**DTU Library**

# Sampling pig farms at the abattoir in a cross-sectional study  Evaluation of a sampling method

**Birkegård, Anna Camilla; Hisham Beshara Halasa, Tariq; Toft, Nils**

Title: Sampling pig farms at the abattoir in a cross-sectional study – evaluation of a sampling method.

Anna Camilla Birkegård[1,*], Tariq Halasa[1], Nils Toft[1]

[1]Department for Diagnostics and Scientific Advice, National Veterinary Institute, Technical University of Denmark, Kemitorvet, Building 204, 2800 Kgs Lyngby, Denmark

*Corresponding author, acbir@vet.dtu.dk

**Abstract**

A cross-sectional study design is relatively inexpensive, fast and easy to conduct when compared to other study designs. Careful planning is essential to obtaining a representative sample of the population, and the recommended approach is to use simple random sampling from an exhaustive list of units in the target population. This approach is rarely feasible in practice, and other sampling procedures must often be adopted. For example, when slaughter pigs are the target population, sampling the pigs on the slaughter line may be an alternative to on-site sampling at a list of farms. However, it is difficult to sample a large number of farms from an exact predefined list, due to the logistics and workflow of an abattoir. Therefore, it is necessary to have a systematic sampling procedure and to evaluate the obtained sample with respect to the study objective.

We propose a method for 1) planning, 2) conducting, and 3) evaluating the representativeness and reproducibility of a cross-sectional study when simple random sampling is not possible. We used an example of a cross-sectional study with the aim of quantifying the association of antimicrobial resistance and antimicrobial consumption in Danish slaughter pigs. It was not possible to visit farms within the designated timeframe. Therefore, it was decided to use convenience sampling at the abattoir. Our approach was carried out in three steps: 1) planning: using data from meat inspection to plan at which abattoirs and how many farms to sample; 2) conducting: sampling was carried out at five abattoirs; 3) evaluation: representativeness was evaluated by comparing sampled and non-sampled farms, and the reproducibility of the study was assessed through simulated sampling based on meat inspection data from the period where the actual data collection was carried out.

1

In the cross-sectional study samples were taken from 681 Danish pig farms, during five weeks from February to March 2015. The evaluation showed that the sampling procedure was reproducible with results comparable to the collected sample. However, the sampling procedure favoured sampling of large farms. Furthermore, both under-sampled and over-sampled areas were found using scan statistics.

In conclusion, sampling conducted at abattoirs can provide a spatially representative sample. Hence it is a possible cost-effective alternative to simple random sampling. However, it is important to assess the properties of the resulting sample so that any potential selection bias can be addressed when reporting the findings.

**Keywords**

Sampling procedure, Pig farm, Cross-Sectional study, Slaughterhouse, Abattoir

**Introduction**

The cross-sectional study is a commonly used design in veterinary epidemiology. Compared to other study designs, it is relatively inexpensive, fast, and easy to conduct. An ideal sample would be obtained through simple random sampling, where samples are obtained by randomly selecting from a list of all units in the target population. In simple random sampling every subject has the same chance to be sampled (Dohoo et al., 2010). Other random sampling procedures such as stratified or systematic random sampling also exist (Dohoo et al., 2010). When sampling from live individuals, the random component of the sampling is often impeded by logistic, financial, or biological constraints. Many that have designed and carried out a study involving sampling of live animals, for example at farms, would agree on that the resulting sample is often not obtained completely at random or the characteristics of the farms have changed after they were chosen for sampling. Therefore, in reality many resort to non-probability sampling when it is not possible to use an exact method for determining a subjects probability for sampling (Dohoo et al., 2010).

Regardless of the sampling procedure, the sample must be representative of the target population or at least the part of the target population that the researcher wishes to draw conclusions about. While careful planning is essential to obtaining a sample that meets the criteria of the study (Houe et al., 2004), an evaluation of the representativeness of the sample after collection could provide valuable information. In the 1960s, John Tukey (1962) introduced the idea that it is important to assess if your assumptions are correct before initiating statistical analyses. When reporting a cross-sectional

2

study, the STROBE statement (von Elm et al., 2007) suggests that the eligibility criteria, as well as the sources and methods of sample selection are defined, together with a descriptive summary of the sample. However, an evaluation of the representativeness of sample and the extent to which it meets the eligibility criteria is not required. Therefore, it could be valuable to evaluate the level of sample bias that might be introduced as a result of limited access to the entire target population or due to the sampling procedure, particularly in studies where practical constraints prohibit simple random sampling. For example, the most practical way to obtain a random sample of pig farms would be to sample at the farms. However, due to the strict biosecurity measures at pig farms (Boklund et al., 2004), it is difficult to sample a large number of farms in a short period of time. Depending on the purpose of the study, a convenience sampling procedure (such as sampling pigs at the abattoir) could be a valuable substitute. Convenience sampling is a non-probability approach where the samples are collected because they are easy to obtain rather than sampling at random (Dohoo et al., 2010). Although convenience sampling at the abattoir has previously been explored (DANMAP, 2016; van den Bogaard et al., 2000; Wegener et al., 2003), this study describes how the idea can be used on a larger scale in order to obtain samples from a bigger number of farms. Convenience sampling at the abattoir is inexpensive compared to farm visits and has no negative impact on animal welfare, as sampling from slaughtered pigs would not harm the pigs. However, the consequence of sampling over a short time period at the abattoir is that it is not possible to plan in advance which farms to sample. Records of the farms sending pigs to slaughter are only available close to the date of slaughter. The representativeness of the sample must therefore be evaluated after sampling.

The objective of the study was to present a method for planning and conducting a cross-sectional study when simple random sampling is not possible and to evaluate if the study criteria have been fulfilled for the obtained sample. Our approach is exemplified using a cross-sectional study of pig farms sampled at the abattoir with the aim of quantifying the association of antimicrobial resistance and antimicrobial consumption in Danish pigs.

