



Sample Size Estimation in Veterinary Epidemiologic Research

Mark A. Stevenson*

Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia

In the design of intervention and observational epidemiological studies sample size calculations are used to provide estimates of the minimum number of observations that need to be made to ensure that the stated objectives of a study are met. Justification of the number of subjects enrolled into a study and details of the assumptions and methodologies used to derive sample size estimates are now a mandatory component of grant application processes by funding agencies. Studies with insufficient numbers of study subjects run the risk of failing to identify differences among treatment or exposure groups when differences do, in fact, exist. Selection of a number of study subjects greater than that actually required results in a wastage of time and resources. In contrast to human epidemiological research, individual study subjects in a veterinary setting are almost always aggregated into hierarchical groups and, for this reason, sample size estimates calculated using formulae that assume data independence are not appropriate. This paper provides an overview of the reasons researchers might need to calculate an appropriate sample size in veterinary epidemiology and a summary of sample size calculation methods. Two approaches are presented for dealing with lack of data independence when calculating sample sizes: (1) inflation of crude sample size estimates using a design effect; and (2) simulation-based methods. The advantage of simulation methods is that appropriate sample sizes can be estimated for complex study designs for which formula-based methods are not available. A description of the methodological approach for simulation is described and a worked example provided.

OPEN ACCESS

Edited by:

Heinzpeter Schwemmer,
Federal Food Safety and Veterinary
Office (FSVO), Switzerland

Reviewed by:

Nikolaos Pandis,
University of Bern, Switzerland
Ane Nødtvedt,
Norwegian University of Life
Sciences, Norway

*Correspondence:

Mark A. Stevenson
mark.stevenson1@unimelb.edu.au

Specialty section:

This article was submitted to
Veterinary Epidemiology and
Economics,
a section of the journal
Frontiers in Veterinary Science

Received: 01 March 2020

Accepted: 30 November 2020

Published: 17 February 2021

Citation:

Stevenson MA (2021) Sample Size
Estimation in Veterinary Epidemiologic
Research. *Front. Vet. Sci.* 7:539573.
doi: 10.3389/fvets.2020.539573

Keywords: sampling, epidemiology, multilevel—hierarchical clustering, veterinary science, biostatistics

INTRODUCTION

In the design of intervention and observational epidemiological studies sample size calculations are used to provide estimates of the minimum number of observations that need to be made to ensure that the stated objectives of a study are met (1, 2). Peer reviewed journals require investigators to provide justification of the number of subjects enrolled into a study and details of the assumptions and methodologies used to derive sample size estimates are now a mandatory component of grant application processes (3). Studies lacking in justification of sample size run the risk of failing to identify differences among treatment or exposure groups if a difference in those groups actually exist (4). Selection of a number of study subjects greater than that actually required results in a wastage of time and resources (2).

Methods for sample size estimation vary depending on the type of study being carried out i.e., observational (non-experimental) or interventional (experimental). Formula-based approaches for sample size estimation are often preferred by investigators because: (1) they

are relatively quick and simple to implement; (2) their widespread use makes peer review challenge less likely; and (3) the ability to use standard formulae goes hand in hand with “standard” study designs (i.e., randomized clinical trials, cross-sectional studies, case-control studies or cohort studies). Use of a standard study design implies the use of established approaches for data collection and analysis, again reducing the likelihood of challenge during peer review. In veterinary epidemiology the aggregation of animals into often several levels of hierarchy (e.g., cows within pens, pens within herds, herds within farms, and farms within regions) complicates sample size calculations due to lack of data independence arising from study subjects being aggregated into groups (e.g., pens, herds, farms, and regions). While modifications to standard sample size formulae are available, their flexibility to handle the range of real-world data situations is often limited.

The aim of this paper is to provide an overview of sample size estimation methods and their usage in applied veterinary epidemiological research. The structure of the paper is as follows. In the first section an overview of formula-based approaches for sample size estimation in epidemiological research is provided. In the second section, formula-based approaches for calculation of appropriate samples sizes for clustered data are presented. In the third and final section simulation-based approaches are presented as a means for estimating an appropriate sample size for hierarchical study designs for which formula-based methods are not available. Examples are provided throughout the paper to

illustrate and support the concepts discussed. The supplementary material contains code allowing readers to reproduce the results presented in each of the examples using functions available in the contributed epiR package (5) in R (6).

FORMULA-BASED APPROACHES FOR SAMPLE SIZE ESTIMATION

In veterinary epidemiology sample size calculations are used during the design phase of a study to allow investigators to: (1) estimate a population parameter (e.g., the prevalence of disease); (2) test a hypothesis in an observational setting (using, for example, one of the three main observational study designs: cross-sectional, cohort or case-control); (3) test a hypothesis in an intervention setting (using a randomized clinical trial); and (4) achieve a specified level of confidence that an event will be detected if it is present at a specified design prevalence.

Sample Size Calculations to Estimate a Population Parameter

A summary of formula-based methods for estimation of a population parameter, all assuming data independence, is provided in **Tables 1, 2**. Methods are defined for continuous and binary outcomes with different calculation methods dependent on the proposed sampling design: simple random, stratified

TABLE 1 | Information required to estimate a sample size for each of the common sampling designs, binary or continuous population parameters.

Outcome variable	Sampling design	Arguments	References
Continuous	Simple random	Total number of individual listing units in the population, the relative variance of the continuous variable to be estimated (i.e., the variance divided by the mean squared).	(7) pp. 74, Equation 3.14
Continuous	Stratified random	Total number of individual listing units in each strata, the expected means of the continuous variable to be estimated for each strata, the expected variances of the continuous variable to be estimated for each strata.	(7) pp. 176, Equation 6.25
Continuous	One stage cluster	Total number of clusters in the population, the population mean of the continuous variable to be estimated, the population variance of the continuous variable to be estimated.	(7) pp. 255, Box 9.4
Continuous	Two stage cluster	Number of individual listing units to be sampled from each cluster, the total number of clusters in the population and the number of individual listing units in each cluster, the mean of the continuous variable to be estimated at the first and second stage of sampling, the variance of the continuous variable to be estimated at the first and second stage of sampling.	(7) pp. 289, Equation 10.6
Binary	Simple random sampling	Total number of individual listing units in the population, the expected proportion of individual listing units with the outcome of interest.	(7) pp. 74, Equation 3.16
Binary	Stratified random	Total number of individual listing units in each strata, the expected proportion of individual listing units with the outcome of interest for each strata.	(7) pp. 176, Equation 6.23
Binary	One stage cluster	Total number of clusters in the population, the mean of the proportion of individual listing units in each cluster with the outcome of interest, the variance of the proportion of individual listing units in each cluster with the outcome of interest.	(7) pp. 255 Box 9.4
Binary	Two stage cluster	Number of individual listing units to be sampled from each cluster, the total number of clusters in the population and the number of individual listing units within each cluster, the mean of the denominator variable used to calculate the unknown population proportion at the first and second stage of sampling, the variance of the denominator variable used to calculate the unknown population proportion at the first and second stage of sampling, the variance of the numerator variable used to calculate the unknown population proportion at the first and second stage of sampling, the covariance of the unknown population proportion at the first and second stage of sampling.	(7) pp. 289, Equation 10.7

TABLE 2 | Formulae to estimate a sample size for each of the common sampling designs, binary or continuous population parameters.

