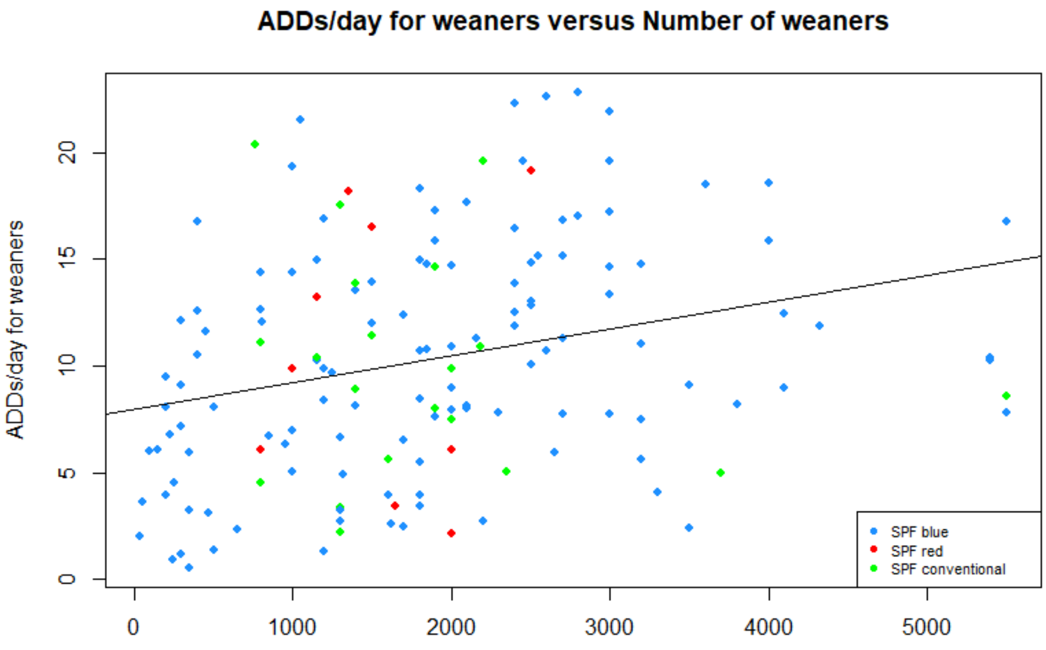
Vaccination against *Lawsonia*:

- Farms vaccinating: 19
- Farms not vaccinating: 138

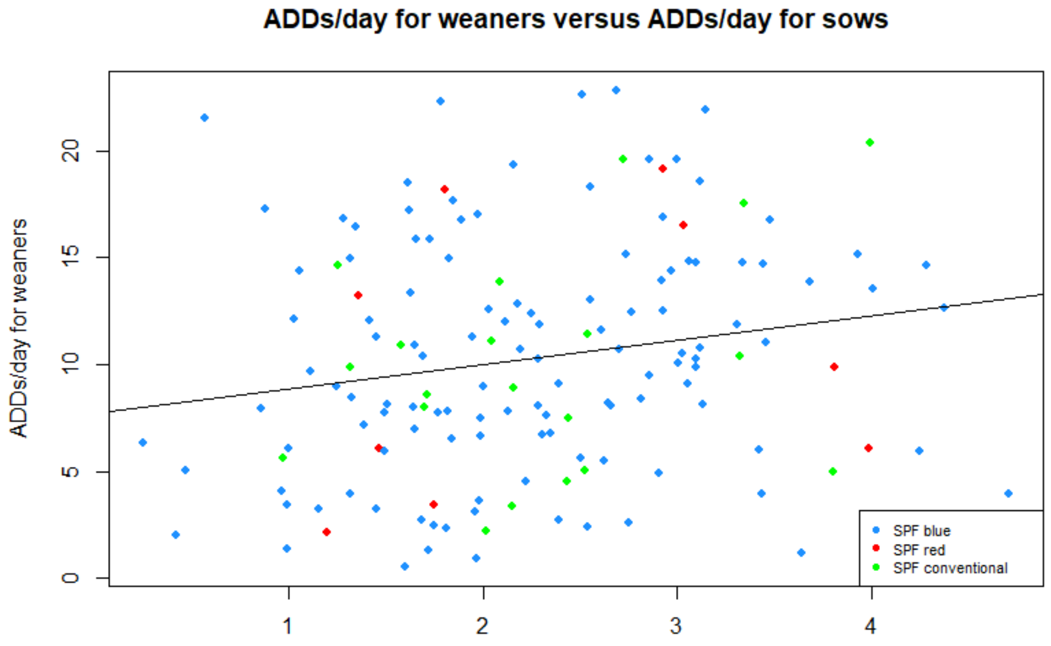
# Appendix II – Significant linear associations between the variables



