

Evidence-Based Medicine: The Design and Interpretation of Noninferiority Clinical Trials in Veterinary Medicine

K.J. Freise, T.-L. Lin, T.M. Fan, V. Recta, and T.P. Clark

Noninferiority trials are clinical studies designed to demonstrate that an investigational drug is at least as effective as an established treatment within a predetermined margin. They are conducted, in part, because of ethical concerns of administering a placebo to veterinary patients when an established effective treatment exists. The use of noninferiority trial designs has become more common in veterinary medicine with the increasing number of established veterinary therapeutics and the desire to eliminate potential pain or distress in a placebo-controlled study. Selecting the appropriate active control and an *a priori* noninferiority margin between the investigational and active control drug are unique and critical design factors for noninferiority studies. Without reliable historical knowledge of the disease response in the absence of treatment and of the response to the selected active control drug, proper design and interpretation of a noninferiority trial is not possible. Despite the appeal of conducting noninferiority trials to eliminate ethical concerns of placebo-controlled studies, there are real limitations and possible ethical conundrums associated with noninferiority trials. The consequences of incorrect study conclusions because of poor noninferiority trial design need careful attention. Alternative trial designs to typical noninferiority studies exist, but these too have limitations and must also be carefully considered.

Key words: Ethics; Placebo; Positive control; Statistical methods.

Clinical trials conducted in subjects with naturally occurring disease are not only considered the gold standard to demonstrate that drugs are safe and effective but also are required to achieve drug approval by both the Food and Drug Administration (FDA) and European Medicines Agency (EMA). In the broadest sense, there are 3 types of trial designs: superiority, equivalence, and noninferiority studies. Superiority studies seek to establish effectiveness by demonstrating an outcome superior to a control product, typically a placebo. Equivalence studies are designed to confirm that there is no clinically relevant difference between treatments. Noninferiority studies are designed to demonstrate that an investigational drug is at least as effective, within an acceptably small margin, as an established, well-studied treatment.

There are several medical, legal, statistical, and ethical factors that influence design selection. A placebo-controlled superiority trial often is pursued when there is no well-studied standard of care. Moreover, when a new drug is being developed for regulatory approval, a placebo control generally is considered ideal unless an existing standard of care, used as an active control, is approved for the species and indication for which the investigational drug is being examined. However, active controlled superiority trials generally are not pursued because it often is not logical that a new treatment will be superior to an old treatment or the

Abbreviations:

CI	confidence interval
EMA	European Medicines Agency
FDA	Food and Drug Administration
H ₀	null hypothesis
H _A	alternative hypothesis
ITT	intent-to-treat
NSAIDS	nonsteroidal anti-inflammatory drugs
PP	per-protocol

expected improvement in treatment is small and thus an appropriately powered study would be prohibitively large. Equivalence trials are designed to demonstrate that the effectiveness of an investigational drug is similar to that of an active control, within a predetermined *indifference zone*,¹ and most commonly are used for pharmacokinetic bioequivalence determination. Noninferiority trials sometimes are conducted solely because of the ethical concerns of administering a placebo in a study when an established effective treatment exists.^{2,3} Another reason to conduct a noninferiority trial is to determine the comparative safety and effectiveness of 2 products.² Even though an effective treatment already exists, a new treatment that is no better than the current treatment may be desirable because of improved safety, decreased cost, improved compliance, or ease of administration.⁴

There have been a number of recently FDA- or EMA-approved drugs for which sponsors have conducted noninferiority trials to demonstrate safety and effectiveness (Table 1). The first noninferiority studies in veterinary medicine were not reported until 2003, and have become more common in the peer-reviewed literature over the past few years. This trend toward noninferiority trials in veterinary medicine may be caused by ethical concerns with placebo-controlled studies. Informed consent cannot be gained from the subject under study (eg, the dog), nor can the subject choose to withdraw of its own will, and therefore trial

From the Nexcyon Pharmaceuticals Inc, Madison, WI (Freise, Lin, Clark); the University of Illinois College of Veterinary Medicine, Urbana, IL (Fan); and the US Food and Drug Administration (FDA) Center for Veterinary Medicine, Rockville, MD (Recta).