**Materials and methods**

**Study set-up**

To demonstrate our approach, a cross-sectional study of the association between antimicrobial resistance genes at farm level in pigs, and factors such as geographical location, farm size, and antimicrobial consumption was used. To meet the objectives of such a study, it was necessary to

3

obtain a random sample of farms that were representative of the Danish pig population in terms of geographical location and farm size.

We conducted our study in three consecutive steps:

1) Planning the cross-sectional study
2) Conducting the cross-sectional study
3) Evaluating the sampling procedure
   a. Evaluation of the collected sample from the cross-sectional study in terms of farm size and spatial randomness
   b. Evaluation of the reproducibility of the sampling procedure by simulating it using meat inspection data. The sampling procedure was assessed to be reproducible if repeating the sampling would result in a sample that was comparable with the collected sample in terms of farm size and spatial distribution

Thereby, two set of samples were considered: a) the sample obtained in the cross-sectional study, where actual faecal samples were collected, hereafter referred to as the collected sample; b) simulated samples, used to examine the reproducibility of the sampling procedure, where information about the farms were obtained from the meat inspection data.

**Register data**

Data from three registers (the Central Husbandry Register (CHR), meat inspection data, and the database for pig movements) were used.

All pig farms in Denmark are registered in the national CHR database with a unique identification number (CHR-number). The CHR is owned and maintained by the Danish Veterinary and Food Administration. Information in the CHR includes Cartesian coordinates (given in "UTM EUREF89 zone 32" format) and the number of pigs for three age groups: breeding animals, weaners (7-30 kg) and finishers (pigs above 30 kg excluding breeding animals)(Anonymous, 2016a). The CHR also contains information on abattoirs, rendering plants, cooling facilities for dead animals, animal fairs and export stables.

The meat inspection data are owned and maintained by the Danish Classification inspection. The data include delivery number, CHR-number, date of slaughter, and the ID of the abattoir (Anonymous, 2016b). The delivery number is tattooed on the ham of the pig and can be read on the carcass on the slaughter line. The delivery number was used during sampling to identify the farm from which the pigs originated.

4

The database for pig movements contains information about of movement of dead and live pigs. The database includes CHR-numbers of receiving and sending farms or holdings, the number of pigs moved and the date of the movement (Anonymous, 2016c).

**Step one: Planning the cross-sectional study**

The cross-sectional study serving as an example for our approach was designed to meet the following criteria:

1. Representativeness based on farm size, which was estimated using two different variables – the productivity and the number of finisher pigs registered in the CHR. Productivity was calculated as the cumulative delivery over 26 weeks of pigs for slaughter, traced back from the first day of sampling. This approximately corresponds to two batches of finishers, in order to reduce the influence of seasonal variation (Moodley et al., 2011). Both number of finishers and productivity were used in order to account for the inaccuracy in the CHR and the fact that not all pigs were slaughtered in Denmark, which would influence the estimated productivity of some farms.

2. Spatial randomness of locations of sampled farms compared to non-sampled farms in the target population.

The target population is the population that we want to draw conclusions on (Houe et al., 2004). For the cross-sectional study the target population was Danish farms with a conventional production of finishers. The target population was defined as farms included in a data extraction from the CHR database on 19th January 2015. Farms that did not have finishers registered or that were registered as organic, free-range, trade or hobby farms were excluded from the target population because these farm types have a farm management system that differ from the conventional production farms. Furthermore, all farms located on Bornholm (124 farms) and farms without coordinates in the CHR data extraction (3 farms) were excluded. The information for the target population was obtained from the CHR which optimally only contain information on active farms. However, it happens that the farmer have not inactivated the farm in the register while the farm is empty. Therefore, we used the database for pig movements to estimate if the farm was active or inactive. The farms were considered to be active if at least one movement (either into or out of the farm) was registered in the movement data during the period between September 2014 and March 2015. Only active farms were included in the target population, resulting in the exclusion of 358 inactive farms. The final target population consisted of 5,654 farms.

The sampling frame did not cover all farms in the target population, as not all farms sent pigs to the sampling sites. The 3,274 farms included in the sampling frame were defined as the study population, i.e. the population that the samples are taken from (Houe et al., 2004). The study population was used to evaluate the choice of sampling site. The remaining farms did not send any pigs to slaughter during the period, sent pigs to slaughter outside of Denmark, or sent pigs to an abattoir that was not in the sampling frame.

It was decided that only five of the seven major Danish-owned abattoirs for finisher pigs would be used as sampling sites. Statistics Denmark estimated that a total of 1,531,600 finishers were slaughtered in Denmark in February 2015 (Statistics Denmark, 2015), and 1,365,963 (90%) of these were slaughtered at the seven abattoirs (calculated based on meat inspection data). The reasons for excluding two of the abattoirs were that one primarily slaughtered pigs from free-range and organic farms, and the other was located on the remote island of Bornholm. By mapping the location of the farms that sent pigs to the sampling sites in 2014, it was apparent that each of the five chosen sampling sites covered a certain region of Denmark (Fig. 1).

The next step in planning the sampling was to determine the number of farms that should be sampled per abattoir. There was a chance of sampling the same farms more than once and a chance of sampling farms that did not match the target population. Given the resources available to the project, we therefore planned to have 800 farms with finishers in the sample. The number of farms per sampling site was calculated based on the number of farms that sent pigs to the abattoir in weeks 6-10 and weeks 46-50 of 2014 using the meat inspection data. These two periods were chosen in order to take into account seasonal variation and a re-organisation of the abattoir structure in the autumn of 2014.