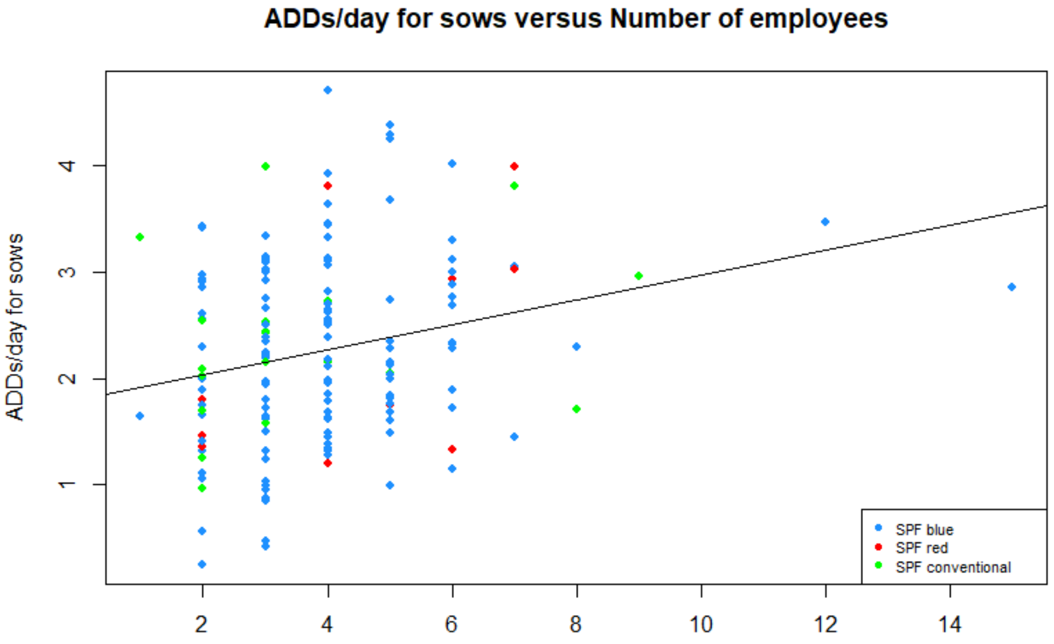
Number of sows
(157 observations, r squared = 0.036, p-value = 0.01728, linear model: y = 0.00061x + 1.89296)



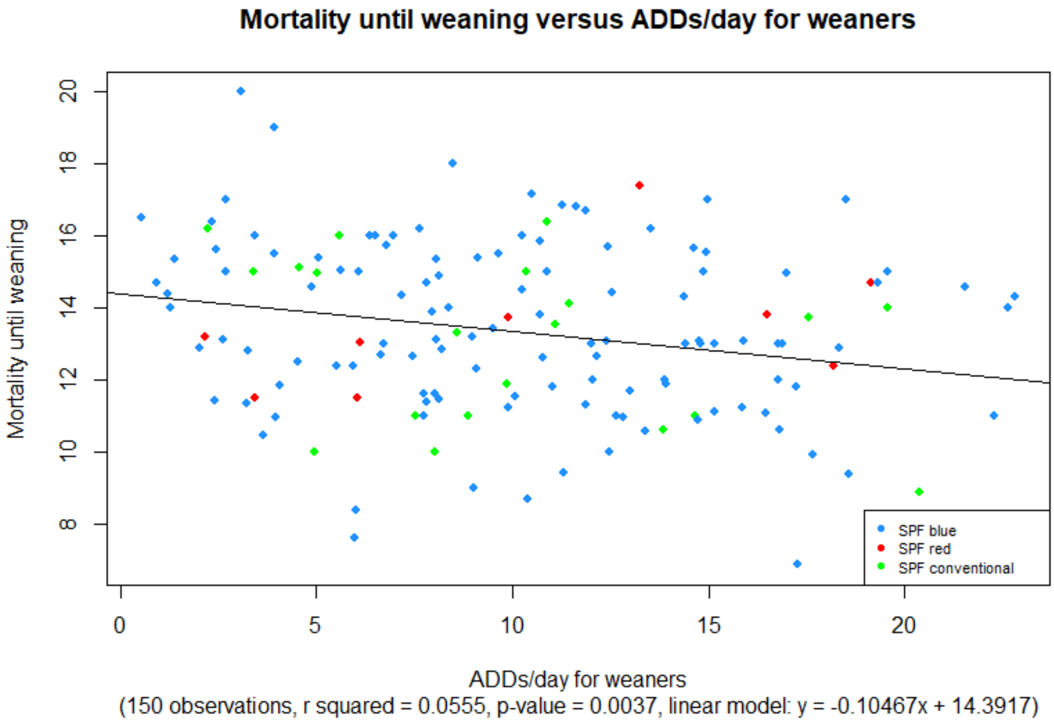
Number of weaners
(150 observations, r squared = 0.0769, p-value = 0.00059, linear model: y = 0.00125x + 7.95232)