Corresponding author: T. Clark, Nexcyon Pharmaceuticals Inc, 644 West Washington Avenue, Madison, WI 53703; e-mail: clarktp@nexcyon.com

Submitted May 6, 2013; Revised August 8, 2013; Accepted August 27, 2013.

Copyright © 2013 by the American College of Veterinary Internal Medicine

10.1111/jvim.12211

Table 1. Recent noninferiority studies in veterinary medicine.^a

Year of Publication	Species	Therapeutic Class	Primary Endpoint	<i>A Priori</i> Selected Margin (Justification)	Noninferiority Margin	Comments
2003 ³²	Pig	Anti-inflammatory	Clinical index score	Yes (no justification)	1.4 score point	
2004 ³³	Cat	Analgesic	Pain intervention rate	No	NA	
2004 ³⁴	Dog	Antibiotic	Cure rate	Yes (no justification)	15%	
2005 ³⁵	Dog	Parasiticide	Cure rate	Yes (clinical judgment)	20%	
2005 ³⁶	Dog	Parasiticide	Cure rate	Yes (clinical judgment)	20%	
2005 ³⁷	Dog	Vaccine	Geometric mean virus titers	No	NA	Used superiority analysis and concluded noninferiority based on no significant difference
2005 ³⁸	Horse	Analgesic	Clinical improvement	Yes (no justification)	15%	
2006 ^{b,39,40}	Dog	Analgesic	Incidence of clinical improvement	Yes (no justification)	15%	
2007 ⁴¹	Cat	Anthelmintic	Success rate	Yes (no justification)	15%	
2007 ⁴²	Cat	Sedative and Analgesic	Success rate	Yes (no justification)	13%	
2007 ⁴³	Cattle	Disinfectant	Infection rate	No, because of lack of historical study	NA	
2007 ⁴⁴	Dog	Antibiotic	Success rate	Yes (clinical judgment)	15%	
2007 ⁴⁵	Dog	Antihypertensive	Success rate	Yes (no justification)	20%	
2007 ⁴⁶	Dog	Inotrope and Vasodilator	Success rate	Yes (no justification)	Not reported	
2008 ⁴⁷	Cat	Antibiotic	Success rate	Yes (clinical judgment)	15%	
2008 ⁴⁸	Cat	Antibiotic	Success rate	Yes (no justification)	15%	
2008 ⁴⁸	Dog	Antibiotic	Success rate	Yes (no justification)	15%	
2009 ⁴⁹	Cattle	Antibiotic	Success rate	Yes (no justification)	15%	
2009 ⁵⁰	Dog	Anthelmintic	Log transformed egg count	Yes (no justification)	.38	
2009 ⁵¹	Dog	Antibiotic and Antifungal	Improvement rate	Yes (no justification)	10%	
2010 ⁵²	Cat	Analgesic	Assessment score ratio	Yes (no justification)	.2	
2010 ⁵³	Cattle	Antibiotic	New infection rate and cure rate	Yes (clinical judgment)	3 and 10%	
2010 ⁵⁴	Dog	Antifungal	Yeast count and score	Yes (clinical judgment)	20% of control	
2011 ⁵⁵	Cat	Antianginal	Cardiac functions	Yes (clinical and statistical judgments based on random effects)	50% of control	
2011 ²⁵	Dog	Analgesic	Glasgow Composite Pain Scale Ratio	Yes (no justification)	.2	
2011 ⁵⁶	Dog	Anthelmintic	Log transformed egg count	Yes (no justification)	.2	
2012 ⁵⁷	Cat	Analgesic	Total clinician score ratio	Yes (based on a previous human study)	.25	
2012 ⁵⁸	Cat	Analgesic	Total clinician score ratio	Yes (based on a previous human study)	.25	
2012 ⁵⁹	Dog	Analgesic	Failure rate	Yes (no justification)	15%	
2012 ⁶⁰	Dog	Analgesic	Failure rate	Yes (based on a previous study of different control)	15%	
2012 ⁶¹	Dog	Analgesic	Global functional disability score ratio	Yes (based on a previous human study)	.25	Lower CI bound of ratio <.75 (point estimate 1.244), but noninferiority still concluded
2012 ⁶²	Dog	Analgesic	Rating scale ratio	Yes (no justification)	.2	
2012 ⁶³	Dog	Antibiotic	Clinical cure rate	Yes (no justification)	20%	