**Step two: Conducting the cross-sectional study**

Sampling took place during weeks 6-10 of 2015. Reaching the sample size required a total of 4-7 hours per day over 16 active sampling days with sampling from one or two abattoirs per day. A previous study (Clasen et al. 2016) showed that a pooled sample of faeces from five individual pigs was sufficient to represent the antimicrobial resistance level of the farm. Our sampling approach therefore required that at least five pigs from the same farm were slaughtered on the same day.

**Step three: Evaluation of the sampling procedure**

Evaluation of the sampling procedure was conducted in two parts. First, the representativeness of the farms in the collected sample was assessed by comparing farm size and spatial distribution with

6

that of the non-sampled farms in the target population and with the non-sampled farms in the study population. Reproducibility was assessed by simulating samples using information from the meat inspection data about farms sending pigs to slaughter on the days when of data collection occurred.

Evaluation of representativeness of the collected sample

In order to evaluate the representativeness of the farms in the collected sample with regard to farm size, the registered number of finishers and estimated productivity (estimated as the number of pigs delivered during 26 weeks for slaughter) of the farms included in the collected sample were compared to the farms in the target and study populations that were not included in the collected sample. A Wilcoxon test was used to compare the median productivity of the farms in the collected sample with both the non-sampled farms in the target population, and the study population as well as to compare the number of finishers at the farms in the collected sample with the non-sampled farms in the study and target population.

To evaluate whether or not the farms in the collected sample were randomly spatially distributed among the non-sampled farms, a purely spatial cluster analysis was performed using the "rsatscan" package (Kleinman, 2015), which allows the spatial cluster analysis to be run in SatScan (Kulldorff M. and Information Management Services Inc, 2015) from R (R Core Team, 2017). The Scan statistics were based on a Bernoulli model (Kulldorff, 1997). The sampled farms were defined as cases and the non-sampled farms in the target and study populations were defined as controls. In the Bernoulli model, the null hypothesis is that there is an equal distribution of cases and controls both inside and outside of the search window (Kulldorff, 1997). The spatial scanning window was set to be circular, include a maximum level of 25% of the population, and centre on the coordinates of each farm in turn. Only secondary spatial clusters with no geographical overlap were allowed. The analysis was run as two-sided tests, scanning for areas with sampling rates that were higher or lower than expected under the null hypothesis. The relative risk (RR) to be sampled for a specific site within an identified cluster in the cross-sectional study was calculated as the ratio of the observed to the expected number of sampled farms within that cluster. The spatial scan statistics used 999 Monte Carlo replications to estimate the significance levels of these clusters. The spatial cluster was considered to be significant if the p-value was below 0.05. Spatial clusters with RR>1 were defined as over-sampled areas, and clusters with RR<1 were defined as under-sampled areas. The SatScan analysis gives the point location of the centre and the radius of the clusters (here defined as circles). The results of the scan statistics were plotted onto a map of Denmark (excluding Bornholm) using the "sp" package (Pebesma and Bivand, 2005) in R (R Core Team, 2017).

7

Evaluating the reproducibility of the sampling procedure through simulation

In order to evaluate the reproducibility of the proposed sampling procedure, a simulation study was conducted by sampling from meat inspection data. To simulate the sampling, the following three assumptions were made:

1. Sampling would take place in the same period as the cross-sectional study
2. For the total number of farms to sample, three scenarios were chosen:
    i. Using the same number of farms to sample per abattoir as in the cross-sectional study.
    ii. Doubling the number of farms to sample compared to the cross-sectional study.
    iii. Recalculating the number of farms to sample per abattoir using meat inspection data from the sampling period instead of using previous data. This resulted in a target of 120, 290, 180, 120 and 90 farms to sample for each of the abattoirs A-E, respectively.
3. Two cut-off values were chosen in order to determine the smallest number of pigs a farm could send to the abattoir on the day of sampling before being included in the sampling:
    a. A cut-off value of 20 pigs sent for slaughter per day. This number was chosen as five was sampled per farm to be sampled and it was estimated that it would be possible to sample every second pig arriving from the same farm, and that the abattoirs would divide each batch into two slaughter lines.
    b. An abattoir-specific cut-off value was defined as the minimum number of pigs sent to slaughter on the day of sampling in the cross-sectional study for the sampled farms. The cut-off values for abattoirs A to E were 21, 10, 15, 8 and 18, respectively.

Assumption 3 was included because we anticipated that it might not be possible to sample five pigs from some farms sending only a small number of pigs to slaughter in one day. Assumptions 2 and 3 were combined in five simulation scenarios: 1: i+a, 2: i+b, 3: ii+a, 4: ii+b, 5: iii+a in order to be able to assess the effect of the choice of setting.

A subset of the meat inspection data (containing only farms delivering more than the chosen cut-off value during the sampling period) was constructed for each simulation scenario. From this subset, the required number of farms were sampled at random, using the "sample" function in R (R Core Team, 2017) without replacement. This was repeated 1,000 times for each simulation scenario. This resulted in 5,000 samples (5 simulation scenarios with 1,000 repetitions of the sampling).

8

The farms that did not match the target population were excluded. For each simulated sample the productivity and number of finishers was calculated and summarized.

To further evaluate the reproducibility of the sampling procedure, we evaluated whether the sampled farms in each of the 5,000 samples were randomly distributed in space. Spatial cluster analyses were conducted for the farms in the simulated samples using the same approach as described for the collected sample, where the sampled farms were compared with their respective set of non-sampled farms.

**Descriptive analyses**

For each of the 5,000 simulated samples summary statistics were calculated for the number of finishers and the productivity for the sampled farms. Hereafter, boxplots were made for the distribution of the first quantile, median, and third quantile within the five simulation scenarios. The summary statistic for the number of finishers and the productivity for the farms in the collected sample, farms in the target population, and farms in the study population were also calculated and the information was added to the graphs.