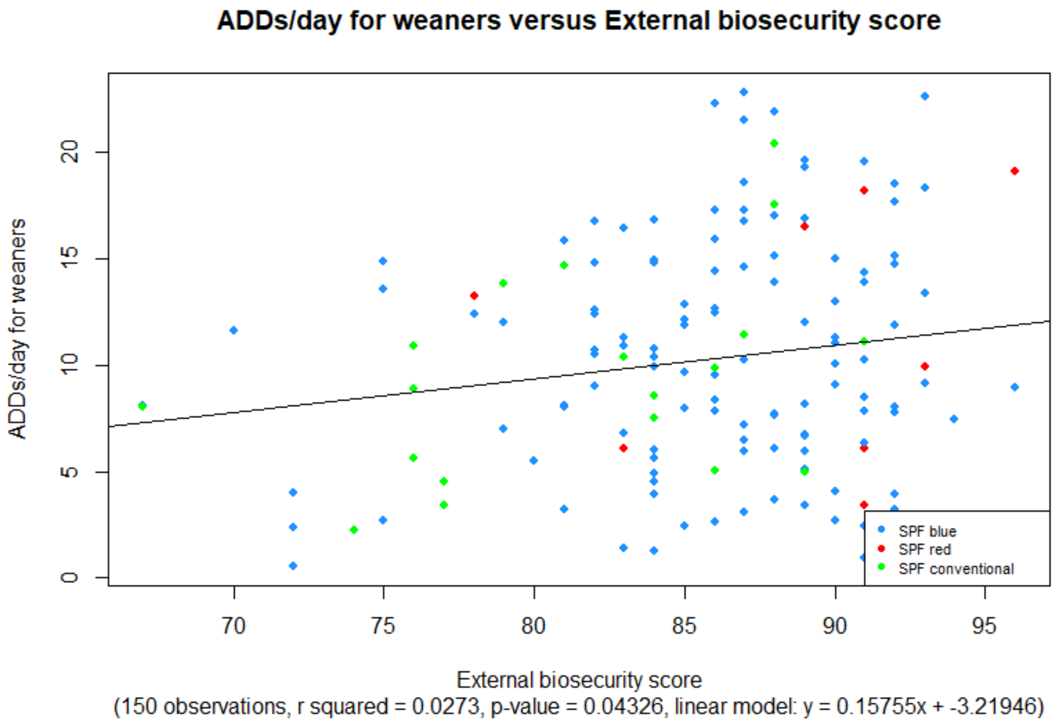


ADDs/day for sows
(150 observations, r squared = 0.0349, p-value = 0.02213, linear model: y = 1.13806x + 7.73031)