Outcome variable	Sampling design	Formula	Arguments
Continuous	Simple random	$n \geq \frac{z_{1-(\alpha/2)}^2 N V_x^2}{z_{1-(\alpha/2)}^2 V_x^2 + (N - 1) \epsilon_r^2}$	<p>n = the number of subjects in the sample. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. N = the population size. V_x = the relative variance (the variance divided by the mean squared). ϵ_r = the relative error.</p>
Continuous	Stratified random	$n \geq \frac{z_{1-(\alpha/2)}^2 \times \frac{N}{1+\gamma} \times V_x^2}{N\epsilon_r^2 + z_{1-(\alpha/2)}^2 \times \frac{V_x^2}{1+\gamma}}$	<p>n = the number of subjects in the sample. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. N = the population size. γ = between strata variance σ_{Dx}^2 divided by the within strata variance. σ_{wx}^2. V_x^2 = the relative variance (the variance divided by the mean squared). ϵ_r = the relative error.</p>
Continuous	One stage cluster	$m = \frac{z_{1-(\alpha/2)}^2 M V_{1x}^2}{z_{1-(\alpha/2)}^2 V_{1x}^2 + (M - 1) \epsilon_r^2}$ $V_{1x}^2 = \frac{\sigma_{1x}^2}{\bar{X}^2}$ $\sigma_{1x}^2 = \frac{\sum_{i=1}^M (X_i - \bar{X})(Y_i - \bar{Y})}{M}$	<p>m = the number of clusters in the sample $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. M = the number of clusters in the population. ϵ_r = the relative error. σ_{1x}^2 = the first stage variance components. \bar{X} = mean level of X per cluster. X_i = level of the ith value of characteristic X.</p>
Continuous	Two stage cluster	$m = \frac{\left(\frac{\sigma_{1x}^2}{\bar{X}^2} \right) \times \left(\frac{M}{M-1} \right) + \left(\frac{1}{\bar{n}} \right) \times \left(\frac{\sigma_{2x}^2}{\bar{X}^2} \right) \times \left(\frac{\bar{N} - \bar{n}}{\bar{N} - 1} \right)}{\frac{\epsilon_r^2}{z_{1-(\alpha/2)}^2} + \frac{\sigma_{1x}^2}{\bar{X}^2(M-1)}}$	<p>m = the number of clusters in the sample. σ_{1x}^2 = the first stage variance components. \bar{X} = mean level of X per cluster. M = the number of clusters in the population. \bar{n} = the number of listing units to be sampled from each cluster. σ_{2x}^2 = the second stage variance components. \bar{N} = the number of listing units in each cluster. ϵ_r = the relative error. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence.</p>
Binary	Simple random sampling	$n \geq \frac{z_{1-(\alpha/2)}^2 N P_y (1 - P_y)}{[(N - 1) \epsilon_r^2 P_y^2] + z_{1-(\alpha/2)}^2 P_y (1 - P_y)}$	<p>n = the number of subjects in the sample. N = the population size P_y = the estimated population prevalence. ϵ_r = the relative error. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence.</p>
Binary	Stratified random	$n \geq \frac{\left(\frac{z_{1-(\alpha/2)}^2}{N^2} \right) \sum_{h=1}^L \frac{N_h^2 P_{hy} (1 - P_{hy})}{\pi_h P_y^2}}{\epsilon_r^2 + \left(\frac{z_{1-(\alpha/2)}^2}{N^2} \right) \left(\sum_{h=1}^L \frac{N_h P_{hy} (1 - P_{hy})}{P_y^2} \right)}$ $\pi_h = \frac{n_h}{n}$	<p>n = the number of subjects in the sample $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. N = the population size. L = the number of strata. N_h = the population size in the hth strata. P_{hy} = the estimated population prevalence in the hth strata. P_y = the estimated population prevalence. ϵ_r = the relative error. π_h = the fraction of samples allocated to strata h (decided in advance).</p>
Binary	One stage cluster	<p>When the number of listing units to be sampled per cluster is the same: $D = 1 + (b - 1)\rho$ When the number of listing units to be sampled per cluster varies: $D = 1 + \{ (CV^2 + 1) \bar{b} - 1 \} \rho$ $n_c \geq \frac{z_{1-(\alpha/2)}^2 P_y (1 - P_y) D}{(P_y \epsilon_r)^2 \bar{b}}$</p>	<p>D = the design effect. b = the number of listing units to be sampled from each cluster. ρ = the intracluster correlation coefficient. CV = the coefficient of variation of the number of listing units to be sampled from each cluster. \bar{b} = the average number of listing units to be sampled from each cluster. n_c = the number of primary sampling units (clusters) to be sampled. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. P_y = the estimated population prevalence ϵ_r = the relative error.</p>
Binary	Two stage cluster	$m = \frac{\left(\frac{\sigma_{1R}^2}{\bar{X}^2} \right) \times \left(\frac{M}{M-1} \right) + \left(\frac{1}{\bar{n}} \right) \times \left(\frac{\sigma_{2R}^2}{\bar{X}^2} \right) \times \left(\frac{\bar{N} - \bar{n}}{\bar{N} - 1} \right)}{\frac{\epsilon_r^2}{z_{1-(\alpha/2)}^2} + \frac{\sigma_{1R}^2}{\bar{X}^2(M-1)}}$	<p>m = the number of clusters in the sample. σ_{1R}^2 = the first stage variance components. \bar{X} = the mean level of characteristic X per listing unit. M = the number of clusters in the population. \bar{n} = the number of listing units to be sampled from each cluster. σ_{2R}^2 = the first stage variance components. \bar{X} = the mean level of characteristic X per cluster. \bar{N} = the average number of listing units per cluster in the population. ϵ_r = the relative error. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence.</p>

BOX 1 | The expected seroprevalence of brucellosis in a population of cattle is thought to be in the order of 15%. How many cattle need to be sampled and tested to be 95% certain that our seroprevalence estimate is within 20% (i.e., $0.20 \times 0.15 = 0.03$, 3%) of the true population value, assuming use of a test with perfect sensitivity and specificity? This formula requires the population size to be specified so we set N to a large number, 1,000,000:

$$n \geq \frac{z_{1-(\alpha/2)}^2 N P_y (1 - P_y)}{[(N - 1) \epsilon_r^2 P_y^2] + z_{1-(\alpha/2)}^2 P_y (1 - P_y)}$$

$$n \geq \frac{1.96^2 \times 1,000,000 \times 0.15 (1 - 0.15)}{[(1,000,000 - 1) 0.20^2 0.15^2] + 1.96^2 \times 0.15 \times (1 - 0.15)}$$

$$n \geq \frac{489,804}{900.489}$$

$$n \geq 545$$

To be 95% confident that our estimate of brucellosis seroprevalence is within 20% of the true population value (i.e., a relative error of 0.20) 545 cattle should be sampled.

TABLE 3 | Information required to estimate a sample size for each of the common observational epidemiological study designs.

Study design	Arguments	References
Cross-sectional	The expected prevalence of the outcome among the exposed, the expected prevalence of the outcome among in the unexposed, the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects, sided test.	(9) pp. 313, Equation 8.14
Case-control	The expected odds ratio, the prevalence of exposure among controls, the required study power, the ratio of the number of control subjects to the number of case subjects, sided test.	(10)
Cohort, count data	The expected outcome incidence risk among the exposed, the expected outcome incidence risk among the unexposed, the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects, sided test.	(9) pp. 313, Equation 8.14
Cohort, time at risk	The expected outcome incidence rate among the exposed, the expected outcome incidence rate among the unexposed, the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects, sided test.	(11)

random, one-stage cluster and two-stage cluster designs. For continuous outcomes the analyst needs to provide an estimate of the mean of the outcome of interest and its expected variability. For binary outcomes only an estimate of the expected population proportion is required, given the variance of a proportion P equals $P \times [1 - P]$ (8). In addition to specifying the required level of confidence in the population parameter estimate (usually 95%) one needs to specify the desired maximum tolerable error. The maximum tolerable error is the difference between the true population parameter and the estimate of the true population parameter derived from sampling. In each of the formula-based approaches listed in **Tables 1, 2** tolerable error is expressed in relative (as opposed to absolute) terms. If one assumes that the true population prevalence of disease is 0.40 and a desired relative tolerable error of 0.10 with 95% confidence is required, this means the calculation will return the required number of subjects to be 95% certain that the prevalence estimate from the study will be anywhere between $0.40 \pm (0.10 \times 0.40)$ that is, from 0.36 to 0.44. Some sample size formulae and/or software packages require maximum tolerable error to be expressed in absolute terms (that is, 0.04 for the example cited above). Analysts should take care to ensure that there is no ambiguity around the input format for tolerable error when using a published formula or software package since the distinction between absolute and relative error is often not clear in either the formula documentation or the graphic user interface, in the case of computer software. Similarly, when making a statement

of the criteria used for sample size calculations when reporting the results of a study, care should be taken to ensure that the “relative” or “absolute” qualifier is used when referring to tolerable error.

In the absence of prior knowledge of the event prevalence in a population a conservative sample size estimate can be made assuming event prevalence is 0.5, since the variance of a prevalence (that is, $P \times [1 - P]$) is greatest when $P = 0.5$ and the absolute tolerable error and level of confidence remains fixed (8).

A worked example of a sample size calculation to estimate a prevalence using simple random sampling is shown in **Box 1**.

With stratified sampling the sampling frame is divided into groups (strata) and a random sample is taken from each stratum. When the variation of the outcome of interest within each stratum is small relative to the variation between strata, stratified random sampling returns a more precise estimate of the population parameter compared with simple random sampling.

Sample Size Calculations to Test a Hypothesis Using an Observational Study Design

Details of the formula-based methods to estimate a sample size for each of the main observational study (i.e., cross-sectional, case-control, and cohort studies) are provided in **Tables 3, 4**. Again, these formulae all assume that data are independent.

TABLE 4 | Formulae to estimate a sample size for each of the common observational epidemiological study designs.