The results of the spatial cluster analyses for the farms in the simulated samples were mapped together with the results of the spatial cluster analyses for the farms in the collected sample.

**Results**

**Step one: planning the cross-sectional study**

A sample size of 800 farms was initially decided as a target. Calculating the number of farms to sample from abattoirs A-E resulted in the following distribution: A: 140, B: 300, C: 160, D: 120 and E: 80 farms.

**Step two: conducting the cross-sectional study**

In the cross-sectional study 800 faecal samples were collected. However, 94 farms were sampled two to four times. In four observations, it was not possible to match the delivery number with a CHR-number, and these were excluded. A further six farms were excluded – four because they did not have any finishers registered in the CHR, and two because they did not match the target population. This resulted in a collected sample of 681 farms.

9

**Step three: Evaluating the sampling procedure**

Evaluation of representativeness of the collected sample

For the cross-sectional study, the median number of finishers was 1,500 pigs for the farms in the collected sample and 800 pigs for the non-sampled farms in the target population, and 1,000 pigs for the non-sampled farms in the study population. The median productivity was 2,439 pigs for the farms in the collected sample and 939 pigs for the non-sampled farms in the target population, and 1,554 pigs for the non-sampled farms in the study population. The differences were significant for all variables and for both populations, with p<0.001.

One under-sampled area in the western part of Jutland and one over-sampled area in the northern part of Jutland were found for the farms in the collected sample (Fig. 2).

Evaluation of random spatial distribution

Repeating the sampling procedure through simulation resulted in samples that were consistent in terms of farm size measured as the number of finishers and the productivity (Fig. 3, boxplots) and comparable with the collected sample (Fig. 3, black lines), but the sampled farms were larger than those in the target population (Fig 3, grey solid lines) and study population (Fig. 3, grey dashed lines).

For the simulation scenarios, different numbers of over- and under-sampled areas were found, and the under-sampled areas were found in a higher percentage of iterations than the over-sampled areas (Table 1). The spatial clusters were often found in the same areas as for the collected sample, and sampling in Zealand seemed to lead to a randomly spatially distributed sample, except when the number of farms to sample per abattoir was recalculated (Fig. 2 – simulation scenario 5).

**Discussion**

This study presents a method for planning, conducting and evaluating a non-probability sampling procedure exemplified through a cross-sectional study of pig farms at the abattoir. The sampling procedure was efficient, simple and cost-effective, as samples were obtained from 681 farms in just 36 man-days, whereas sampling directly at farms would require 681 man-days, representing an 18-fold increase in costs.

Simple random sampling is the most correct way of obtaining a cross-sectional sample. However, it can be difficult to retain a completely random sample due to time, logistic, economic or legislative restrictions. Therefore, a non-probability sampling procedure is often adopted – either on purpose or

10

as an ad hoc solution to sampling difficulties. Dohoo et al. (2010) discuss that using a non-probability sampling procedure limits the external validity of the study. We suggest making a post-sampling evaluation to assess the potential problem with external validity. Our study highlights the importance of evaluating the characteristics for the farms in the resulting sample. We show how this can be achieved using two evaluation criteria: farm size and a spatial component. However, other sampling criteria can be evaluated in a similar manner, in order to assess if appropriate conclusions can be drawn based on the sampled data.

When evaluating the representativeness of the farms in the sample it is important to choose the appropriate evaluation parameters, so they match the objectives of the study. In this study, the representativeness of the farms in the samples was based on farm size. This is an important confounder for several farm-related parameters, such as the level of antimicrobial consumption per pig (Fertner et al., 2015; van der Fels-Klerx et al., 2011), the prevalence of diseases such as PRRSV (Evans et al., 2010), movement patterns (Nöremark et al., 2011) and biosecurity measures (Boklund et al., 2004; Laanen et al., 2013). We showed that the sampling procedure introduced sampling bias, since larger farms were over-represented in the collected sample from the cross-sectional study. Doubling the simulated sample size did not remove this selection bias (Fig. 3 – simulation scenario 3 and 4). It is not surprising that larger farms were over-represented in the sample, as they are expected to deliver pigs for slaughter more often and in larger batches, leading to a higher chance of being sampled at the abattoir. In Denmark, the number of small pig farms is decreasing (Christiansen, 2015). In addition, large farms have a higher impact on human health due to the high number of slaughtered pigs that originate from these farms. Therefore, large farms represent the most relevant segment of the Danish pig population. However, this is still a sampling bias, and results from non-probability sampling at abattoirs should be evaluated with care and the results interpreted accordingly. Our example study aims to explain possible associations between farm size, geographical location, antimicrobial consumption, and the level of antimicrobial resistance at a farm, thus selecting predominantly larger farms will restrict the validity of the conclusions to larger farms.

Although the sampling method aimed to cover the entire country, it appears that an uneven spatial distribution of the sampled farms was unavoidable. Both under- and over-sampled areas were found for the cross-sectional study, and the trend remained in the simulated samples (Fig. 2). The locations of the under- and over-sampled areas were not consistent, and it appears that there is a smaller chance of finding an over-sampled area than an under-sampled area (Table 1). For the farms

in the collected sample, an under-sampled area in the western part of Jutland was found both when comparing the locations of the sampled farms of all Danish pig farms with finishers (target population) and with the only farms that send pigs for slaughter at the sampling sites (study population) (Fig. 2). It is therefore unlikely that the choice of sampling sites caused the uneven spatial distribution of the sampled farms. However, we actually do not have an explanation for this phenomenon.