Table 1. Continued

Year of Publication	Species	Therapeutic Class	Primary Endpoint	<i>A Priori</i> Selected Margin (Justification)	Noninferiority Margin	Comments
2013 ⁶⁴	Cattle	Antibiotic	Cure rate	Yes (no justification)	10%	
2013 ⁶⁵	Dog	Analgesic	Glasgow Composite Pain Scale Ratio	Yes (no justification)	.2	
2013 ⁶⁶	Dog	Analgesic	Success rate	Yes (clinical judgment)	20%	

NA, not applicable; CI, confidence interval.

^aIncludes published, peer-reviewed studies, and studies reported in FDA freedom of information summaries.

^bPeer-reviewed publication year.

designs often err on the side of caution. In the sense that informed consent is not obtained from the patient but from a legal guardian or owner, clinical trials in veterinary medicine share a unique commonality with pediatric trials in human medicine. Conducting negatively controlled clinical trials of analgesics in dogs and cats is especially difficult, because of ethical concerns of knowingly withholding pain medication coupled with the greater availability of analgesics in these species,⁵ and in fact analgesic studies in dogs and cats were the most commonly identified veterinary noninferiority trial (Table 1). Also, similar to the situation in human pediatric medicine, there are many major diseases in veterinary medicine for which a standard of care exists, but no effect size has been established compared with a placebo. As discussed below, without appropriate placebo-controlled studies, there is insufficient information to justify the use of the established treatment as an active control in a noninferiority trial.³

Although there are many study design differences between noninferiority and “classical” negatively controlled superiority trials, the most important of these differences in a noninferiority study are as follows: (1) switching the hypotheses; (2) selecting the appropriate active control; (3) determining the noninferiority margin between the active control and investigational drug; and (4) the appropriate interpretation of the study results. In human medicine, there are a number of national and international regulatory guidance documents on the appropriate conduct of noninferiority trials^{2,6,7} and just recently, the FDA released guidance for veterinary medicine.⁸ Despite this growing body of studies and guidance, most documents fail to provide any specific guidelines on how to choose the active control and how to determine the noninferiority margin.^{2,6,9} Because of the increased use of noninferiority trials in veterinary medicine and the unique design features of these trials, the objective of this review is to outline the key design aspects and proper interpretation of noninferiority trials.

Design of Noninferiority Trials

Switching the Hypotheses

In a superiority trial, the null (H_0) and alternative (H_A) hypotheses about the effect of a new test

treatment (μ_T) relative to the active control treatment (μ_A) are the familiar:

$$H_0 : \mu_T - \mu_A = 0 \quad H_A : \mu_T - \mu_A \neq 0 \quad (1)$$

If H_0 is rejected based on the study results, it is concluded that treatment has a significant effect. However, if the study fails to reject H_0 , the conclusion is not that the treatments are equal, but that there is insufficient evidence to conclude that H_0 is not true. In other words, failing to reject H_0 does *not* result in the conclusion that H_0 is true.¹⁰ If it were, one could conclude that 2 treatments are the same by failing to reject the null hypothesis, an erroneous conclusion reached in some small studies in veterinary medicine. To test if there is no difference in the treatments, H_0 and H_A must be reversed to:

$$H_0 : \mu_T - \mu_A \neq 0 \quad H_A : \mu_T - \mu_A = 0 \quad (2)$$

With the hypotheses reversed, rejection of the null hypothesis results in the conclusion that there is no difference between the treatments. However, demonstrating that the 2 treatments are identical by showing the difference in their outcome effect is zero is not possible nor is it usually of clinical interest. What is usually of interest is determining if the difference in treatments is not too much different from zero. More relevant to a noninferiority trial, the clinical interest is determining that the test treatment is at least as effective as the active control treatment. That is, the investigational drug is considered noninferior to an active control within a predetermined acceptable difference. This allowed difference between active control and test responses is referred to as noninferiority margin or margin of indifference (δ). In the case where a larger effect indicates better effectiveness of a treatment (eg, cure rate), the noninferiority trial null and alternative hypotheses become:

$$H_0 : \mu_T - \mu_A \leq -\delta \quad H_A : \mu_T - \mu_A > -\delta, \delta > 0 \quad (3)$$

And in the case where a smaller effect indicates better effectiveness (eg, mortality rate), the hypotheses become:

$$H_0 : \mu_T - \mu_A \geq \delta \quad H_A : \mu_T - \mu_A < \delta, \delta > 0 \quad (4)$$

In either case, if H_0 is rejected based on the study results, it is concluded the test treatment is noninferior to the active control treatment.

Another common set-up of hypotheses for noninferiority trials is the comparison of ratios (eg, risk, odds, or hazard). For simplicity of this review, only absolute treatment differences will be discussed. However, the design and interpretation of noninferiority trials with the hypotheses set up as ratios can be logically extended from the discussion below.

Type I and Type II Errors

Although the hypotheses are reversed, as described above, the definitions of Type I and Type II errors remain the same. Only the interpretations are different from those under the conventional hypotheses (Equation 1). Under the noninferiority hypotheses, the Type I error is to wrongly conclude that the investigational drug is noninferior to the active control, whereas it is actually inferior to the active control. Conversely, the Type II error is to wrongly conclude that the investigational drug is inferior to the active control, whereas it actually is noninferior.

Sample Size Considerations

It is common that the sample size of a noninferiority trial should be of the same order of magnitude as that of a superiority trial.¹¹ If the sample size calculation in a noninferiority trial is performed under the alternative hypothesis that the investigational drug is as effective as the active control (ie, $\mu_T - \mu_A = 0$), then the noninferiority margin is analogous for powering purposes to the assumed difference between treatments under the alternative hypothesis in a superiority trial. Thus, the smaller the margin is, the larger the sample size that is needed. Nevertheless, it has been shown that the sample size may greatly increase if the investigational drug is actually slightly less effective than the active control.¹⁰

Assay Sensitivity

An essential property of a noninferiority trial is assay sensitivity.^{2,6-9} Assay sensitivity is the ability of a trial to distinguish an effective treatment from a less effective or ineffective treatment. A successful superiority trial, in which the new treatment is determined to be significantly better than the negative or positive control, has inherent assay sensitivity. In contrast, if the hypothesis test in a superiority trial fails to reject the null hypothesis of equal treatment effects, it could be because the new treatment is truly ineffective or because the trial lacked assay sensitivity. However, even if a noninferiority trial is successful and the new treatment is determined to be noninferior to the active control, it could be that (1) both treatments are similarly effective, (2) both treatments are similarly ineffec-

tive, or (3) the trial lacks assay sensitivity. In other words, the assay sensitivity of the trial is unknown based on the results of the study alone. In a noninferiority study, the assay sensitivity of the trial is assessed from historical evidence of sensitivity of a trial design to the active control treatment's effects, based on 1 or more well-designed, previously conducted study of the active control compared with a placebo. In the absence of a placebo-controlled superiority trial with sensitivity, at minimum, there needs to be rigorous historical knowledge of the course of the disease in the absence of treatment and (separate) knowledge of the disease response when the active control is administered. Primarily because of the elimination of the concern of assay sensitivity within the trial, placebo-controlled superiority studies generally are preferred over noninferiority studies when ethically appropriate because they more conclusively and less subjectively demonstrate a treatment's effectiveness.¹²