Study design	Formula	Arguments
Cross-sectional	$n \geq \frac{r+1}{r(\lambda-1)^2 \pi^2} \left[Z_{1-(\alpha/2)} \sqrt{(r+1)\rho_c(1-\rho_c)} + Z_{1-\beta} \sqrt{\lambda\pi(1-\lambda\pi) + r\pi(1-\pi)} \right]^2$	<p>n = the number of subjects in the sample. r = the anticipated number of subjects in the exposed group divided by the anticipated number of subjects in the unexposed group. λ = the expected prevalence ratio. π = the expected prevalence of the outcome among the non-exposed. $Z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $Z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. ρ_c = the common prevalence over exposed and unexposed groups. $Z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $Z_{1-\beta} = -0.84$ for 80% power.</p>
Case-control	$n \geq \frac{\rho_0}{r+1} \left(\frac{r\lambda}{1+(\lambda-1)\rho_0} + 1 \right) \frac{(r+1)(1+(\lambda-1)\rho_0)^2}{r\rho_0^2(\rho_0-1)^2(\lambda-1)^2} \left[Z_{1-(\alpha/2)} \sqrt{(r+1)\rho_c^*(1-\rho_c^*)} + Z_{1-\beta} \sqrt{\frac{\lambda\rho_0(1-\rho_0)}{[1+(\lambda-1)\rho_0]^2} + r\rho_0(1-\rho_0)} \right]^2$	<p>n = the number of subjects in the sample. ρ_0 = the expected prevalence of exposure among the controls. r = anticipated number of subjects in the control group divided by the anticipated number of subjects in the case group. λ = the expected odds ratio. $Z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $Z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. $Z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $Z_{1-\beta} = 0.84$ for 80% power.</p>
Cohort, count data	$n \geq \frac{r+1}{r(\lambda-1)^2 \pi^2} \left[Z_{1-(\alpha/2)} \sqrt{(r+1)\rho_c(1-\rho_c)} + Z_{1-\beta} \sqrt{\lambda\pi(1-\lambda\pi) + r\pi(1-\pi)} \right]^2$	<p>n = the number of subjects in the sample. r = the anticipated number of subjects in the exposed group divided by the anticipated number of subjects in the unexposed group. λ = the expected incidence risk ratio. π = the expected prevalence of the outcome among the non-exposed. $Z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $Z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. ρ_c = the common prevalence over exposed and unexposed groups. $Z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $Z_{1-\beta} = 0.84$ for 80% power.</p>
Cohort, time at risk	$\lambda'_0 = \frac{\lambda_0^3 FT}{\lambda_0 FT - 1 + \exp(-\lambda_0 FT)} \quad \lambda'_1 = \frac{\lambda_1^3 FT}{\lambda_1 FT - 1 + \exp(-\lambda_1 FT)} \quad \bar{\lambda}' = \frac{\bar{\lambda}^3 FT}{\bar{\lambda} FT - 1 + \exp(-\bar{\lambda} FT)}$ $n_A = m_B n_A \geq \frac{\left(Z_{1-(\alpha/2)} \sqrt{(1+r)\bar{\lambda}'} + Z_{1-\beta} \sqrt{r \times \lambda'_1 + \lambda'_0} \right)^2}{r(\lambda'_1 - \lambda'_0)^2}$	<p>n_A = the number of subjects in the sample. λ_0 = the expected incidence rate among the unexposed. λ_1 = the expected incidence rate among the exposed. $\bar{\lambda}' = (\lambda_0 + \lambda_1) / 2$ FT = the expected follow-up period for the study. r = anticipated number of subjects in the exposed group divided by the anticipated number of subjects in the unexposed group. $Z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $Z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. $Z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $Z_{1-\beta} = 0.84$ for 80% power.</p>

BOX 2 | A prospective cohort study of dry food diets and feline lower urinary tract disease (FLUTD) in mature male cats is planned. A sample of cats will be selected at random from the population and owners who agree to participate in the study will be asked to complete a questionnaire at the time of enrolment. Cats enrolled into the study will be followed for at least 5 years to identify incident cases of FLUTD. The investigators would like to be 0.80 certain of being able to detect when the risk ratio of FLUTD is 1.4 for cats habitually fed a dry food diet, using a 0.05 significance test. Previous evidence suggests that the incidence risk of FLUTD in cats not on a dry food (i.e., “other”) diet is around 50 per 1000 per year. Assuming equal numbers of cats on dry food and other diets are sampled, how many cats should be enrolled into the study?

$$\lambda'_0 = \frac{\lambda_0^3 FT}{\lambda_0 FT - 1 + \exp(-\lambda_0 FT)} \quad \lambda'_1 = \frac{\lambda_1^3 FT}{\lambda_1 FT - 1 + \exp(-\lambda_1 FT)} \quad \bar{\lambda}' = \frac{\bar{\lambda}^3 FT}{\bar{\lambda} FT - 1 + \exp(-\bar{\lambda} FT)}$$

$$n_A \geq \frac{\left(z_{1-(\alpha/2)} \sqrt{(1+r)\bar{\lambda}'} + z_{1-\beta} \sqrt{(r \times \lambda'_1 + \lambda'_0)} \right)^2}{r(\lambda'_1 - \lambda'_0)^2}$$

$$n_A \geq \frac{\left(1.96 \sqrt{(1+1)0.0264^2} + 0.84 \sqrt{(1 \times 0.313 + 0.0217)} \right)^2}{1(0.07 - 0.05)^2}$$

$$n_A \geq \frac{(0.4509 + 0.1935)^2}{0.0004}$$

$$n_A \geq 1040$$

A total of 2,080 male cats need to be sampled to meet the requirements of the study (1,040 cats habitually fed dry food and 1,040 cats habitually fed “other” diet types).

BOX 3 | A case-control study of the association between white pigmentation around the eyes and ocular squamous cell carcinoma in Hereford cattle is planned. A sample of cattle with newly diagnosed squamous cell carcinoma will be compared for white pigmentation around the eyes with a sample of controls. Assuming an equal number of cases and controls, how many study subjects are required to detect an odds ratio of 2.0 with 0.80 power using a two-sided 0.05 test? Previous surveys have shown that around 0.30 of Hereford cattle without squamous cell carcinoma have white pigmentation around the eyes.

$$n \geq \frac{(r+1)(1+(\lambda-1)\rho_0)^2}{r\rho_0^2(\rho_0-1)^2(\lambda-1)^2} \left[z_{1-(\alpha/2)} \sqrt{(r+1)\rho_0^2(1-\rho_0^2)} + z_\beta \sqrt{\frac{\lambda\rho_0(1-\rho_0)}{[1+(\lambda-1)\rho_0]^2} + r\rho_0(1-\rho_0)} \right]^2$$

$$n \geq \frac{(1+1)(1+(2-1)0.3)^2}{1 \times 0.3^2(0.3-1)^2(2-1)^2} \left[1.96 \sqrt{(1+1)0.38(1-0.38)} + 0.84 \sqrt{\frac{2 \times 0.3(1-0.3)}{[1+(2-1)0.3]^2} + 1 \times 0.3(1-0.3)} \right]^2$$

$$n \geq \frac{3.38}{0.0441} [1.346 + 0.569]^2$$

$$n \geq 282$$

If the true odds for squamous cell carcinoma in exposed subjects relative to unexposed subjects is 2.0, we will need to enrol 141 cases and 141 controls (282 cattle in total) to reject the null hypothesis that the odds ratio equals one with probability (power) 0.80. The Type I error probability associated with this test of this null hypothesis is 0.05.

Note that the sample size formulae for cross-sectional studies, cohort studies using count data and cohort studies using time at risk require the analyst to provide an estimate of prevalence, incidence risk and incidence rate (respectively) for both risk factor exposed and unexposed groups. **Box 2** provides a worked example for a prospective cohort study, with a fixed follow-up time.

In contrast to sample size formulae for cross-sectional and cohort studies, the sample size formula for case-control studies requires provision of an estimate of the prevalence of exposure amongst controls (**Box 3**). An additional consideration when estimating an appropriate sample size for a case-control study is specification of the design – either matched or unmatched (12). The process of matching provides a means for controlling for

the effect of a known confounder with the added benefit of an increase in statistical efficiency (12, 13).

Sample Size Calculations to Test a Hypothesis Using a Randomized Clinical Trial

A superiority trial is a study in which the aim is to show that a treatment intervention provides a better therapeutic outcome than a known reference (often a placebo) and the statistical procedure to provide this evidence is called a superiority test (14).

In situations where an established treatment already exists a study comparing a new treatment to a placebo (effectively, no treatment) will be unethical. In this situation interest lies

TABLE 5 | Information required to estimate a sample size for equivalence, superiority and non-inferiority trials.

Outcome variable	Study design	Arguments	References
Continuous	Equivalence trial	The expected mean of the outcome variable in the treatment and control groups, the expected population standard deviation of the outcome variable, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15–17).
Binary	Equivalence trial	The expected proportion of successes in the treatment and control groups, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15–17).
Continuous	Superiority trial	The expected mean of the outcome variable in the treatment and control groups, the expected population standard deviation of the outcome variable, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15).
Binary	Superiority trial	The expected proportion of successes in the treatment and control groups, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15)
Continuous	Non-inferiority trial	The expected mean of the outcome variable in the treatment and control groups, the expected population standard deviation of the outcome variable, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15–17)
Binary	Non-inferiority trial	The expected proportion of successes in the treatment and control groups, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(16, 17)

TABLE 6 | Formulae to estimate a sample size for equivalence, superiority and non-inferiority trials, binary or continuous population parameters.

Outcome variable	Study design	Formula	Arguments
Continuous	Equivalence trial	$n_A = rn_B$ $n_B = \left(1 + \frac{1}{r}\right) \left(\sigma \frac{Z_{1-(\alpha/2)} + Z_{1-\beta/2}}{ \mu_A - \mu_B - \delta}\right)^2$ $1 - \beta = 2 \left[\Phi(Z - Z_{1-(\alpha/2)}) + \Phi(-Z - Z_{1-(\alpha/2)})\right] - 1$ $Z = \frac{ \mu_A - \mu_B - \delta}{\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$	<p>μ_A = the expected mean of the outcome in the treatment group. μ_B = the expected mean of the outcome in the control group. σ = the expected standard deviation of the outcome across treatment and control groups. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin.</p>
Binary	Equivalence trial	$n_A = rn_B$ $n_B = \left(\frac{\rho_A(1 - \rho_A) + \rho_B(1 - \rho_B)}{r}\right) \left(\sigma \frac{Z_{1-(\alpha/2)} + Z_{1-\beta/2}}{ \rho_A - \rho_B - \delta}\right)^2$ $1 - \beta = 2 \left[\Phi(Z - Z_{1-(\alpha/2)}) + \Phi(-Z - Z_{1-(\alpha/2)})\right] - 1$ $Z = \frac{ \rho_A - \rho_B - \delta}{\sigma \sqrt{\frac{\rho_A(1 - \rho_A)}{n_A} + \frac{\rho_B(1 - \rho_B)}{n_B}}}$	<p>ρ_A = the expected probability of success in the treatment group. ρ_B = the expected probability of success in the control group. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin.</p>
Continuous	Superiority trial or non-inferiority trial	$n_A = rn_B$ $n_B = \left(1 + \frac{1}{r}\right) \left(\sigma \frac{Z_{1-(\alpha/2)} + Z_{1-\beta}}{\mu_A - \mu_B - \delta}\right)^2$ $1 - \beta = \Phi(Z - Z_{1-\alpha}) + \Phi(-Z - Z_{1-\alpha})$ $Z = \frac{\mu_A - \mu_B - \delta}{\sigma \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$	<p>μ_A = the expected mean of the outcome in the treatment group. μ_B = the expected mean of the outcome in the control group. σ = the expected standard deviation of the outcome across treatment and control groups. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin.</p>
Binary	Superiority trial or non-inferiority trial	$n_A = rn_B$ $n_B = \left(\frac{\rho_A(1 - \rho_A) + \rho_B(1 - \rho_B)}{r}\right) \left(\frac{Z_{1-(\alpha/2)} + Z_{1-\beta}}{\rho_A - \rho_B - \delta}\right)^2$ $1 - \beta = \Phi(Z - Z_{1-(\alpha/2)}) + \Phi(-Z - Z_{1-(\alpha/2)})$ $Z = \frac{\rho_A - \rho_B - \delta}{\sqrt{\frac{\rho_A(1 - \rho_A)}{n_A} + \frac{\rho_B(1 - \rho_B)}{n_B}}}$	<p>ρ_A = the expected probability of success in the treatment group. ρ_B = the expected probability of success in the control group. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin. $Z_{1-(\alpha/2)}$ = values from the standard normal curve corresponding to the desired level of confidence. Use $Z_{1-(\alpha/2)} = 1.96$ for 95% (two side) confidence. $Z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $Z_{1-\beta} = 0.84$ for 80% power.</p>

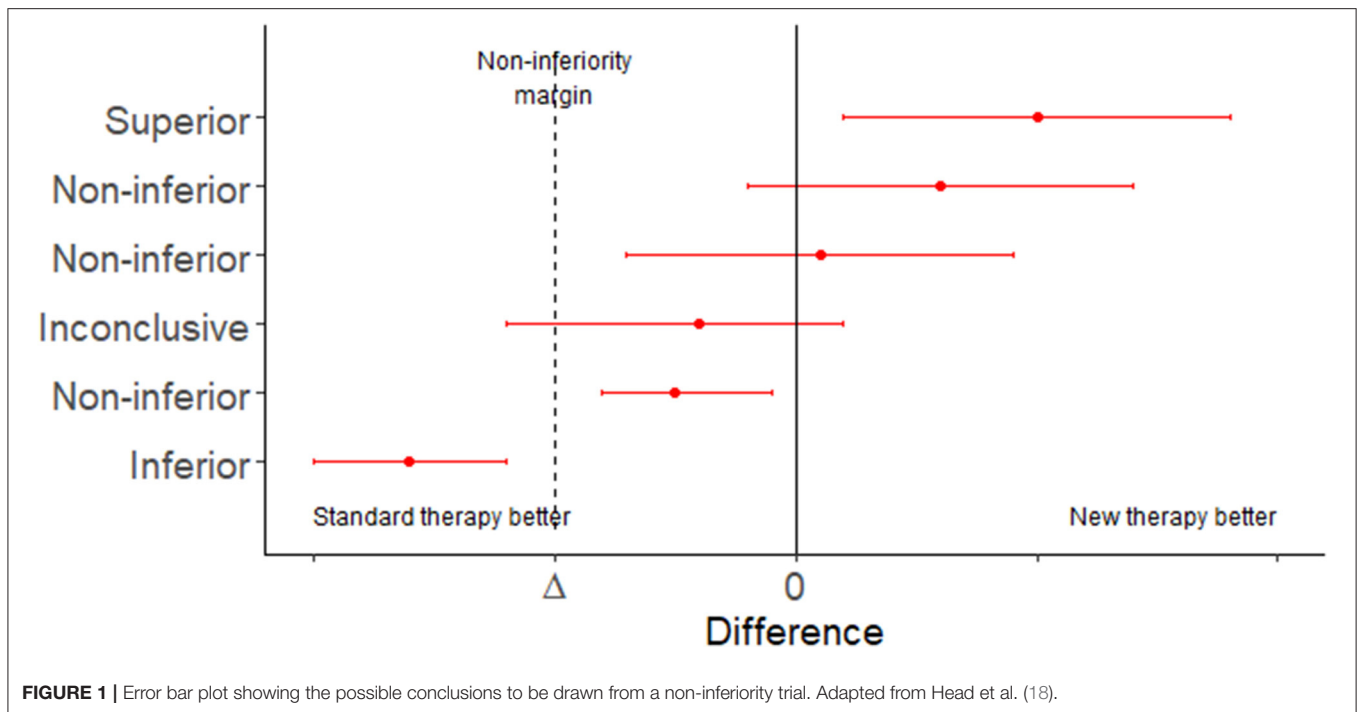


FIGURE 1 | Error bar plot showing the possible conclusions to be drawn from a non-inferiority trial. Adapted from Head et al. (18).

BOX 4 | Suppose a pharmaceutical company would like to conduct a clinical trial to compare the efficacy of two antimicrobial agents when administered orally to patients with skin infections. Assume the true mean cure rate of the treatment is 0.85 and the true mean cure rate of the control is 0.65. We consider a difference of <0.10 in cure rate to be of no clinical importance (i.e., $\delta = -0.10$). Assuming a one-sided test size of 5% and a power of 80% how many subjects should be included in the trial?

$$n_B = \left(\frac{p_A(1-p_A) + p_B(1-p_B)}{r} \right) \left(\frac{Z_{1-\alpha/2} + Z_{1-\beta}}{p_A - p_B - \delta} \right)^2$$

$$n_B = \left(\frac{0.85(1-0.85) + 0.65(1-0.65)}{1} \right) \left(\frac{1.96 + 0.84}{0.85 - 0.65 - (-0.10)} \right)^2$$

$$n_B = \left(\frac{0.355}{1} \right) \left(\frac{2.48}{0.30} \right)^2$$

$$n_B = 25$$

A total of 50 subjects need to be enrolled in the trial, 25 in the treatment group and 25 in the control group.

in determining if the new treatment is: (1) either the same as, or better than, an established treatment using a non-inferiority trial; or (2) equivalent to an existing treatment within a specified range, using an equivalence trial (Tables 5, 6 and Figure 1). Equivalence trials are not to be confused with bioequivalence trials where generic drug preparations are compared to currently marketed formulations with respect to their pharmacokinetic parameters.

Superiority, non-inferiority and equivalence trials require the analyst to specify an equivalence margin. The equivalence margin is the range of values for which the treatment efficacies are close enough to be considered the same (19). Expressed in another way, the equivalence margin is the maximum clinically acceptable difference one is

willing to accept in return for the secondary benefits of the new treatment.

Equivalence margins can be set on the basis of a clinical estimation of a minimally important effect. This approach is subjective and, as a result, it is possible to set the equivalence margin to be greater than the effect of the established treatment, which could lead to potentially harmful treatments classified as non-inferior. A second approach is to select an equivalence margin with reference to the effect of the established treatment in trials where a placebo has been used. When the equivalence margin is chosen in this way, there is some objective basis on which to claim that a positive non-inferiority trial implies that a new treatment is, in fact, superior to the established treatment (assuming the effect of the established treatment in the current

TABLE 7 | Summary of sample size formulae to estimate the probability of disease freedom or estimate surveillance system sensitivity.

Study design	Outcome	Arguments	References
Representative sampling	Probability of disease freedom assuming imperfect test sensitivity and perfect test specificity.	The assumed population size, an estimate of the prior probability that the population is free of disease, an estimate of the probability of disease introduction, the population level design prevalence, the desired probability that the population is free of disease, the sensitivity of the diagnostic test used at the surveillance unit level.	(22, 23)
Representative sampling	Surveillance system sensitivity assuming a single risk factor and varying test sensitivity.	The assumed population size, the population level design prevalence, the desired surveillance system sensitivity, the sensitivity of the diagnostic test used at the surveillance unit level.	(24, 25)
Representative sampling	Surveillance system sensitivity assuming two stage sampling, imperfect test sensitivity and perfect test specificity.	The number of clusters in the population, the number of surveillance units within each cluster, the cluster level design prevalence, the desired cluster level sensitivity, the population level design prevalence, the desired population level sensitivity, the sensitivity of the diagnostic test used at the surveillance unit level.	(24–27)
Representative sampling	Surveillance system sensitivity, imperfect test sensitivity and imperfect test specificity.	The assumed population size, the population level design prevalence, the desired population level sensitivity, the desired population level specificity, the sensitivity of the diagnostic test at the surveillance unit level, the specificity of the diagnostic test at the surveillance unit level.	(26–28)
Representative sampling	Surveillance system sensitivity assuming pooled sampling giving rise to imperfect test sensitivity and imperfect test specificity.	The number of surveillance units that contribute to each pool, the population level design prevalence, the sensitivity of the diagnostic test at the pooled level, the specificity of the diagnostic test at the pooled level, the desired population level sensitivity.	(29)
Risk-based sampling	Surveillance system sensitivity assuming imperfect test sensitivity and perfect test specificity.	The population level design prevalence, relative risk estimates for each strata, the population proportions for each strata, the surveillance proportions for each strata, the desired population level sensitivity, the sensitivity of the diagnostic test at the surveillance unit level.	(5)
Risk-based sampling	Surveillance system sensitivity assuming risk-based 2-stage sampling on one risk factor at the cluster level assuming imperfect test sensitivity and perfect test specificity.	Relative risk values for each strata in the population, the population proportions in each strata, the planned number of units to be sampled from each strata, the cluster level design prevalence, the desired cluster level sensitivity, the surveillance unit level design prevalence, the sensitivity of the diagnostic test at the surveillance unit level, the desired surveillance system (population-level) sensitivity.	(5)
Risk-based sampling	Surveillance system sensitivity assuming risk-based 2-stage sampling on two risk factors at either the cluster level, the unit level or both, imperfect test sensitivity and perfect test specificity.	The number of risk strata defining the relative risk values at the cluster level, the population proportions at the cluster level, the planned surveillance proportions at the cluster level, the cluster level design prevalence, the desired cluster level sensitivity, the number of risk strata defining the relative risk values at the surveillance unit level, the population proportions at the surveillance unit level, the planned surveillance proportions at the surveillance unit level, the surveillance unit level design prevalence, the sensitivity of the diagnostic test at the surveillance unit level, the desired surveillance system (population-level) sensitivity.	(5)

trial is similar to its effect in the historical trials). An example sample size calculation for a non-inferiority trial is presented in **Box 4**.

Sample Size Calculations to Detect the Presence of an Event

Sampling of individuals to either detect the presence of an event (usually the presence of disease or the presence of infection) or provide evidence that disease is absent from a jurisdiction are frequent activities in veterinary epidemiology. Typical scenarios include: (1) shipment of live animals from one country to another where the country receiving the shipment might request that testing is carried out on a sample of individuals, as opposed to testing every animal; and (2) a country wishing to re-gain official disease freedom status following an infectious disease outbreak.

Details of formula-based sample size estimation methods to detect the presence of an event are provided in **Table 7**. Sample size estimation methods can be categorized into two groups: (1) to ensure sufficient units are sampled to return a desired (posterior) probability of disease freedom; and (2) to ensure sufficient units are sampled to ensure a surveillance system has a

desired system sensitivity. For the surveillance system sensitivity methods sampling can be either representative or risk based. All of the methods listed in **Table 7** account for imperfect diagnostic test sensitivity at the surveillance unit level.

When tests with both imperfect diagnostic sensitivity and specificity are used, diseased individuals can be missed because of imperfect diagnostic sensitivity but at the same time disease negative individuals can be incorrectly identified as disease positive because of imperfect specificity. Cameron and Baldock (26) describe an approach to estimate the number of animals to be sampled from a finite population using a test with imperfect diagnostic sensitivity and specificity using the hypergeometric distribution. This method returns the number of individuals to be sampled and the estimated probability that the population is diseased for 1 to n individuals that return a positive test result. This allows an analyst to make a statement that they can be (for example) 95% confident that the prevalence of disease in the population of interest is less than the stated design prevalence if the number of (surveillance) units with a positive test result is less than or equal to a specified cut-point. A worked example of this approach is provided in the **Supplementary Material**.

BOX 5 | An aid project has distributed cook stoves in a single province in a resource-poor country. At the end of 3 years, the donors would like to know what proportion of households are still using their donated stove. A cross-sectional study is planned where villages in a province will be sampled and all households (~75 per village) will be visited to determine if the donated stove is still in use. A pilot study of the prevalence of stove usage in five villages showed that 0.46 of householders were still using their stove and the ICC for stove use within villages is in the order of 0.20. If the donor wanted to be 95% confident that the survey estimate of stove usage was within 10% of the true population value, how many villages (clusters) need to be sampled?

$$D = 1 + (b - 1)\rho$$

$$D = 1 + (75 - 1) \times 0.20$$

$$D = 15.8$$

$$n_c \geq \frac{z_{1-(\alpha/2)}^2 P_y (1 - P_y) D}{(P_y \epsilon_r)^2 b}$$

$$n_c \geq \frac{1.96^2 0.46 (1 - 0.46) 15.8}{(0.46 \times 0.10)^2 \times 75}$$

$$n_c \geq \frac{15.077}{0.1587}$$

$$n_c \geq 96$$

A total of 96 villages need to be sampled to meet the requirements of the study.

BOX 6 | Continuing the example presented in Box 5, we are now told that the number of households per village varies. The average number of households per village is 75 with a 0.025 quartile of 40 households and a 0.975 quartile of 180. Assuming the number of households per village follows a normal distribution the expected standard deviation of the number of households per village is in the order of $(180 - 40) \div 4 = 35$. How many villages need to be sampled? In the formula below, CV stands for coefficient of variation defined as the standard deviation of the cluster sizes divided by the mean of the cluster sizes.

$$D = 1 + \{(CV^2 + 1) b - 1\} \rho$$

$$D = 1 + \{(0.467^2 + 1) 75 - 1\} 0.2$$

$$D = 19.1$$

$$n_c \geq \frac{z_{1-(\alpha/2)}^2 P_y (1 - P_y) D}{(P_y \epsilon_r)^2 b}$$

$$n_c \geq \frac{1.96^2 0.46 (1 - 0.46) 19.1}{(0.46 \times 0.10)^2 \times 75}$$

$$n_c \geq \frac{18,194}{0.1587}$$

$$n_c \geq 115$$

A total of 115 villages need to be sampled to meet the requirements of the study.

BOX 7 | Continuing the example provided in Box 1, being seropositive to brucellosis is likely to cluster within herds. Otte and Gumm (20) cite the intracluster correlation coefficient for *Brucella abortus* in cattle to be in the order of 0.09. We now adjust our sample size estimate of 545 to account for clustering at the herd level. Assume that, on average, $b = 20$ animals will be sampled per herd:

$$D = 1 + (b - 1)\rho$$

$$D = 1 + (20 - 1) \times 0.09$$

$$D = 2.71$$

After accounting for the presence of clustering at the herd level we estimate that a total of $(545 \times 2.71) = 1,477$ cattle need to be sampled to meet the requirements of the survey. If 20 cows are sampled per herd this means that a total of $(1,477 \div 20) = 74$ herds are required.

BOX 8 | Dohoo et al. (21) provide details of an observational study of the reproductive performance of dairy cows on Reunion Island. If this study were to be repeated, how many lactations would need to be sampled to be 95% confident that the estimated logarithm of calving to conception interval was within 5% of the true population value?

From (21) the standard deviations of the random effect terms from a multilevel model of factors influencing log transformed calving to conception interval at the herd, cow and lactation level were 0.1157, 0.1479, and 0.5116, respectively. The ICC for lactations within herds (Equation 3):

$$\rho_2 = \frac{\sigma_3^2}{\sigma_3^2 + \sigma_2^2 + \sigma_1^2}$$

$$\rho_2 = \frac{0.1157^2}{0.1157^2 + 0.1479^2 + 0.5116^2}$$

$$\rho_2 = 0.0451$$

and the ICC for lactations within cows (Equation 4):

$$\rho_1 = \frac{\sigma_3^2 + \sigma_2^2}{\sigma_3^2 + \sigma_2^2 + \sigma_1^2}$$

$$\rho_1 = \frac{0.1157^2 + 0.1479^2}{0.1157^2 + 0.1479^2 + 0.5116^2}$$

$$\rho_1 = 0.1188$$

The mean and standard deviation of the logarithm of calving to conception interval was 4.59 and 0.54, respectively. What is the required sample size assuming the data are independent?

$$m_1 \geq \frac{1.96 \times 1,000,000 \times (0.54^2 \div 4.59^2)}{1.96 (0.54^2 \div 4.59^2) + ([1,000,000 - 1] 0.05^2)}$$

$$m_1 \geq \frac{54057.4}{2500.025}$$

$$m_1 \geq 22$$

Assuming the data are independent a total of 22 lactations are required to be 95% confident that our estimate of the logarithm of calving to conception interval is within 5% of the true population value.

We elect to sample two lactations per cow. How many lactations are required to account for clustering of lactations within cows?

$$n_1 = 2$$

$$D_1 = 1 + \rho_1 (n_1 - 1)$$

$$D_1 = 1 + 0.1188 (2 - 1)$$

$$D_1 = 1.1188$$

$$m_2 = D_1 \times m_1$$

$$m_2 = 1.1188 \times 22$$

$$m_2 = 25$$

A total of 25 lactations are required accounting for clustering of lactations within cows. How many cows are required?

$$n_2 = m_2 / n_1$$

$$n_2 = 25 / 2$$

$$n_2 = 13$$

A total of 13 cows are required if we sample two lactations per cow (26 lactations in total).

We now consider clustering at the herd level. How many lactations are required to account for clustering of cows within herds?

$$D_2 = 1 + (n_1 \times (n_2 - 1) \times \rho_2) + ((n_1 - 1) \times \rho_1)$$

$$D_2 = 1 + (2 \times (13 - 1) \times 0.0451) + ((2 - 1) \times 0.1188)$$

$$D_2 = 2.2016$$

$$m_3 = D_2 \times m_1$$

$$m_3 = 2.2016 \times 22$$

$$m_3 = 49$$

Continued

BOX 8 | Continued

Accounting for clustering of lactations within cow and cows within herds, a total of 39 lactations are required. How many herds are required?

$$\begin{aligned}n_3 &= m_3 / (n_1 \times n_2) \\n_3 &= 49 / (2 \times 13) \\n_3 &= 2\end{aligned}$$

A total of 2 herds are required if we sample 13 cows from each herd and 2 lactations from each cow. The total number of lactations required is therefore:

$$\begin{aligned}n_{total} &= (n_1 \times n_2 \times n_3) \\n_{total} &= (2 \times 13 \times 2) \\n_{total} &= 52\end{aligned}$$

To account for lack of independence in the data arising from clustering of lactations within cows and cows within herds 52 lactations (2 lactations from 13 cows from 2 herds) are required to meet the requirements of the study.

The required sample size assuming the data were independent was 22. The required sample size accounting for lack of independence in the data was 52, a 2.5-fold difference.

SAMPLE SIZE CALCULATIONS FOR CLUSTERED DATA

Aggregation of individual sampling units into groups (“clusters”) for example farms, households or villages violates the assumption of independence that is central to the sample size calculation methods described so far. When individuals are aggregated into clusters there are two sources of variation in the outcome of interest. The first arises from the effect of the cluster; the second from the effect of the individual. This means that individuals selected from the same cluster are more likely to be similar compared with those sampled from the general population (30). For this reason, the effective sample size when observations are made on randomly selected individuals from the same cluster will be less than that when observations are made on individuals selected completely at random from the general population. For studies where the objective is to estimate a population parameter (e.g., a prevalence) a reduction in effective sample size increases the uncertainty around the estimate of the population parameter. For studies where the objective is to test a hypothesis a reduction in effective sample size results in a reduction in statistical power, in effect the ability to detect a statistically significant difference in event outcomes for exposure positive and exposure negative individuals given a true difference actually exists.

With one-stage cluster sampling a random sample of clusters is selected first and then all individual listing units within each cluster are selected for study. With two-stage cluster sampling a random sample of clusters is selected first and then a random sample of individual listing units within each cluster is selected. The primary advantage of cluster sampling is logistics. In animal health, where animals are typically managed within clusters (e.g., herds or flocks) it is easier to select clusters first and then, from each selected cluster, take a sample of individual animals. This contrasts with a simple random sampling approach which would require an investigator to travel to a large number of herds-flocks, sampling small numbers of animals from each. As explained

above, the main disadvantage of cluster sampling is a reduction in the effective sample size due to animals from the same cluster being more homogenous (similar) compared with those from different clusters.

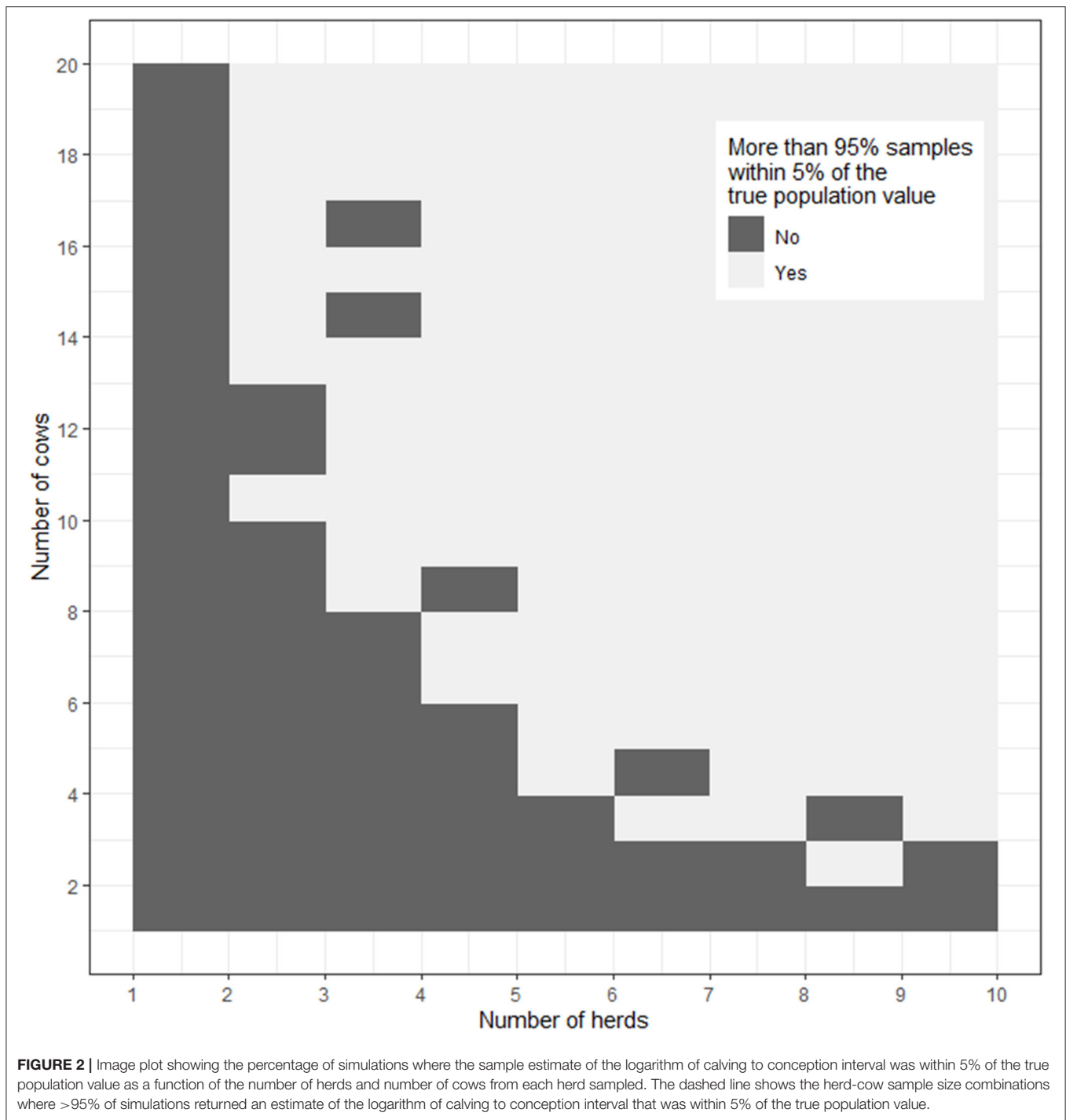
To compensate for this lack of precision Donner et al. (31) proposed that a sample size estimate calculated assuming complete independence (using the formulae presented in **Tables 1–7**) can be inflated by a value known as the design effect (D) to achieve the level of statistical power achieved using independent sampling. For a single level of clustering (e.g., the situation where cows are clustered within herds) the design effect is calculated as:

$$D = 1 + (b - 1)\rho \quad (1)$$

In Equation 1 b equals the number of animals to be sampled from each cluster (not to be confused with the total number of animals eligible for sampling within each cluster) and ρ is the intraclass correlation coefficient (ICC). The value of ρ equals the between-cluster variance σ_B^2 divided by the between-cluster variance plus the within-cluster variance ($\sigma_B^2 + \sigma_W^2$):

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \quad (2)$$

When there is little variation in an outcome within a cluster (e.g., observations made on individual cows within herds are “similar,” σ_W^2 will be small) ρ will be close to 1 and the design effect will therefore be large. When there is wide variation within a cluster (e.g., observations made on individual cows within herds showing a similar variability to the general population, σ_W^2 will be large) ρ will be close to 0 and, therefore, the design effect will be close to unity. Using the definition provided above (Equation 2), ρ ranges between 0 and +1 with typical values ranging from 0 to 0.05 for non-communicable diseases and values >0.4 uncommon. Papers providing ICC estimates for various outcomes in the human



and veterinary literature have been published: see, for example, (20, 32, 33). Researchers should be aware of the importance of publishing estimates of ICC since high quality empirical data are necessary to provide credible sample size estimates for future studies. More importantly, for the same outcome measure, ICC estimates will vary from one research setting to another so access to a likely range of ICC measures is desirable.

A number of methods are available to estimate ρ from empirical data (34, 35) ranging from one-way analysis of variance (36) to regression-based approaches using mixed effects models (37). Eldridge and Kerry (38) provide a comprehensive review of appropriate techniques.

An example of how the ICC can be used to estimate the number of primary sampling units for a one-stage cluster design is provided in **Box 5**.

The example shown in **Box 5** is somewhat unrealistic in that it is assumed that the number of households in each village is a constant value of 75. Eldridge, Ashby and Kerry (39) provide an approach to estimate a sample size using a one-stage cluster design when the number of individual listing units per cluster varies (**Box 6**).

An example showing how a crude sample size estimate (i.e., a sample size calculated assuming independence) can be adjusted to account for clustering using the design effect is provided in **Box 7**.

Three levels of clustering are relatively common in veterinary epidemiological research (much more so than in human epidemiology) where, for example, lactations (level 1 units) might be sampled within cows (level 2 units) which are then sampled within herds (level 3 units). The total variance in this situation is made up of the variance associated with lactations within cows within herds σ_1^2 , the variance between cows within herds σ_2^2 , and the variance between herds σ_3^2 . Two ICCs can be calculated: lactations within herds:

$$\rho_2 = \frac{\sigma_3^2}{\sigma_3^2 + \sigma_2^2 + \sigma_1^2} \quad (3)$$

and lactations within cows:

$$\rho_1 = \frac{\sigma_3^2 + \sigma_2^2}{\sigma_3^2 + \sigma_2^2 + \sigma_1^2} \quad (4)$$

In a study comprised of three levels the required sample size, accounting for clustering equals (40):

$$n_3 n_2 n_1 = DE \times m \quad (5)$$

In Equation 5, m is the number of lactations to be sampled to meet the requirements of the study assuming the data are completely independent and n_3 , n_2 , and n_1 are the number of units to be sampled at the herd, cow and lactation level (respectively). The design effect for three levels of clustering equals:

$$DE = 1 + n_1 (n_2 - 1) \rho_2 + (n_1 - 1) \rho_1 \quad (6)$$

Box 8 provides a worked example of a sample size calculation for the three-level clustering scenario.

SIMULATION-BASED APPROACHES FOR SAMPLE SIZE ESTIMATION

In applied veterinary epidemiological research it is common for study designs not to conform to the standard study designs for which sample size formulae are available. Typical examples include situations where study subjects are organized into more than three levels of aggregation and in clinical trials where a treatment might be applied at the group level and a second treatment applied at the individual level. Where there are multiple levels of aggregation researchers may elect to apply a

more conservative design effect multiplier than that used when study subjects are kept in simpler cluster groups. While this approach is an attempt to address the problem, it can result in the final sample size estimate being larger than the final sample size required if the design effect was known more precisely.

For complex study designs simulation-based approaches provide an alternative for sample size estimation that is relatively easy to implement using modern statistical software (41). In the text that follows a worked example is provided, where simulation is used to estimate the number of lactations, cows and herds to be sampled to provide an estimate of log calving to conception interval, using the scenario presented in **Box 8**.

The general approach when using simulation to estimate a sample size to estimate a population parameter is to: (1) simulate a population data set that respects clustering of the outcome variable within the population of interest; (2) define a series of candidate sample size estimates; (3) repeatedly sample the simulated population using each of the candidate sample size estimates to determine the proportion of occasions the estimate of the population parameter is within the prescribed relative error of the true population value. When estimating a population prevalence and assuming the level of confidence specified by the analyst has been set to 95%, the required sample size is the combination of level 1, 2, 3, ... n units sampled that returns an estimate of the outcome variable that is within the prescribed relative error of the true population value on 95% of occasions. Note that several different combinations of units sampled at each level might achieve the stated objectives of the study.

When the study aim is to test a hypothesis, an additional step is to assign the exposure variable (e.g., a treatment) to members of the population and then to estimate the effect of the exposure on the outcome of interest using a regression approach. Arnold et al. (41) provide a worked example of this approach using a two-treatment factorial trial in rural Bangladesh as an example. In this study children <6 months of age were randomly assigned to one of four treatment groups: control, sanitation mobilization, lipid-based nutrient supplementation, and sanitation plus lipid-based nutrient supplementation. The design of this study made sample size and study power calculations difficult for two reasons: (1) treatments were deployed at two levels (sanitation mobilization at the community level and lipid supplementation at the individual level); and (2) there were two sources of correlation in the outcome: at the community level and the individual child level.

Generation of the population data set involves: (1) defining the mean and standard deviation of the outcome of interest (for continuous outcomes) or the expected population prevalence (for binary outcomes); (2) defining the number of level 1, 2, 3, ... n units in the population; (3) defining the level 1, 2, 3, ... n variance terms; (4) simulating a population of individuals eligible for sampling based on the specified number of level 1, 2, 3, ... n units; (5) assignment of a value for the outcome variable to each member of the simulated population, and; (6) adjustment of the value of the outcome variable for each individual to account for clustering using the level 1, 2, 3, ... n variance terms.

Code written in the R programming language (6) to generate a population data set for the Reunion Island dairy cow reproduction example (25) and estimate a sample size to meet the requirements of the study is provided in the Supplementary material accompanying this paper.

Figure 2 is an image plot showing the proportion of simulations where the sample estimate of the logarithm of calving to conception interval was within 5% of the true population value as a function of the number of sampled herds and the number of cows sampled from within each herd. In **Figure 2** the superimposed contour line shows the herd-cow sample size combinations where >95% of simulations returned an estimate of the logarithm of calving to conception interval that was within 5% of the true population value, in agreement with the requirement for 2 lactations from 13 cows from 2 herds ($n = 52$ lactations) calculated using the formula-based approach presented in **Box 8**. When the results of simulations are presented in this way one can appreciate that there is some flexibility in the combinations of herd and cow numbers that need to be sampled to meet the requirements of the study. For example, **Figure 2** shows that the estimate of mean logarithm of calving to conception interval would be within 5% of the true population value if a smaller number of cows (e.g., $n = 6$) were sampled from a larger number of herds (e.g., $n = 6$).