An over-sampled area was found when the sampled farms were compared to the target population, but not when they were compared to the study population. This might indicate that the farms in the northern part of Jutland primarily sent pigs to the two abattoirs located in that area that was also included in the study. One possible explanation for the occurrence of the under- and over-sampled areas is that the number of farms to sample per abattoir was calculated using historical meat inspection data. The first calculation might have influenced the results and been misleading, as the distribution of farms between the abattoirs might have changed in the meantime. Recalculating the number of farms to sample per abattoir did not remove the uneven spatial distribution of sampled farms in relation to non-sampled farms. However, the location of the clusters was changed and there was a reduction in the percentage of samples in which a significant spatial cluster was found (Table 1 and Fig. 3 – simulation scenario 5).

The sampling procedure was shown to be reproducible, as simulation of the sampling procedure resulted in simulated samples that were comparable (in terms of farm size) with the collected sample. In addition, there was little variation in farm size among the farms in the simulated samples (Fig. 3).

Using the non-probability sampling strategy inevitably meant sampling some farms more than once. In our collected sample, 119 observations were not used for further analysis. This is equivalent to a loss of 17% of the observations, meaning that if a predetermined number must be met, roughly extra 20% should be added to the sample size.

In conclusion, sampling at abattoirs in order to obtain a cross-sectional sample of finisher farms (with no prior knowledge of the farms sending pigs for slaughter) is a fast, inexpensive and logistically less demanding option than on-farm sampling. However, the sampling procedure will target larger farms, additional samples may be needed in order to reach the minimum required number of farms, and a potential bias may occur due to the farm-size effect. We have described a method for post-sample evaluation, thus making it possible to assess and adjust for potential

12

selection bias. In our example, we used farm size and spatial randomness as criteria, but it is possible to test any property of the sample using our methodology.

13

# References

Anonymous, 2016a. Order (DK) 1408/2016 Regarding registration of herds in CHR [in Danish], of November 29, 2016.

Anonymous, 2016b. The Danish Classification Inspection [in Danish] [WWW Document]. URL http://www.klassificeringskontrollen.dk/ (accessed 2.10.16).

Anonymous, 2016c. Order (DK) 1448/2016 Regarding marking, registration, and movement of cattle, pigs, sheep or goats [in Danish], of December 1, 2016.

van den Bogaard, A.E.J.M., London, N., Stobberingh, E.E., 2000. Antimicrobial resistance in pig faecal samples from the Netherlands (five abattoirs) and Sweden. J. Antimicrob. Chemother. 45, 663–671.

Boklund, A., Alban, L., Mortensen, S., Houe, H., 2004. Biosecurity in 116 Danish fattening swineherds: descriptive results and factor analysis. Prev. Vet. Med. 66, 49–62. doi:10.1016/j.prevetmed.2004.08.004

Christiansen, M.G., 2015. Structural development in Danish Pig production [in Danish].

DANMAP, 2016. DANMAP 2015 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark.

Dohoo, I., Martin, W., Stryhn, H., 2010. Veterinary Epidemiologic Research, 2nd ed. National library of Canada Cataloguing in Publication, Charlottetown.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Med. 4, e296. doi:10.1371/journal.pmed.0040296

Evans, C.M., Medley, G.F., Creasey, S.J., Green, L.E., 2010. A stochastic mathematical model of the within-herd transmission dynamics of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV): fade-out and persistence. Prev. Vet. Med. 93, 248–57. doi:10.1016/j.prevetmed.2009.11.001

Fertner, M., Sanchez, J., Boklund, A., Stryhn, H., Dupont, N., Toft, N., 2015. Persistent spatial clusters of prescribed antimicrobials among Danish pig farms--A register-based study. PLoS One 10, e0136834. doi:10.1371/journal.pone.0136834

van der Fels-Klerx, H.J., Puister-Jansen, L.F., van Asselt, E.D., Burgers, S.L.G.E., 2011. Farm factors associated with the use of antibiotics in pig production. J. Anim. Sci. 89, 1922–9. doi:10.2527/jas.2010-3046

Houe, H., Ersbøll, A.K., Toft, N., 2004. Introduction to Veterinary Epidemiology. Samfundslitteratur.

Kleinman, K., 2015. rsatscan: Tools, classes, and methods for interfacing with SaTScan stand-alone software.

Kulldorff, M., 1997. A spatial scan statistic. Commun. Stat. - Theory Methods 26, 1481–1496. doi:10.1080/03610929708831995

Kulldorff M. and Information Management Services Inc, 2015. SaTScanTM ver. 9.4.1: Software for the spatial and space-time scan statistics.

Laanen, M., Persoons, D., Ribbens, S., de Jong, E., Callens, B., Strubbe, M., Maes, D., Dewulf, J., 2013. Relationship between biosecurity and production/antimicrobial treatment characteristics in pig herds. Vet. J. 198, 508–12. doi:10.1016/j.tvjl.2013.08.029

Moodley, A., Nielsen, S.S., Guardabassi, L., 2011. Effects of tetracycline and zinc on selection of methicillin-resistant Staphylococcus aureus (MRSA) sequence type 398 in pigs. Vet. Microbiol. 152, 420–3. doi:10.1016/j.vetmic.2011.05.025

Nöremark, M., Håkansson, N., Lewerin, S.S., Lindberg, A., Jonsson, A., 2011. Network analysis of cattle and pig movements in Sweden: measures relevant for disease control and risk based surveillance. Prev. Vet. Med. 99, 78–90. doi:10.1016/j.prevetmed.2010.12.009

Pebesma, E., Bivand, R.S., 2005. Classes and methods for spatial data : the sp package. doi:10.1.1.160.9361

R Core Team, 2017. R: A Language and Environment for Statistical Computing.