There are several particularly important design features that must be followed to gain confidence that the planned noninferiority trial maintains assay sensitivity. The first is that the trial must utilize a similar treatment regimen as the prior trial(s) of the active control against the placebo. For example, data from a single-dose study of the active control would not be supportive of a planned noninferiority trial examining multiple doses over time. Second, the planned noninferiority trial also must study a similar patient population as the placebo-controlled trial of the active control. If the active control was studied in a population of patients with mild disease, this would not be supportive in the planned noninferiority study in a population of patients with severe disease (Fig 1). Third, the selected primary outcome measure and the time of evaluation in the noninferiority trial must be an endpoint measured and reported in a placebo-controlled trial of the selected active control because this is the outcome that has been shown to demonstrate sensitivity to treatment effects. For example, if the active control effect size was measured with a visual analog scale, it would be problematic to assume that a noninferiority study using a Glasgow-modified pain scale (Fig 2) would be equally sensitive.

The study quality of the trial(s) conducted with the selected active control compared with the placebo is also important. A high-quality, well-described trial demonstrating the effectiveness of the active control ensures a high degree of confidence that the effect is reproducible (ie, assay sensitivity) and that trial design can be recreated. If a single low-quality study of the active control is all that exists, the strength of evidence from the noninferiority trial will be undermined, no matter how well the noninferiority trial is conducted.

Constancy Assumption

Implicit in the choice of active control treatment is the assumption that the difference between the active control and the placebo is constant from study to study (Fig 3). This constancy assumption is needed, so

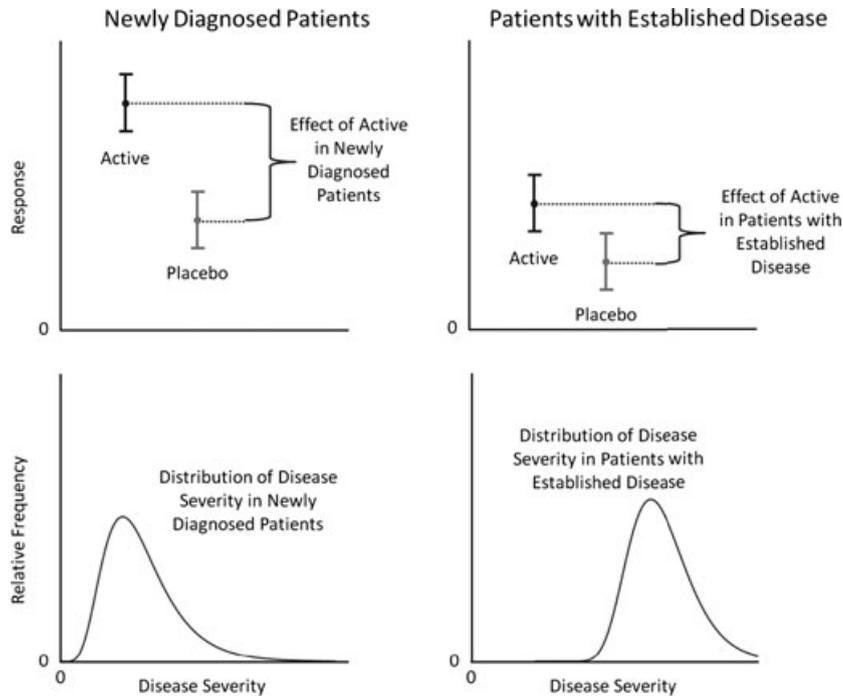


Fig 1. The effect that changes in disease severity enrollment criteria (newly diagnosed versus established disease) may have on the response of the active control. If the distribution of the disease shifts to greater severity with different enrollment criteria, the effectiveness of the active control compared with the placebo may be reduced.

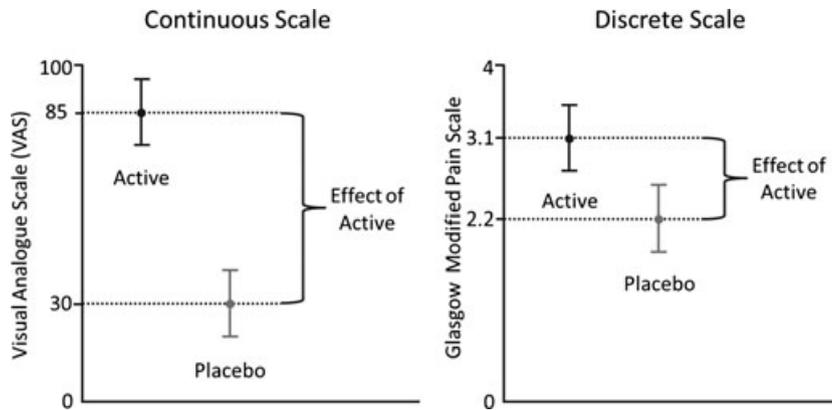


Fig 2. The impact that changing the clinical endpoint from one scale to another may have on the effect of the active control. The differences between the active control and placebo cannot be transformed from the visual analog scale to the Glasgow-Modified Pain Scale.

that the appropriate noninferiority margins can be determined from the historical placebo-controlled trials. The factors that are favorable to assay sensitivity, as discussed in the previous section, also support the constancy assumption. In addition, the conditions under which the active control's effect was evaluated should be relatively recent to ensure that clinical practice has not changed substantially since the placebo-controlled trial(s) was(were) conducted. If clinical practice has not changed substantially, it is expected that the effect of the active control compared with a placebo will be consistent with historical data. The constancy assumption may be problematic in the case

of subjective and highly variable responses,⁸ such as animal pain or behavioral endpoints that result in subjective clinical scores. If the effect of the active control is not relatively constant and consistent from study to study, it cannot be used for a noninferiority study because of the inability to establish a reliable effect of the active control and thus ascertain the assay sensitivity of any trial conducted. In the case where the effect size can be inconsistent attributable to a subjective and variable clinical endpoint, the disease state itself may not lend itself to a noninferiority study. There is a risk that a new drug could be demonstrated as noninferior to an active drug that is ineffective in the clinical trial

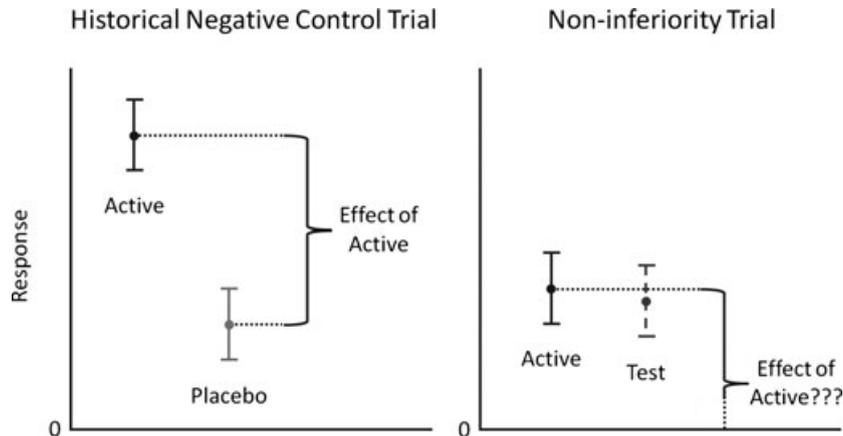


Fig 3. The potential impact if the constancy assumption is violated. As the effect of the active control is less in the noninferiority trial than in the historical negative control trial used for trial design, the effect size of the active control is unknown and the test drug may not be effective at all, even if noninferiority is concluded.

and lead to the incorrect conclusion that the new drug is effective. Wide clinical use of an ineffective treatment introduces a much broader ethical dilemma than conducting a single placebo-controlled study. For this reason, regulatory bodies have required placebo-controlled, superiority trials to increase the degree of confidence in the decision about the effectiveness of treatment.

Determining the Noninferiority Margin

A key consideration in the design of noninferiority trials is to select *a priori* the noninferiority margin (δ).³ The noninferiority margin is the maximum difference in treatment effectiveness that can be allowed. The choice of δ should be based both on statistical reasoning and clinical judgment, should reflect uncertainties in the evidence on which the choice is based, and should be suitably conservative.² Furthermore, the margin cannot be greater than the smallest effect size that the active control would be reliably expected to have compared with a placebo. The process of selecting a margin can be broken down logically into 2 steps. The first step uses statistical reasoning to determine the smallest reliable effect size of the active control compared with placebo, and the second step uses clinical judgment to determine the largest loss of effectiveness of the new treatment that would be considered clinically insignificant while minimally maintaining the smallest effect size, resulting in a single clinically and statistically relevant noninferiority margin.

Smallest Reliable Effect Size. Determining the smallest reliable effect size is the more objective step of determining the noninferiority margin, although some subjectivity still is incorporated into the process. The smallest reliable effect size, denoted here as δ_0 , also is referred to as margin 1.^{7,8} It is determined based on prior experience with the active control compared with placebo or no treatment. The smallest reliable effect size is considered to represent the whole effect of the

active control treatment and it is critical because if the difference between the active control and the new treatment is greater than the effect size, the new treatment may have no effect at all.

Many different approaches have been proposed for establishing a margin based on statistical reasoning,^{4,7,13,14} but there is no single accepted method. All methods involve using the results of 1 or more placebo-controlled studies of the active treatment to determine the difference between the active control and placebo. The most straightforward and probably most commonly used procedure is to calculate the lower bound of the 2-sided 90% or 95% confidence interval (CI) of the absolute value of the mean difference between the active drug and placebo. The CI can be constructed from a single well-designed study, or if multiple placebo-controlled studies of the active treatment have been conducted, it can be constructed using a random effects model approach. The use of the lower bound of the 95% CI conservatively incorporates the uncertainty of effect of the active treatment, thus ensuring with a high degree of confidence that if the magnitude of the difference in treatments is less than this value, the new treatment has an effect relative to the placebo.

Clinical Judgment. To determine the noninferiority margin (δ), clinical judgment as well as statistical reasoning needs to be incorporated into its selection. This margin incorporating clinical judgment also is sometimes referred to as margin 2.^{7,8} Incorporating clinical judgment into δ is subjective and little has been published on a recommended procedure. The clinically relevant margin of indifference often is described as determining the largest loss of effectiveness of the new treatment that would be considered clinically insignificant. Obviously, the selected value may vary substantially among individuals, and may likely vary within an individual on a case-by-case basis. For example, consider an active control treatment that has adverse effects that cannot be tolerated by a particular individual. This patient would consider a much larger

loss of effectiveness as clinically insignificant compared with a patient that tolerates the active control treatment well. Furthermore, the severity of the disease state being treated strongly determines the magnitude of δ . To ensure that the new treatment has some effect, the selected δ cannot be greater than the smallest reliable effects size (ie, $\delta \leq \delta_0$). Therefore, δ_0 must be determined before selecting δ . Ensuring that δ is $\leq \delta_0$ also allows for straightforward incorporation of statistical reasoning into the selection of the noninferiority margin. Probably the most common procedure for selecting δ is to halve δ_0 to maintain at least half of the clinical effect of the active control.^{7,8} An alternative, less conservative, approach is to halve the point estimate of the mean difference between the active control and placebo, and then select the smaller of this value and δ_0 to ensure that δ is $\leq \delta_0$. Another common “rule-of-thumb” for selecting δ is to select the smaller of the δ_0 and 1/2 an SD of the response, as differences in less than that magnitude generally are not considered to be clinically relevant.⁹ However, in reality, there is no “one-size-fits-all” approach to selecting δ because of inherent subjectivity in using clinical judgment.

