The Veterinary Journal 205 (2015) 233–237

Contents lists available at ScienceDirect

The Veterinary Journal

journal homepage: www.elsevier.com/locate/tvjl

Review

Epidemiological and statistical considerations for interpreting and communicating oncology clinical trials

Jane Heller *

School of Animal and Veterinary Sciences, Charles Sturt University, Wagga Wagga 2650 Australia

ARTICLE INFO

Article history:
Accepted 13 February 2015

Keywords:
Oncology
Epidemiology
Randomised controlled trial
Statistics
Risk
Veterinary

ABSTRACT

The use of randomised controlled trials (RCTs) in veterinary oncological research and practice is increasing as is the number of relevant scientific publications. While clear guidelines exist for the reporting of RCTs, a thorough understanding of statistical and epidemiological concepts is required in order to accurately interpret and then impart the results of such trials, and to make balanced decisions regarding the uptake of published findings. This review presents the most important epidemiological and statistical considerations that are needed in order to interpret and communicate with confidence the results of oncology clinical trials.

Introduction

The implementation and reporting of randomised controlled (clinical) trials (RCTs) is imperative in order to progress our knowledge in the field of veterinary oncology. The ability of clinicians not only to interpret the results of RCTs accurately, but also to impart these results to owners, is essential in establishing appropriate, evidence-based therapy recommendations and uptake.

Well-recognised guidelines exist to guide researchers in the reporting of RCTs. These include the CONSORT[1] (Consolidated Standards of Reporting Trials that are published for human RCTs but are applicable to companion animals) and the supplementary REFLECT[2] guidelines (Reporting guidElines For randomized controLled trials for livEstoCk and food safeTy that are designed for trials in livestock and food safety). In addition, a number of efficacy guidelines exist that have been published electronically by the International Conference on Harmonisation (ICH) covering the design, conduct, safety and reporting of clinical trials specifically related to human pharmaceutical product development and registration (Interrnational Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2015). [3]

Statements outlining the CONSORT guidelines were first published in 1996 (Begg et al., 1996) and were subsequently revised in 2001 (Moher et al., 2001) and 2010 (Moher et al., 2010; Schulz et al., 2010), with the supplementary REFLECT guidelines published in 2010 which remain the most recent revision (Sargeant et al, 2010). These guidelines, including a checklist and flow diagram, are intended to "*assist authors in writing reports of randomised controlled trials, editors and peer reviewers in reviewing manuscripts for publication, and readers in critically appraising published articles*" (Schulz et al., 2010).

As this well-published structure is in place to guide the process of reporting and, by extension, to facilitate appropriate methodology, the quality of RCTs that are published should be excellent. However, given that it is explicitly identified within the CONSORT 2010 statement that this statement "*does not include recommendations for designing, conducting, and analysing trials*", one must remember that while the reporting may be transparent the content still requires in-depth critical appraisal.

This review identifies, describes and aims to demystify the most important statistical and epidemiological considerations required to accurately interpret, appraise and communicate the results of oncological RCTs.

Statistical considerations

Sample size and power

One of the first things that should be assessed in the reading of any scientific paper that reports statistical outcomes is the evidence for the sample size used. The reported sample size calculation provides readers with the ability to identify whether the study is indeed capable of detecting associations between the variables of interest (i.e. the study has adequate **power** to identify an association) by the analysis that was performed. In order to calculate an

* Tel.: +61 269332839.
E-mail address: jheller@csu.edu.au.

[1] See: http://www.consort-statement.org/ Accessed 24 January 2015.
[2] See http://www.reflect-statement.org/statement/ Accessed 24 January 2015.
[3] See: Interrnational Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 2015. ICH Guidelines. http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html Accessed 5 January 2015.

http://dx.doi.org/10.1016/j.tvjl.2015.02.014
1090-0233/

Table 1
– Type I and II errors. H_0, null hypothesis.

		Truth	
		Difference between groups	No difference between groups
Conclusion of statistical analysis	Difference between groups (Reject H_0)	Correct	Type I error
	No difference between groups (Do not reject H_0)	Type II error	Correct

appropriate sample size, the authors need to have defined the level of **significance** that they would like to achieve (usually set at $\alpha = 0.05$), the power to detect an effect (usually set at $\beta = 80\%$ or 90%) and, minimally, the effect size that the authors would like to detect (i.e. what is deemed a biologically important difference in the outcome) (Table 1). It is this estimate of effect size that is the critical factor in interpretation of results.

While it is likely that readers who are not trained in statistics will struggle to follow the (sometimes complex) calculations that have been performed, the assumptions that have been made in relation to sample size calculations (including the provision for increased sample size in order to allow multivariable rather than univariable analysis, considered further below) should be evident and transparent. If a study has a limited sample size, the results require interpretation in light of this limitation. It should be specifically noted that a small sample size will directly affect the ability to detect a difference and result in an increase in Type II error (Table 1). Therefore, if a statistically significant difference is not identified it may be solely due to the **lack of power**, rather than a true lack of difference or association. It serves well in these circumstances to recall that "*absence of evidence is not evidence of absence*" (Altman and Bland, 1995; Alderson, 2004) and, with respect to this statement, any 'negative' finding in the literature should not be interpreted as evidence of lack of clinical importance.

As an example of difficulties in interpreting non-significant findings, if a sample size has been calculated based on the ability to detect a difference of 10% mortality between treatment and control groups and no significant difference is reported (i.e. $P > 0.05$), one cannot surmise that a difference of 9.9% or less does not exist. In addition, if the result is reported with a wide confidence interval, the marked uncertainty that is associated with the estimate is not imparted through the simple reporting of a 'negative finding', or lack of significance and, indeed, the discussion of such a result without taking into account such uncertainty will be very misleading.

Identification of appropriate outcomes/endpoints

While it appears obvious that the outcome or endpoint of interest should be appropriately identified, the simplicity of this statement is not always evident in practice. The outcome of interest must reflect clinical relevance and the outcomes to be examined should be identified in the study hypothesis before any results are presented.

In oncology, endpoints may be related to survival time, tumour symptoms or measures of welfare (Pilz et al., 2012). The most common endpoints related to survival time are overall survival (OS), progression free survival (PFS) and disease-free interval (DFI), which encompass the time between diagnosis or recruitment until: (1) death from any cause (OS); (2) death from any cause OR progression/relapse of the tumour (PFS), or (3) relapse/recurrence of the tumour (DFI), respectively. These outcomes include a time component and are therefore denoted as 'time-to-event' measures, requiring the use of a survival analysis in order to estimate the risk of the outcome. Survival analyses account for data collected from animals that have not experienced the events during the time under study, through the use of censoring (see below).

Endpoints may also be set to reflect tumour symptoms, such as partial or complete remission over a specified period of time, and may be represented as dichotomous outcomes and analysed using standard techniques for comparison of proportions. Other measures such as pain or weight loss over or at a specified time after treatment may also be used to measure symptom-related outcomes and analyses will depend on the nature of the data generated (i.e. continuous or categorical). Finally, owner-reported quality of life measures may also be used.

The sample size calculation must be appropriate for the endpoint selected and the estimation of a minimal effect size for detection should also be relevant to that measure. Consideration of issues that might be associated with differing outcome measures need to be appraised on an individual study/neoplastic process basis.

Censoring

Use of censoring is imperative for accurate and appropriate interpretation of studies where the outcome is survival or time-to-event. Censoring allows for the fact that some animals may not have experienced the outcome of interest (e.g. death or relapse) during the period under study. It allows for the inclusion of animals that are lost to follow-up, have died from alternate causes or are still alive at the end of the study (Clark et al., 2003). Criteria for censoring should be rigorously identified in the reported work and, if censoring is not present, results should be interpreted cautiously. It should also be acknowledged that censoring is uninformative, i.e. those animals which are censored should have the same likelihood of undergoing a subsequent event as those remaining in the study (Clark et al., 2003).

Univariable and multivariable statistical analyses

Univariable statistical analysis, where the association between a single explanatory variable of interest and outcome is explored, is an important first step in any analytical process. However, progression to the use of multivariable statistical analyses is required in order to estimate the effect of the intervention that has been trialled in the presence of other explanatory variables that may also affect the outcome. The type of analysis that is performed is dependent on the type of data that is collected for the outcome (or endpoint) of interest. As a general rule, if time-to-event data are collected, a survival analysis will be performed (e.g. Cox proportional hazards regression). In contrast, for outcome measures that are dichotomous (e.g. remission either occurs or does not occur), a logistic regression will be performed and for continuous data (e.g. weight loss in kg), linear regression should be used. Mixed models may also be used for data with either continuous or categorical outcomes and these allow for the additional inclusion of random effects that can account for repeated measures, clustering or hierarchical structure within a study (Dohoo et al., 2009).

Multivariable modelling is often used to control for confounding as it allows for the effect of the variable of interest (in the case of RCTs, this is the intervention) to be estimated while all other factors (including confounders) that may affect the outcome are held constant. The term 'confounding' refers to the mixing together of the effects of multiple variables (Dohoo et al., 2009). If one is trying to estimate the magnitude of effect of an intervention, then it is important to be able to attribute the estimated effect to the intervention and not in part to an unidentified confounding variable. Confounders have an effect on both the exposure (in this case intervention) and outcome of interest, and failing to adjust for this effect (which can be done by inclusion of confounding variables in a multivariable