Tukey, J.W., 1962. The Future of Data Analysis. Ann. Math. Stat. 33, 1–67. doi:10.1214/aoms/1177704711

Wegener, H.C., Hald, T., Lo Fo Wong, D., Madsen, M., Korsgaard, H., Bager, F., Gerner-Smidt, P., Mølbak, K., 2003. Salmonella control programs in Denmark. Emerg. Infect. Dis. 9, 774–80. doi:10.3201/eid0907.030024

15

**Figure labels**

Figure 1: Location of farms sending pigs to the five abattoirs (A to E) included as sampling sites. The maps are showing the location of the farms that have sent pigs to slaughter at the abattoirs included in the study. Black dot indicates the location of a pig farm.

Figure 2: Results of the spatial cluster analyses for the cross-sectional study and the simulated samples when compared to either the target population or the study population. The sampling procedure was simulated using five different sampling scenarios. These were different combinations of how many farms to sample per sampling site (i: using the same number of farms to sample per abattoir as in the cross-sectional study, ii: doubling the number of farms to sample compared to the cross-sectional study, iii: recalculating the number of farms to sample per abattoir by using meat inspection data from the sampling period) and how many pigs a farm should send to slaughter, that the farm would be sampled (a: 20 pigs sent for slaughter per day or b: abattoir-specific cut-off value). The five simulated samples were, 1: i+a, 2: i+b, 3: ii+a, 4: ii+b, 5; iii+a. The numbers on the left side of the figure refer to the simulation scenarios. Each simulated scenario was repeated for1000 iterations and for each iteration; the spatial cluster analysis was performed. The grey circles show the over- and under-sampled areas found from the spatial cluster analysis of the collected sample. The black circles show the locations of significant clusters (over- and under-sampled areas) found for the farms from the simulated samples. Each black circle represents the result from one iteration out of the 1000 iterations per simulated scenario.

Figure 3: Comparison of farm sizes between the collected sample and the simulated samples. The sampling procedure was simulated using five different sampling scenarios. These were different combinations of how many farms to sample per sampling site (i: using the same number of farms to sample per abattoir as in the cross-sectional study, ii: doubling the number of farms to sample compared to the cross-sectional study, iii: recalculating the number of farms to sample per abattoir by using meat inspection data from the sampling period) and how many pigs a farm should send to slaughter, that the farm would be sampled (a: 20 pigs sent for slaughter per day or b: abattoir-specific cut-off value). The five simulated samples were, 1: i+a, 2: i+b, 3: ii+a, 4: ii+b, 5; iii+a. Each simulated scenario was repeated in 1000 iterations and for each iteration summary statistics was calculated for the number of finishers and the productivity of the farms in the resulting sample. Per simulated sample, the distribution (boxplots) in the first quantile, median and third quantile for

