It is essential that veterinary nurses conducting research have a basic knowledge of statistics, thus being able to provide useful and solid analyses that can be read with confidence. Usually the statistical methods to be used are dependent on the study design; depending on the type of study design (as longitudinal or cross-sectional, prospective or retrospective, or matched or unmatched case-control), the statistical methods to be applied for the analysis of data might be different. Therefore, a very detailed plan of the study should be performed before the study starts (Kirkwood and Sterne, 2003).

An essential part of planning any research is to decide how many objects (i.e. animals, people, among others) are going to be studied. Choosing a suitable sample size is a crucial step that must be considered primarily, justifying that the study is capable of answering the question posed and is now a component of research proposals required by most funding agencies (Evans and O’Connor, 2007; Fosgate, 2009; Boyd et al, 2011; Pandis et al, 2011 Ayeni et al, 2012). Sample sizes that are insufficient cannot generate trustworthy answers to the research questions or hypotheses that need to be tested (Fitzner and Heckinger, 2010), while sample sizes that are too large can be a waste of time, money and resources, often raising ethical questions (Noordzij et al, 2011). As such, a suitable sample size uses time and resources in the most gainful manner and is vital to producing valuable research outcomes (Fitzner and Heckinger, 2010; Scott et al, 2011; Scott et al, 2012). Unfortunately, as frequently observed in systematic reviews and meta-analysis studies, sample size calculation is regularly based on assumptions that are recurrently inaccurate, and are inadequately stated and often flawed (Charles et al, 2009). Thus, this paper attempted to highlight this knowledge gap by describing some of the statistical considerations that are appropriate when designing a clinical or an epidemiological research study in veterinary nursing, with a keen focus on sample size calculation.

Calculation of required sample size
To calculate an adequate sample size it is important to quantify the objectives of the study, for example:
- It is necessary to state how much will be considered as the smallest effect or risk between the groups studied
- The significance level, or p-value, that is the strength of the evidence should be decided
- The probability of achieving the significance level, which is called the power of the study should also be stated.

For example, it might be important to decide if it is worthwhile to compare the risk of hypoglycaemia among formula-fed and breast-fed puppies if there is a 90% probability of demonstrating the difference, at a 5% significance, if the true risk was as high as 2; the number of dogs needed would be then calculated. Or if there is a maximum of 300 puppies available, it might be important to calculate the power of the study (strength of the study) that is required to detect a three times higher risk of hypoglycaemia at a 5% significance level.
Sample size calculations can be based on several general methods. Overall, the most often chosen sample size procedures or techniques are based on frequentist statistics (Donner, 1984; Aitken, 1999; Carlin and Doyle, 2002) but these type of calculi can also be found in diagnostic test validation (Alonzo et al, 2002; Georgiadis et al, 2005) and in epidemiologic/surveillance using Bayesian statistics (Suess et al, 2002; Branscum et al, 2006).

**Estimating sample size for clinical studies using formulae**

Equation (1) (Noordzij et al, 2011) is one of the simplest that can be used for calculating sample size in a clinical study using continuous variables.

\[
N = \frac{2[(\alpha + \beta)^2 \sigma^2]}{(\mu_1 - \mu_2)^2} \tag{1}
\]

where, \( N \) is the sample size in each of the groups, \( \mu_1 \) is the population mean in treatment group 1, \( \mu_2 \) is the population mean in treatment group 2, \( \mu_1 - \mu_2 \) is the minimal clinically relevant difference, \( \sigma \) is the population standard deviation, \( \alpha \) is the significance level, and \( \beta \) is the complement of the power of the study (1- power) (Noordzij et al, 2011).

**Minimal clinically relevant difference \((\mu_1 - \mu_2)\)**

The minimal clinically relevant difference \((\mu_1 - \mu_2)\) is the smallest effect between the studied groups that the investigator wants to be able to detect. It is the difference that the investigator believes to be clinically relevant and biologically plausible. The concept of a minimal clinically relevant difference is extremely important and is offered as the new standard for determining effectiveness of a given research hypothesis, such as therapeutic option or the patient prognosis in reference to that treatment. The minimal clinically relevant difference can be a numerical difference (e.g. the choice of a 1°C decrease in pyretic animals under the effect of a given drug) or a binary outcome (yes/no) (e.g. the relevance of adverse effects when given a certain drug, assuming a difference of 20% between the percentage of adverse effects in the treatment group and in the control group as the minimal clinically relevant difference) (Noordzij et al, 2011).