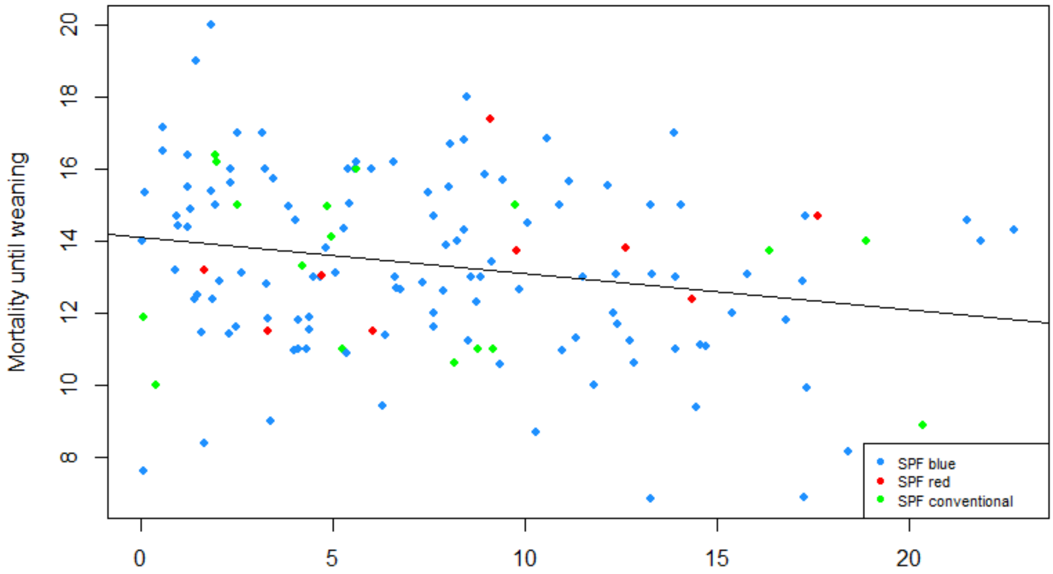


Number of employees
(157 observations, r squared = 0.0607, p-value = 0.00188, linear model: y = 0.11745x + 1.8015)


Mortality until weaning versus ADDs/day for weaners
Mortality until weaning
8
10
12
14
16
18
20
0
5
10
15
20
SPF blue
SPF red
SPF conventional
ADDs/day for weaners
(150 observations, r squared = 0.0555, p-value = 0.0037, linear model: y = -0.10467x + 14.3917)

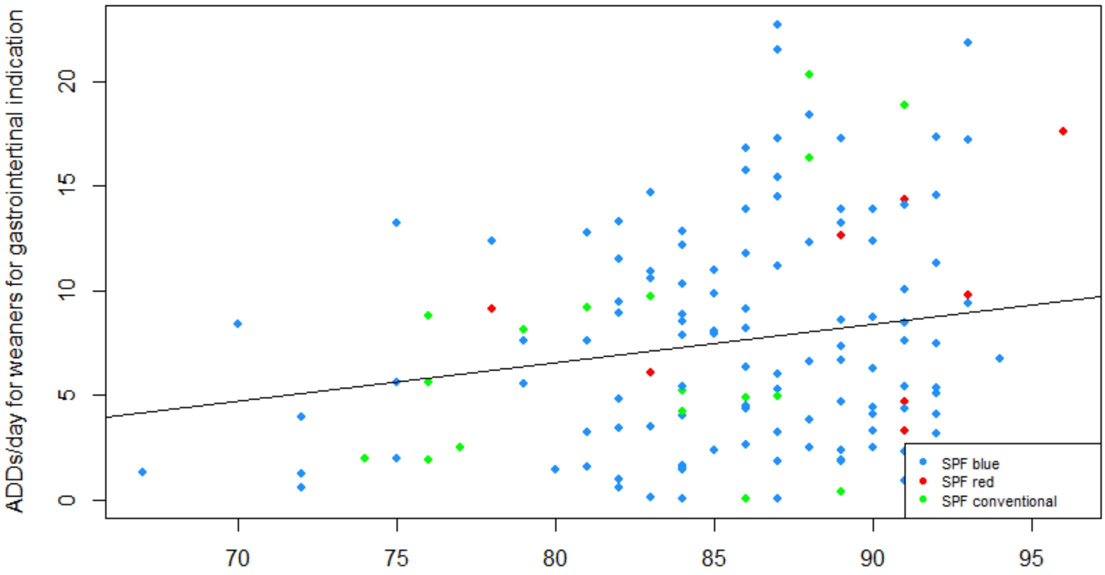

ADDs/day for weaners versus External biosecurity score
ADDs/day for weaners
0
5
10
15
20
70
75
80
85
90
95
SPF blue
SPF red
SPF conventional
External biosecurity score
(150 observations, r squared = 0.0273, p-value = 0.04326, linear model: y = 0.15755x + -3.21946)

## Mortality until weaning versus ADDs/day for weaners for gastrointertinal indication



ADDs/day for weaners for gastrointertinal indication
(143 observations, r squared = 0.0495, p-value = 0.00755, linear model: y = -0.09971x + 14.09139)