16

the number of finishers registered in CHR (top) and estimated productivity (delivery of pigs to slaughter over 26 weeks) (bottom) is shown. The black lines refer to the farms in the collected sample, the grey solid lines to the farms in the target population, and the grey dashed lines to the farms in the study population.
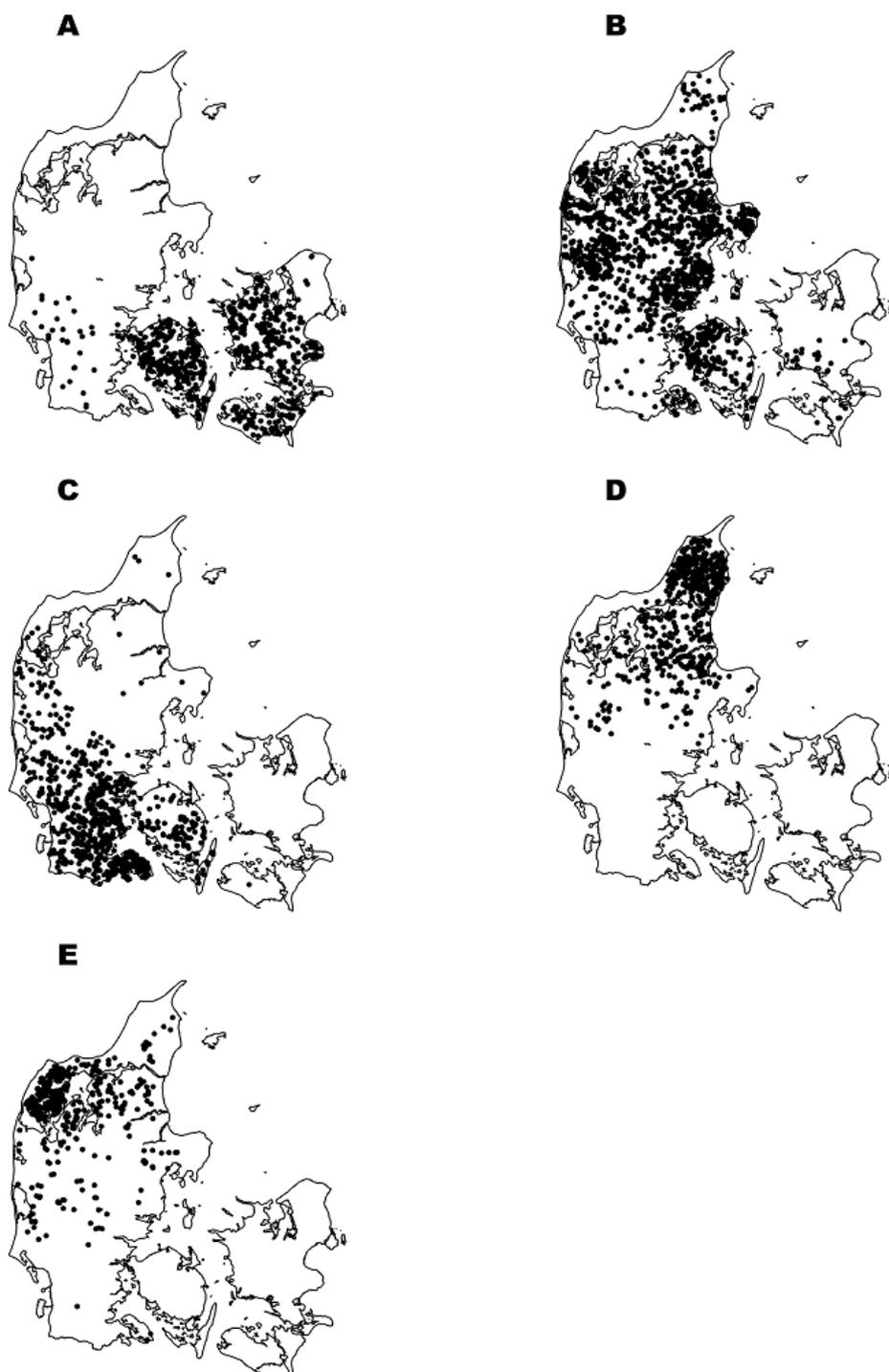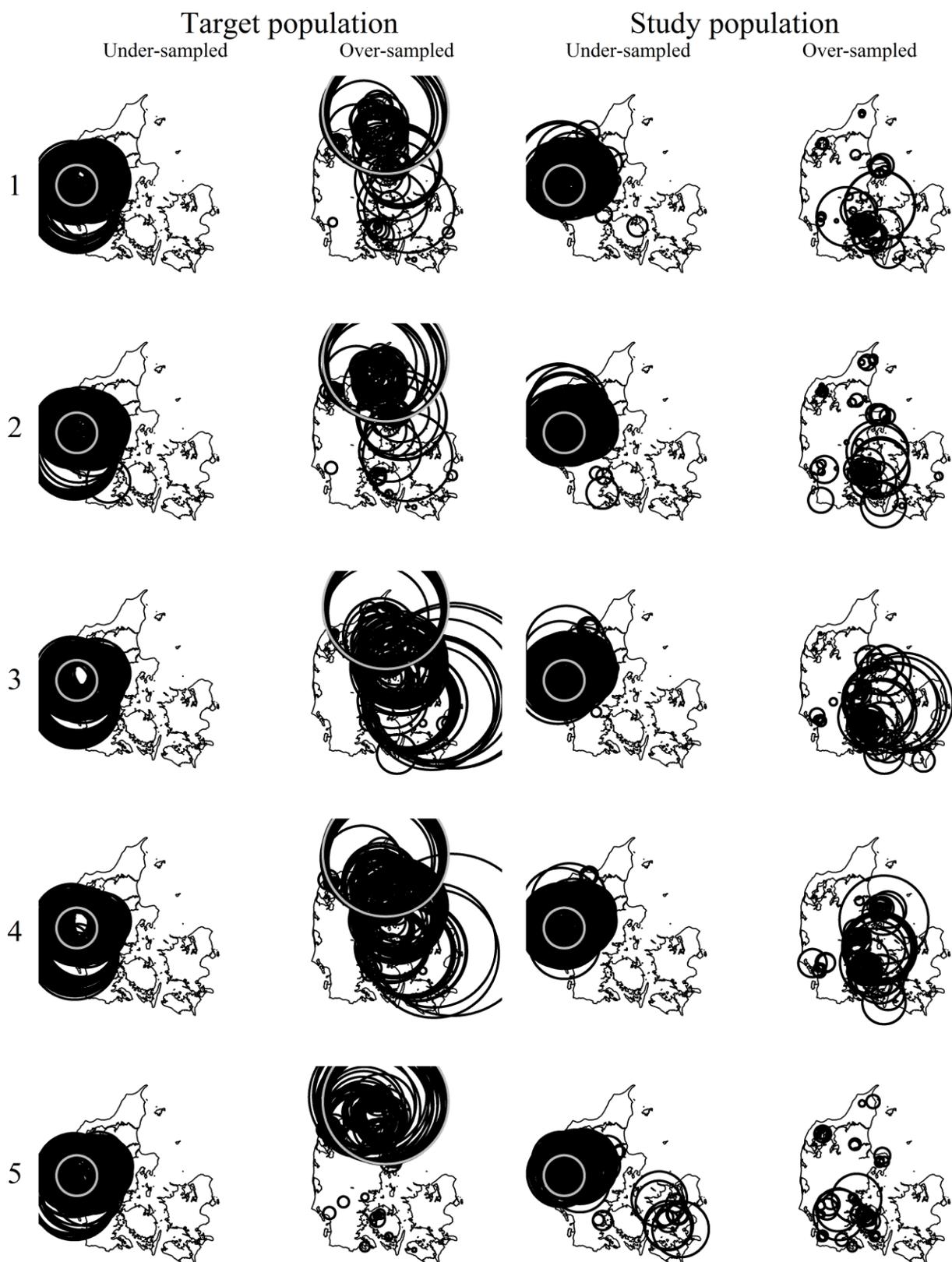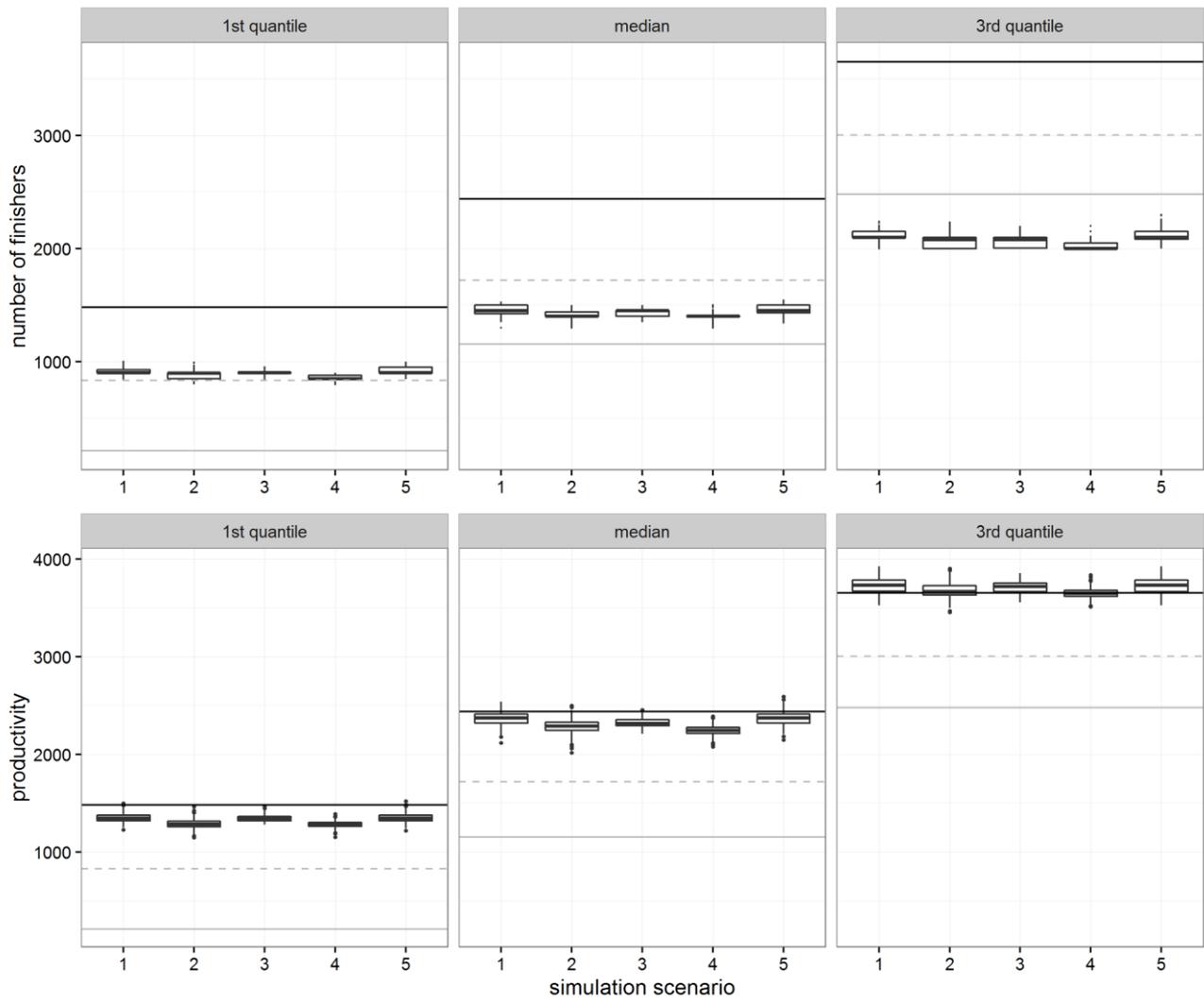
**Table 1: Results from the spatial cluster analyses**