In summary, the process of simulation replaces the time and effort to derive a formula-based approach for a complex study design with basic programming and computer simulation time. An additional positive side effect is that the process of simulation requires investigators to define the structure of their study population, the expected value and variability of the outcome of interest and how the results of the study will be analyzed once

the data are collected. This reduces the likelihood of investigators exploring alternative analytical approaches in the presence of negative findings, consistent with CONSORT guidelines (42).

CONCLUSIONS

This paper has provided an overview of the reasons researchers might need to calculate an appropriate sample size in veterinary epidemiology and a summary of different sample size calculation methods. In contrast to human epidemiology individual study subjects in veterinary epidemiology are almost always aggregated into hierarchical groups (43) and, for this reason, sample size estimates calculated using simple formulae that assume independence are usually not appropriate in a veterinary setting. This paper provides details of two approaches for dealing with this problem: (1) inflation of a crude sample size estimate using a design effect; and (2) use of a simulation-based approaches. The key advantage of simulation-based approaches is that appropriate sample sizes can be estimated for complex study designs for which formula-based methods are not available.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2020.539573/full#supplementary-material>

REFERENCES

- Detsky A, Sackett D. When was a “negative” clinical trial big enough? How many patients you needed depends on what you found. *Arch Intern Med.* (1985) 145:709–12. doi: 10.1001/archinte.1985.00360040141030
- Fitzner K, Heckinger E. Sample size calculation and power analysis: a quick review. *Diab Educ.* (2010) 36:701–7. doi: 10.1177/0145721710380791
- Cook J, Julious S, Sones W, Hampson L, Hewitt C, Berlin J, et al. DELTA(2) guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *Br Med J.* (2018) 363:k3750. doi: 10.1136/bmj.k3750
- Delucchi K. Sample size estimation in research with dependent measures and dichotomous outcomes. *Am J Public Health.* (2004) 94:372–7. doi: 10.2105/AJPH.94.3.372
- Stevenson M, Sergeant E, Nunes T, Heuer C. epiR: tools for the analysis of epidemiological data. In: *Faculty of Veterinary and Agricultural Sciences.* Melbourne, Vic: The University of Melbourne (2020). Available online at: <https://cran.rproject.org/web/packages/epiR/index.html>
- R Core Team. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria (2020).
- Levy P, Lemeshow S. *Sampling of Populations Methods and Applications.* London: Wiley Series in Probability and Statistics (2008).
- Kirkwood B, Sterne J. *Essential Medical Statistics.* London: Wiley-Blackwell (2003).
- Woodward M. *Epidemiology Study Design and Data Analysis.* New York, NY: Chapman & Hall CRC (2014).
- Dupont W. Power calculations for matched case-control studies. *Biometrics.* (1988) 44:1157–68. doi: 10.2307/2531743
- Lemeshow S, Hosmer D, Klar J, Lwanga S. *Adequacy of Sample Size in Health Studies.* New York, NY: John Wiley and Sons (1990).
- Breslow N, Day N. *Statistical Methods in Cancer Research Volume 1: The Analysis of Case-Control Studies IARC Scientific Publication No. 32.* Lyon: International Agency for Research on Cancer (1980).
- Rose S, Laan M. Why match? Investigating matched case-control study designs with causal effect estimation. *Int J Biostat.* (2009) 5:1. doi: 10.2202/1557-4679.1127
- Lesaffre E. Superiority, equivalence, non-inferiority trials. *Bull NYU Hosp Joint Dis.* (2008) 66:150–4.
- Chow S-C, Shao J, Wang H, Lohknygina Y. *Sample Size Calculations in Clinical Research.* London: CRC Press (2018).
- Julious S. *Sample Sizes for Clinical Trials.* London: CRC Press.
- Machin D, Campbell M, Tan S, Tan S. *Sample Size Tables for Clinical Studies.* London: John Wiley & Sons (2009).
- Head S, Kaul S, Bogers A, Kappetein A. Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J.* (2012) 33:1318–24. doi: 10.1093/eurheartj/ehs099
- Walker E, Nowacki A. Understanding equivalence and noninferiority testing. *J Gen Int Med.* (2011) 26:192–6. doi: 10.1007/s11606-010-1513-8
- Otte J, Gumm I. Intra-cluster correlation coefficients of 20 infections calculated from the results of cluster-sample surveys. *Prev Vet Med.* (1997) 31:147–50. doi: 10.1016/S0167-5877(96)01108-7
- Dohoo I, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in

- Reunion Island. *Prev Vet Med.* (2001) 50:127–44. doi: 10.1016/S0167-5877(01)00191-X
22. Martin P, Cameron A, Barfod K, Sergeant E, Greiner M. Demonstrating freedom from disease using multiple complex data sources 2: Case study—Classical swine fever in Denmark. *Prev Vet Med.* (2007) 79:98–115. doi: 10.1016/j.prevetmed.2006.09.007
 23. Martin P, Cameron A, Greiner M. Demonstrating freedom from disease using multiple complex data sources 1: a new methodology based on scenario trees. *Prev Vet Med.* (2007) 79:71–97. doi: 10.1016/j.prevetmed.2006.09.008
 24. MacDiarmid S. Future options for brucellosis surveillance in New Zealand beef herds. *N Z Vet J.* (1988) 36:39–42. doi: 10.1080/00480169.1988.35472
 25. Martin S, Shoukri M, Thorburn M. Evaluating the health status of herds based on tests applied to individuals. *Prev Vet Med.* (1992) 14:33–43. doi: 10.1016/0167-5877(92)90082-Q
 26. Cameron A, Baldock F. A new probability formula for surveys to substantiate freedom from disease. *Prev Vet Med.* (1998) 34:1–17. doi: 10.1016/S0167-5877(97)00081-0
 27. Cameron A. *Survey Toolbox — A Practical Manual and Software Package for Active Surveillance of Livestock Diseases in Developing Countries.* Canberra: Australian Centre for International Agricultural Research (1999).
 28. Cameron A, Baldock F. Two-stage sampling in surveys to substantiate freedom from disease. *Prev Vet Med.* (1998) 34:19–30. doi: 10.1016/S0167-5877(97)00073-1
 29. Christensen J, Gardner I. Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Prev Vet Med.* (2000) 45:83–106. doi: 10.1016/S0167-5877(00)00118-5
 30. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol.* (1978) 108:100–2. doi: 10.1093/oxfordjournals.aje.a112592
 31. Donner S, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol.* (1981) 114:906–14. doi: 10.1093/oxfordjournals.aje.a113261
 32. Pagel C, Prost A, Lewycka S, Das S, Colbourn T, Mahapatra R, et al. Intraclass correlation coefficients and coefficients of variation for perinatal outcomes from five cluster-randomised controlled trials in low and middle-income countries: results and methodological implications. *Trials.* (2011) 12:51. doi: 10.1186/1745-6215-12-151
 33. Singh J, Liddy C, Hogg W, Taljaard M. Intraclass correlation coefficients for sample size calculations related to cardiovascular disease prevention and management in primary care practices. *BMC Res Notes.* (2015) 8:89. doi: 10.1186/s13104-015-1042-y
 34. Donner A. A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *Int Stat Rev.* (1986) 54:67–82. doi: 10.2307/1403259
 35. Ridout M, Demétrio C, Firth D. Estimating intraclass correlation for binary data. *Biometrics.* (1999) 55:137–48. doi: 10.1111/j.0006-341X.1999.00137.x
 36. Searle S. A biometrics invited paper. Topics in variance component estimation. *Biometrics.* (1971) 27:1–76. doi: 10.2307/2528928
 37. Goldstein H, Browne W, Rasbash J. Partitioning variation in multilevel models. *Underst Stat.* (2002) 1:223–32. doi: 10.1207/S15328031US0104_02
 38. Eldridge S, Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research.* London: John Wiley & Sons (2012).
 39. Eldridge S, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol.* (2006) 35:1292–300. doi: 10.1093/ije/dyl129
 40. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *Int J Epidemiol.* (2015) 44:1051–67. doi: 10.1093/ije/dyv113
 41. Arnold B, Hogan D, Colford J, Hubbard A. Simulation methods to estimate design power: an overview for applied research. *BMC Med Res Methodol.* (2011) 11:94. doi: 10.1186/1471-2288-11-94
 42. Schulz K, Altman D, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Br Med J.* (2010) 340:c332. doi: 10.1136/bmj.c332
 43. Stevenson M, Firestone S. Veterinary epidemiology and 'One Health' in Australasia. *Australas Epidemiol.* (2015) 22:35–38.
- Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2021 Stevenson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.