**Standard deviation \(\sigma\)**

Sample size depends on population variance of the result variable. As such, to confirm if the ob-
Power of the study (1-β)

β (beta) is the likelihood of coming to a false-negative conclusion (Type II error), assuming that there is no difference between two groups when in fact there is, thus the opposite from the above mentioned (Type I error). It is most commonly fixed at a level of 0.20, meaning that a researcher desires less than 20% probability of a false-negative inference. The beta value is the complement of the power. If the beta is 0.30, the power is 0.70, meaning there is a 70% probability of detecting a specified effect if this effect actually exists (Kirkwood and Sterne, 2003; Noordzij et al, 2011) (Table 1).

Estimating sample size for epidemiological studies using formulae

If the calculated sample size is smaller than or equal to 5% of the population size, the Equation (2) (Daniel, 1999) can be used:

\[ N = \frac{z^2(1-P)}{d^2} \]  

where \( N \) is the sample size, \( Z \) is the score for a level of confidence, \( P \) is the expected proportion or proportion (in proportion of one; if 30%, \( P = 0.3 \)), and \( d \) is the precision (in proportion of one; if 5%, \( d = 0.05 \)).

If the calculated sample size is larger than 5%, a formula with finite population correction should be used (Daniel, 1999) as follows in Equation (3):

\[ N' = \frac{Nz^2P(1-P)}{d^2(N-1) + z^2P(1-P)} \]  

where \( N' \) is the sample size with finite population correction, \( N \) is the population size, \( Z \) is the score for a level of confidence, \( P \) is the expected proportion and \( d \) is the precision, as explained above (Daniel, 1999).

Z score

The researcher needs to consult a table with the Z score, which depends on the significance level. It is conventional to use a level of confidence of 95%, which corresponds to a Z score of 1.96. If eventually more stringent results are expected and a higher level of confidence is assumed, the Z score increases. For example, if a confidence level of 99% is assumed, the Z score is 2.58 (Daniel, 1999).

Expected prevalence (P)

This value corresponds to what is expected to be achieved by the prevalence study. When first confronted with this variable one might raise the causality dilemma commonly identified as ‘which came first, the chicken or the egg’. If researchers knew the

<table>
<thead>
<tr>
<th>Table 1. Types of error in significance tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conclusion of significance test</strong></td>
</tr>
<tr>
<td><strong>Hypothesis is true</strong></td>
</tr>
<tr>
<td>Reject hypothesis</td>
</tr>
<tr>
<td>Accept hypothesis</td>
</tr>
</tbody>
</table>
Choosing a suitable sample size is a crucial step in a study design. Sample sizes that are too large are often a waste of time, money and resources, often raising ethical issues.

Key Points

- Choosing a suitable sample size is a crucial step in a study design.
- Sample sizes that are insufficient cannot generate trustworthy answers to the research question.
- Sample sizes that are too large often are a waste of time, money and resources, often raising ethical issues.

References


Call for research papers

The Veterinary Nurse provides nurses with an opportunity to document and publish their research. Unsolicited manuscripts are welcome and should be submitted via our online submission site at: http://www.epress.ac.uk/vet/webforms/author.php

Precision ($d$)

The precision ($d$) represents a measure of accuracy by considering the potential data dispersion that one may accept feasible for the study. Considering that the prevalence will be calculated with 95% confidence interval, if the prevalence in a sample is 80% and the 95% confidence interval falls within 75% and 85%, the precision for this estimate is 5% (80% ± 5%) (Daniel, 1999).

Adjustments for loss to follow up, confounding and interaction

It is often the case that it is not possible to include the data of all the recruited sample, because they have either withdrawn from the study, got sick or because some key information is missing. Therefore, the calculated sample size should be increased to allow for some variability on the outcome variable. One must oversample by estimating the percentage of potential drop outs that will not complete the study and then sum it to the sample size (Kirkwood and Sterne, 2003).

Conclusions

When conducting research one must plan its study wisely in order to provide useful and solid data. Sample size calculation is a very important step that needs to be carefully considered in the course of designing a solid research study. VN