| Simulation scenario | Target population | | | | Study population | | | |
|---|---|---|---|---|---|---|---|---|
| | Over-sampled area, %[1] | Under-sampled area, %[1] | Mean number of farms in cluster (+/- sd) | Mean radius (+/- sd), km | Over-sampled area, %[1] | Under-sampled area, %[1] | Mean number of farms in cluster (+/- sd) | Mean radius (+/- sd), km |
| 1 | 23.7 (1.6) | 94.2 (0.8) | 863 (+/- 475) | 47 (+/- 21) | 7.2 (0) | 32.9 (0.2) | 496 (+/- 288) | 42 (+/- 20) |
| 2 | 24.8 (2.4) | 95.5 (1.8) | 840 (+/- 464) | 46 (+/- 20) | 9.4 (0.2) | 35.6 (0.9) | 373 (+/- 293) | 40 (+/- 21) |
| 3 | 46.2 (3.8) | 99.8 (2.0) | 881 (+/- 446) | 60 (+/- 27) | 11.0 (0.1) | 41.2 (1.6) | 354 (+/- 272) | 42 (+/- 24) |
| 4 | 52.8 (6.2) | 100 (3.1) | 803 (+/- 466) | 50 (+/- 27) | 13.6 (0.5) | 51.0 (1.0) | 381 (+/- 262) | 44 (+/- 22) |
| 5 | 17.9 (0.7) | 89.5 (0.1) | 812 (+/- 490) | 44 (+/- 23) | 7.4 (0.4) | 31.9 (0.8) | 331 (+/- 275) | 37 (+/- 20) |

The sampling procedure was simulated using five different sampling scenarios. These were different combinations of how many farms to sample per sampling site (i: using the same number of farms to sample per abattoir as in the cross-sectional study, ii: doubling the number of farms to sample compared to the cross-sectional study, iii: recalculating the number of farms to sample per abattoir by using meat inspection data from the sampling period) and how many pigs a farm should send to slaughter, that the farm would be sampled (a: 20 pigs sent for slaughter per day or b: abattoir-specific cut-off value). The five simulated samples were, 1: i+a, 2: i+b, 3: ii+a, 4: ii+b, 5; iii+a. The numbers on the left refer to the simulation scenarios. Each simulated scenario was repeated for 1000 iterations and for each iteration, the spatial cluster analysis was performed.

[1]Over- and under-sampled areas: percentage of iterations where a significant spatial cluster was found out of 1000 iterations. Numbers in parentheses indicate the percentage of iterations where more than one spatial cluster was found.

21