## ADDs/day for weaners for gastrointertinal indication versus External biosecurity score



External biosecurity score
(143 observations, r squared = 0.0354, p-value = 0.02444, linear model: y = 0.18526x + -8.26967)

# Appendix III – Frequency of arc retrieval for the sow model boostraping

| | SPF blue | SPF red | SPF conventional | Number of sows | ADDs/day for sows | ADDs/day for weaners | Mortality in sows |
|---|---|---|---|---|---|---|---|
| **SPF blue** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SPF red** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SPF conventional** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Number of sows** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **ADDs/day for sows** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **ADDs/day for weaners** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Mortality in sows** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Feed used for sows** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Farrowing %** | 0 | 0 | 0 | 0 | 0 | 0 | 591 |
| **% sows unsuccess. insemin.** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Litters/sow/year** | 0 | 0 | 0 | 278 | 0 | 0 | 0 |
| **Piglets/litter** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Stillborn piglets/litter** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Mortality until weaning** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Weaned pigs/litter** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **External biosecurity** | 0 | 0 | 928 | 840 | 0 | 0 | 0 |
| **Internal biosecurity** | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The table above shows the frequency at which each arc between two given nodes was retrieved, over the 5000 searched executed during the parametric bootstrapping. Columns indicate parent variables and rows represent child variables. Arcs were considered as robust enough to be featured in the final DAG, should the sum of the number of arcs retrieved between two nodes in any of the two possible directions be over 2500 (50% of the searches). Due to the sheer size of the table, it was necessary to spread it across two adjacent pages.

| Fee use fo sow | Farrowing % | % sows unsuccessful Insemination | Litters / sow/ year | Piglets / litter | Stillborn piglets/litter | Mortality until weaning | Weaned pigs/litter | External biosecurity | Internal biosecurity |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1912 | 0 |
| 0 | 0 | 0 | 2945 | 0 | 0 | 0 | 0 | 1945 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2862 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 291 | 378 | 0 | 0 | 0 | 4381 | 0 | 0 |
| 0 | 4709 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 4610 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 |
| 0 | 10 | 0 | 0 | 4986 | 0 | 4984 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

## Appendix IV – Frequency of arc retrieval for the weaner model boostraping

| | Part of SPF system | Number of weaners | PCV2 vaccination | Mycoplasma vaccination | Actinobacillus vaccination | PRRS vaccination |
|---|---|---|---|---|---|---|
| **Part of SPF system** | 0 | 0 | 0 | 0 | 3221 | 2435 |
| **Number of weaners** | 0 | 0 | 0 | 0 | 0 | 0 |
| **PCV2 vaccination** | 0 | 0 | 0 | 1819 | 0 | 0 |
| **Mycoplasma vaccination** | 0 | 0 | 1360 | 0 | 0 | 2989 |
| **Actinobacillus vaccination** | 716 | 0 | 0 | 0 | 0 | 0 |
| **PRRS vaccination** | 543 | 0 | 0 | 756 | 0 | 0 |
| **Lawsonia vaccination** | 0 | 0 | 0 | 0 | 0 | 0 |
| **ADDs/day for sows** | 0 | 0 | 0 | 0 | 0 | 0 |
| **ADDs/day for weaners** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Weaned pigs/litter** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Mortality in weaners** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Weight gain** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Feed conversion ratio** | 0 | 0 | 0 | 0 | 0 | 0 |
| **External biosecurity** | 149 | 0 | 0 | 0 | 0 | 0 |
| **Internal biosecurity** | 0 | 0 | 0 | 0 | 0 | 0 |

The table above shows the frequency at which each arc between two given nodes was retrieved, over the 5000 searched executed during the parametric bootstrapping. Columns indicate parent variables and rows represent child variables. Arcs were considered as robust enough to be featured in the final DAG, should the sum of the number of arcs retrieved between two nodes in any of the two possible directions be over 2500 (50% of the searches). Due to the sheer size of the table, it was necessary to spread it across two adjacent pages.

| Lawsonia vaccination | ADDs/day for sows | ADDs/day for weaners | Weaned pigs/litter | Mortality in weaners | Weight gain | Feed conversion ratio | External bio security | Internal bio security |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3976 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|  |  |  |  |  |  |  |  |  |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |