Food and Agriculture Organization
of the United Nations

Collection and estimation of population size data for risk classification in DAD-IS
A sampling methodology

I. Introduction

The Domestic Animal Diversity Information System (DAD-IS) is a global communication tool supporting the development of strategies for animal genetic resources (AnGR) management, considered as the clearing house mechanism for AnGR data. It provides National Coordinators with a secure means to control the entry, updating and accessing of their national data. Throughout the existence of DAD-IS, population size data have been lacking for many breeds. This information is critical for guiding the national management of animal genetic resources (AnGR) and the calculation of SDG Indicator 2.5.2, for which FAO is custodian. The indicator is based on the breeds' risk of extinction which is calculated using the breeds' population sizes. As of September 2021, across the world, when excluding extinct breeds, 61 percent of local breeds are classified as of unknown risk status, 29 percent as at risk, and 11 percent as not at risk of extinction. The main reason for the data gap is lack of population size data on country level. National livestock censuses, if carried out at all, are expensive and thus not sufficiently frequent relative to the annual SDG reporting. Moreover, they usually do not provide information on breed level. Therefore, countries need cost efficient solutions allowing them to estimate regularly the population sizes of their AnGR.

The document presents a methodology and a simple tool that have been developed to facilitate the collection of information and estimation of population size, in a cost effective manner. This approach has been tested in two pilot projects involving several countries in Latin America and North Africa.

In a first section, the general principles of the approach are presented, in terms of requirements, sampling and statistical approaches. The second part focuses on practical considerations, for the implementation in the field. A third section presents an application that has been developed to facilitate the analysis of the data collected and an estimation of population sizes.

II. Principles of the approach

1. General considerations and prerequisites

In absence of an exhaustive registration system or large scale livestock census, different methods can be used to estimate livestock population sizes (FAO, 2011) based on sampling. Those methods can be divided into the following categories:

- Counting directly the number and type of animals based on a sample of landscapes, farms, or households visited. Direct counting can be very demanding in terms of time, cost, capacity requirements, and applicability depending on the situation on the field.
- Appraisals (questionnaire surveys, household or group interviews, key informants and farmers organizations/breed societies' interviews) are based on estimations provided by relevant stakeholders at various scales. Those methods are relatively quick to implement and less costly than censuses, yet the challenge with those methods lies with the difficulty to quantify, standardize and pool data.

The approach presented here is based on various assumptions and experiences. First, local appraisal approaches based on random or semi-random sampling should be a more cost-effective strategy to estimate breed population size, compared to an exhaustive or large scale livestock census.

For sampling, it needs to be considered that for a given species, the distribution of breeds is expected to be heterogeneous within the country, in relation to agroecological and/or sociocultural factors. Therefore, an effective sampling approach needs to consider this heterogeneity, while also taking into account the administrative stratification of the country for practical considerations related to the collection of data in the field.

The principle of the approach is to assess the repartition of the species among breeds in different agroecological or administrative strata within the country, following the principle of stratified sampling, to obtain reliable population estimates in a cost-effective manner. With adequate stratification, stratified sampling can indeed provide estimates with greater precision than a simple random sample of the same size, increasing therefore cost-effectiveness.

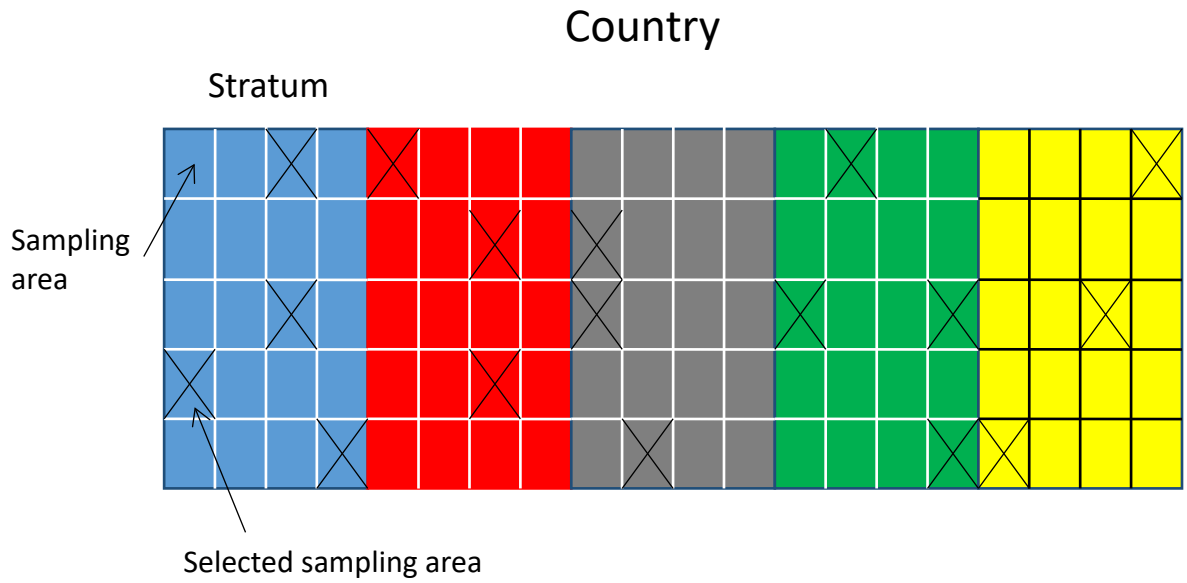
The estimation at stratum level can be based either on direct counts or appraisals, eventually considering a combination of methods. Independently from the methods selected, three major prerequisites have to be considered:

- First, relatively recent national livestock censuses or other reliable estimates at species level must be available on national and sub-national basis (i.e for strata, such as agroecological zones

or administrative regions), and at levels below strata (e.g. municipal level). Indeed, small-scale direct counts or appraisals are not expected to provide precise estimates of numbers at national level. However if the population size at species level is already known, those approaches may provide a good estimate for breed repartition, and therefore allow the estimation of the corresponding population sizes of breeds.

- Second, main breeds of interest should already have been characterized, with a consensus on how those breeds can be differentiated in the field (FAO, 2012).
- Third, whatever the methods chosen, there is a need to have institutional capacities and support at stratum level, through institutions like regional ministry offices, veterinary and extension services, or farmer associations.

If one of those conditions is not fulfilled the proposed sampling approach cannot be applied.



Source: Author's own elaboration.

Figure 1. Stratified sampling scheme over different strata within a country

Within each stratum, sampling areas are chosen in a random or semi-random way (Figure 1), thus being representative for the total stratum but also being easily accessible by the implementation partner on the ground. For each sampling area, breed population sizes will be derived based on information that is either already available from institutional partners (e.g. appraisal by local breeders' or farmers' associations, estimates provided by key informants, information gathered at slaughter houses, etc.), or on field work (e.g. direct counting, household interviews). In a next step those estimates from the sample sites are used to derive population sizes on strata level and further for the total country.

2. Steps to be considered

The implementation of the approach requires several consecutive steps:

Step 1 is dedicated to the planning of the strategy, i.e. how the data collection process is organised, including the choice of species to be covered, the stratification principle (agroecological zones or administrative units), the methods to be used for assessing the breed distribution in the stratum (direct counts or appraisals), the partners to be involved. The strategy needs to be discussed depending on budget, time, human resources available and constraints in the field like accessibility of certain areas (e.g. during raining season). This step involves meetings with national and local stakeholders to ensure their participation, and eventually training of field staff for the identification of breeds.

Step 2 deal with the collection of information on the ground in the different strata and sampling areas following the strategy and methods designed in step 1.

Step 3 focus on the estimation of breed population sizes at national level, on the basis of information collected in each sampling site.

During Step 4, estimations should be validated involving experts in the field of livestock and statistics. Depending on eventual discrepancies identified in step 3 and 4, complementary work may be required to consolidate the estimations. Lessons learned and adaptation necessary should be discussed and documented to improve the approach for future use.

Once validated, data should entered into DAD-IS by the National Coordinator.

3. Statistical estimation of breed population size

Stratified sampling is a sampling approach, which is considered to be of interest, because, in comparison to simple random sampling, the precision of the sampling is improved in relation to reduction of sampling error.

Supposing a country is subdivided into I strata. The total population size of a given species within the country is estimated to be N, with N_i the species population size for stratum i. Within stratum i, n_i animals are sampled, among which a proportion p_{ik} belongs to the breed k of interest. The formulas below (Table 1) show how to estimate the population size and confidence interval for breeds.

Table 1. Formulas used for the estimation of population size

Estimate	Formula
w_i: Weight of stratum *i*	$\frac{N_i}{N}$
$\overline{p_k}$: breed *k* demographic proportion estimated at country level	$\sum_{i=1}^{I} p_{ik} \times w_i$
$\overline{n_k}$: population size estimate for breed k	$\sum_{i=1}^{I} p_{ik} \times N_i$
s_{ik} : Standard deviation of the proportion estimate of breed *k* within the stratum *i*	$(p_{ik} \times (1 - p_{ik}))^{\frac{1}{2}}$
S_{ek}: standard error of proportion estimate $\overline{p_k}$	$\sum_{i=1}^{I} w_i^2 \times (\frac{N_i - n_i}{N_i}) \times (\frac{s_{ik}^2}{n_i})^{\frac{1}{2}}$
Upper limit 95% of confidence interval for population size $\overline{n_k}$	($\overline{p_k}$ + 1.96 x S_{ek}) x *N*
Lower limit 95% of confidence interval for population size $\overline{n_k}$	($\overline{p_k}$ - 1.96 x S_{ek}) x *N*

Source: Author's own elaboration.

III. Implementation in practice

This design and implementation of the strategy require preliminary discussions with experts in livestock breeds in the field, as well as with experts in statistics, given the impact those decisions will have on population size estimation and the precision of those estimations. Data need to be collected from a representative sample of the designated stratum to allow adequate estimation of the breed distribution.

1. Choice of the species

The approach can be applied on selected species and aims at estimating the population size of all breeds present in the country for the selected species. The choice of species will depend on the demographic and socio-economic importance of the species, as well as on capacities for monitoring and management. For the choice of species to be covered, the total number of animals belonging to the species, specific cultural importance, existing governance (e.g. existence of farmers and breed associations) and capacities (field expert that can be mobilized) can play a role.

The choice of species determines the stakeholders to be involved in the fieldwork.

2. Sampling objectives in theory and practice

Determining the sampling size, in terms of number of animals, depends on the objectives of the study, as well as on considerations of statistical (precision and structure of the environment) and practical nature (environmental constraints, availability of funds, time, and human resources…) (Thrusfield, 1995).

Table 2 shows the theoretical confidence interval expected for the estimation of a breed frequency according to the sample size, in an ideal case (large population under random sampling). This interval is expected to increase (and thus the precision is expected to decrease) with the frequency of the breed getting close to 50 percent. If for a given species, there are only a few breeds with each having relative high frequencies (>10 percent), sampling a few hundred animals can provide a rough idea of their importance within the species in theory. In practice, the precision of the estimate is expected to be reduced due to non-random nature of field sampling (non-randomness of sampling) and potential difficulties to identify breeds correctly. Therefore, it is advised to sample a minimum of 5 000 to 10 000 animals within the species if the objective is to monitor breed populations on a regular basis. Yet even with 10 000 animals sampled, the precision for a breed with an expected frequency of only 1 percent will be close to 0.2 percent i.e. 20 percent of its expected frequency. Therefore the sampling may need to be adapted, on *a priori* knowledge, to better consider small local breeds (see following sections).

Table 2. Sample size required to estimate the frequency of a breed with the desired fixed confidence limits (with a level of confidence of 95%)

	Desired absolute precision		
Expected frequency	10%	5%	1%
10%	35	138	3457
20%	61	246	6157
30%	81	323	8067
40%	92	369	9220
50%	96	384	9604
60%	92	369	9220
70%	81	323	8067
80%	61	246	6157
90%	35	138	3457

Source: Thrusfield, M. 2005. Veterinary epidemiology. 2nd Edition, Blackwell Science, Oxford, 178-198.

3. Choice of strata

Ideally, stratification should be designed to match optimally the inhomogeneous breed distribution across the country. As indigenous breeds are often described to be adapted to specific environmental conditions, breed distribution patterns might match to some extent agroecological areas. Therefore, the division of the country in agroecological zones can be the optimal stratification strategy. However, administrative regions within the country may also constitute relevant stratification units, as those units might match with socio-environmental differences, while also corresponding to administrative and governance units. If the sampling implementation strategy relies a lot on regional institutions, it can make sense to use administrative units rather that agroecological ones. Depending on the circumstances, other factors such as production systems or spatial distribution of major ethnic groups may eventually be considered for the definition of the stratification (FAO, 2011).

Table 3. Repartition of the sampling efforts based of two theoretical scenarios

	Total species population size	Sampling size scenario A (no assumption)	Sampling size scenario B (stratum 3 heterogeneous with numerous small breeds)
Stratum 1	600 000	6000	4500
Stratum 2	200 000	2000	1500
Stratum 3	200 000	2000	4000

Source: Author's own elaboration.

In theory, the sampling efforts should be proportional to the demographic importance of a given stratum. Specific importance may however be given to sampling in a certain stratum if the area is expected to be particularly heterogeneous or rich in indigenous breeds of small population size. An example is provided in table 3, where it was decided to sample a total of 10 000 animals. Without specific assumptions, sampling efforts will be proportional for each stratum, but more emphasis may be given with a priori information that a specific stratum is known to be heterogenous and hosts numerous small breeds compared to other strata (see section on sampling frame).

4. Choice of the estimation method

Estimation methods have been described extensively in the FAO guidelines “Surveying and Monitoring Animal Genetic Resources” (FAO, 2011). The development of an effective strategy aiming at collecting breed population data in a regular manner requires decisions on methods to be used and if and how they can be best combined. While appraisals are usually quicker, less costly, and less time consuming, they are less precise, as numbers are extrapolated from a third party who may have for instance a different view on how breeds are defined. Partners making appraisals should also provide the scale of the appraisal, i.e. the number of animals from which the breed repartition is made. Among appraisal methods, questionnaire surveys may allow to increase the number of partners involved and thus the

sample size, in a formalized manner, however they are also less flexible and maybe sometimes prone to bias due to logistic and sociologic issues (degree of literacy of farmers…).

Depending on the local context, different methods may be chosen and eventually mixed, as for a given country, a method can be relevant for a stratum and not for another one. Implementing locally two different methods in the same area can also be a way to check, once in a while, the reliability of the estimate (see Box 1). In practice, in both South American and North Africa pilot projects, the strategy relied mostly on a mix of direct counts, household surveys and key informant appraisals. The validation meeting (see above Step 4) is designed to consolidate the estimates made in the field through complementary appraisal approaches.

Box 1. Comparing sampling results with other estimates in Colombia

During the population assessment in Colombia, it was possible to compare, for a small number of breeds, the results provided by the sampling approach with estimates from breeder associations and conservation institutions. In two out of the four breeds, the estimates from expert fall within the 95% range from the sampling estimate, while for one breed (cattle Casanareño), there was a quite large difference between both estimates, thus requiring some further validation.

Table 4. Population estimate based on sampling approach, compared with other estimates

Species	Breed	Sampling estimate (95% Upper and Lower limit)	Other estimates
Cattle	Casanareño	3100 (2873 - 3327)	977
Cattle	Romosinuano	3197 (2911 - 3483)	3709
Cattle	Sanmartinero	2360 (2162 - 2558)	2547
Goat	Criolla	9673 (8976 - 10370)	9440

Source: Author's own elaboration.

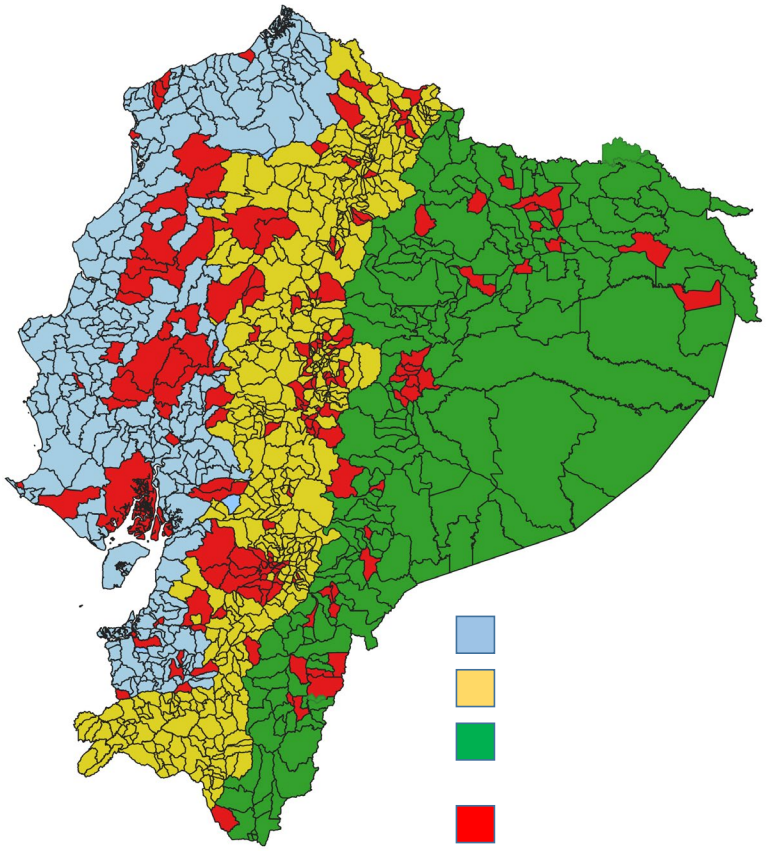
5. Practical considerations on the sampling frame

Depending on agro-ecological features, management systems and economic and socio-cultural factors, while considering also estimation methods used, sample sites may either be constituted of herds, farms households, villages or even small administrative units. Of course, direct count methods are more difficult to apply for large sample sites. The sampling frame within a given stratum may consider either one, two or even three layers, for instance a first layer could be a municipality, with a second layer corresponding to farms within the municipality, as sampling areas.

As previously underlined, the sampling frame needs to be defined according to the local context and capacities. In theory, the selection of sampling areas within a given stratum is expected to be random. In practice however, it is often difficult to select those areas in a completely random manner, as the fieldwork has also to match ground constraints, such as accessibility and availability of field implementation partners. Sampling areas may eventually be chosen in a semi-random way, on the basis of a trade-off between representativeness for the stratum, its more or less homogeneous structure, relevance and accessibility of sampling sites (see Box 2). Quantitatively, a homogenous stratum of limited size will require a smaller number of sample sites than a more heterogenous one of large size. Based on preliminary discussion with local experts, sampling efforts (i.e. more sites and animal sampled) may be directed to consider heterogeneity in breed distribution within the stratum. Also the different production systems (e.g. communal, pastoral, commercial…) within the strata maybe needed to be considered on the sampling (Scholtz et al., 2008). Further layers of stratifications may eventually be considered depending on the means, time and capacities, as well as the heterogeneity of the different strata (Rowland et al., 2003, Ayalew et al., 2004).

Box 2. Sampling stratification in Ecuador

For the assessment of breed populations in Ecuador, the sampling frame was based on three regions linked to different agroecological areas, namely Costa (pacific coastal plain), Sierra (highland and mountainous area), and Oriente (from eastern Andean slopes to Amazonian rainforest). Parishes were considered as sampling areas with semi-random sampling as illustrated by Figure 2. For instance, sampling has been limited in the eastern part of Oriente region, considering the extremely low livestock density in this area, corresponding to Yasuni national park and some indigenous territories.



Source: United Nations Geospatial. 2020. Map of the World. United Nations. Cited 22 August 2022. www.un.org/geospatial/file/3420/download?token=TUP4yDmF modified by authors.

The boundaries and names shown and the designations used on these map do not imply the expression of any opinion whatsoever on the part of FAO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers and boundaries. Dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Figure 2. Parishes sampled for the assessment of breed populations in three agroecological regions of the Ecuador

Determining the sample sizes within sample sites is a complex task, and has to be balanced considering expected heterogeneity between (difference in breed repartition among villages) and within sampling sites (e.g. differences in breeds owned among households interviewed within a given village). Increasing the number of sampling sites is expected to reduce the variance related to the first component, while increasing the number of animals counted within sampling sites will reduce the second.

Table 5. Sampling frame and species coverage in three South America countries

Country	Strata	Sampling units	Animals sampled (population coverage)
Colombia	5 agroecological areas	25 Municipios (300 surveys)	5908 chickens (0.004%) 43401 cattle (0.19%) 430 buffalo (0.14%) 3921 goats (0.34%) 2834 horses (0.2%) 3311 sheep (0.23%) 1194 pigs (0.02%)
Ecuador	3 agroecological areas	113 Parroquias (1100 surveys)	197442 chickens (0.34%) 108347 cattle (2.51%) 74764 guinea pigs (1.48%) 777 goats (1.96%) 6000 horses (2.07%) 3151 sheep (0.81%) 43 camelids (0.11%)
Panama	5 agroecological areas	43 Corregimientos (7476 surveys)	21828 cattle (1.26%) 469 goats (5.61%) 553 sheep (2.96%) 7648 pigs (2.37%)

Source: Author's own elaboration.

Table 5 provides an illustration of the sampling frames applied in Panama, Colombia and Ecuador. The number of species and agroecological areas differed according to country. For the three countries, small administrative units (called Municipios, Parroquias and Corregimientos, respectively) were defined as sampling units, with a coverage of 2 percent, 8 percent and 6 percent of the total number of sampling units in Colombia, Ecuador, and Panama, respectively. Field operators were tasked to visit each farm within those sampling sites, using direct interviews and field questionnaires, the population coverage being rarely exhaustive in practice. Additionally, consultations were carried out with veterinarians and local extension agents, as well as with breeders' associations, completing the field information with these records. This resulted in a number of surveys larger than the number of sampling units. Depending on species, the total number of animals sampled ranged from 43 (camelids in Ecuador) to 197442 (chicken in Ecuador). As a consequence of the limited sampling size for some species, the process did

not result in population estimates for all breeds and species. For instance Ecuador actually reported population estimates in DAD-IS for 36 breeds of only 4 of the 8 species sampled, mostly cattle.

6. Specific cases

The approach is flexible allowing adaptation of methods or sampling strategies depending on the context. For instance, small local breeds known to be concentrated into a limited geographic area may need to be treated separately, to avoid that those breeds are overlooked in the general sampling process. In the Latin American pilot project, such populations of small size found in isolated territories have been reported, with an estimation based on direct counting in the field, followed by direct inclusion of those figures in the total national estimation.

Stakeholder groups (like farmers' or breeders' organizations responsible for a specific breed or stratum), may be able to provide precise numbers for the population estimation in their area of expertise (breed or location). Further, some breeds may have a part of their population very well monitored, e.g. through nucleus farms, or registration schemes for commercial systems. The corresponding subpopulation data can be directly used and added to the estimate as an independent stratum with exhaustive knowledge.

III. Tool to estimate population sizes for breeds

To facilitate the estimation of breed population sizes, a computer application has been developed by Red Conbiand to support the storage of information and estimation of population sizes in the pilot project in Latin America. This Access database is Windows compatible and available in three languages.

The following types of data need to be entered:

- Definition of Strata/Agroecological regions of the country, and areas where sampling takes place.
- Animal species and breeds of study.
- Censuses of each species on country level and on strata/agro-ecological regions' level
- Number of animals per breed and sampling area.

Sampling must be done in a stratified manner and the application bases the calculation on:

- The estimation of the proportions of each breed with respect to the species: a) in the strata, b) in the country.
- Calculates the confidence interval of the estimated population size for both, the region and in the country.

The software is available on request by writing to DAD-IS@fao.org.

Bibliography

Ayalew, W., Dorland, A. V., & Rowlands, G. J. 2004. Design, execution and analysis of the livestock breed survey in Oromiya Regional State, Ethiopia.

FAO. 2011. Surveying and monitoring of animal genetic resources. FAO Animal Production and Health Guidelines. No. 7. Rome.

FAO. 2012. Phenotypic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. No. 11 Rome.

Rowlands, J., Nagda, S., Rege, E., Mhlanga, F., Dzama, K., Gandiya, F., Hamudikwanda, H., Makuza, S., Moyo, S., Matika, O., Nangomasha, E. & Sikosana, J. 2003. A report to FAO on the design, execution and analysis of livestock breed surveys – a case study in Zimbabwe. Nairobi, International Livestock Research Institute (available at http://agtr.ilri.cgiar.org/sites/all/files/library/docs/FAOAndILRIZimbabewReport.pdf).

Scholtz, M. M., Bester, J., Mamabolo, J. M., & Ramsay, K. A. 2008. Results of the national cattle survey undertaken in South Africa, with emphasis on beef. Appl. Anim. Husb. Rural Dev, 1, 1-9.

Thrusfield, M. 2005. Veterinary epidemiology. 2nd Edition, Blackwell Science, Oxford, 178-198.