Interpreting Trial Results

Statistical Analysis

The statistical analysis for a noninferiority trial typically proceeds by constructing a 2-sided 95% CI of the difference between the effects of the active control and the test treatment (as given by Equations 3 or 4). If a smaller effect indicates better effectiveness, then the upper limit of the CI is compared with the *a priori* selected noninferiority margin (Fig 4). When this upper limit is less than δ (Fig 4, Case A), the null hypothesis given by Equation 4 is rejected and it is concluded that the new treatment is noninferior to the active control. If instead, the upper limit of the CI is $\geq \delta$ (Fig 4, Case B and C), there is insufficient evidence to reject the null hypothesis and demonstration of

noninferiority of the new treatment to the active control is not achieved. Figure 4 shows different scenarios when a smaller effect indicates better effectiveness and Equation 4 is considered. Similar interpretations can be made when a larger effect indicates better effectiveness and Equation 3 is considered by using the lower CI and comparing it to $-\delta$. For example, when a larger effect indicates better effectiveness and lower limit of the CI is $>-\delta$, the null hypothesis given by Equation 3 is rejected and noninferiority is concluded.

Although for both case B and C in Figure 4, noninferiority of the new treatment to the active is not demonstrated, the interpretation of the results is somewhat different. In Case B, the upper 95% CI of the difference is $>\delta$ but $<\delta_0$; so even though the new treatment has not been demonstrated to be noninferior to the active control, it still has an effect greater than the smallest reliable effect of the active control (ie, δ_0), indicating that the new treatment is effective. In other words, there is statistically significant (indirect) evidence that the new treatment does have some effect (ie, compared with a placebo or negative control), even if it has not been demonstrated to be as effective as the active control by an acceptable margin. Thus, if the upper CI is only slightly $>\delta$ but still $<\delta_0$, some regulatory bodies are willing to consider a new treatment effective and approve the product,⁷ especially if it has other important attributes such as improved dosing convenience, compliance, or a better safety profile. The objective of the noninferiority trial also may only be to demonstrate that the new treatment has some effect through “indirect superiority” to a placebo.^{6,7} In that case, this objective should be prospectively stated in the protocol and only δ_0 would need to be specified. In contrast to case B in Figure 4, for case C, the upper 95% CI of the difference is greater than both δ and δ_0 , thus not only is there insufficient evidence to demonstrate that the new treatment is noninferior to the active control but also there is insufficient evidence that the new treatment has any effect at all.

In the case in which the null hypothesis given by Equations 3 or 4 is rejected and it is concluded that

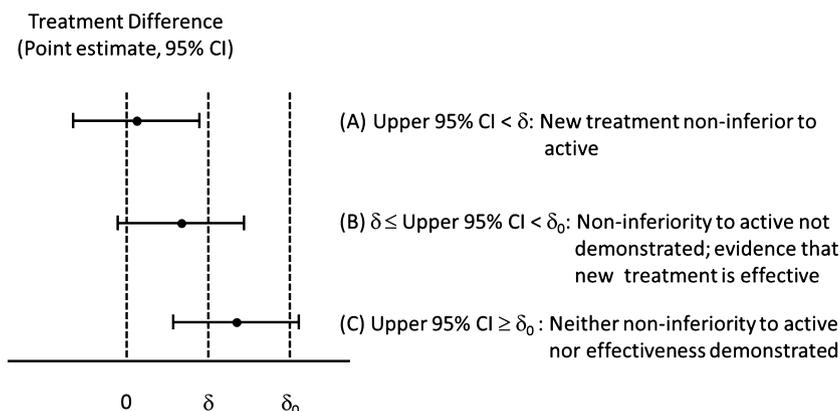


Fig 4. Three different noninferiority trial results (Case A, B, and C) when a smaller effect indicates better effectiveness and the subsequent study interpretations⁷; δ , noninferiority margin; δ_0 , smallest reliable effect size.

the new treatment is noninferior to the active control, a second alternative hypothesis testing of superiority of the new treatment to the active control can be conducted.^{7,15,16} This second alternative hypothesis for testing superiority is $\mu_T - \mu_A > 0$ if a larger effect indicates better effectiveness, or $\mu_T - \mu_A < 0$ if a smaller effect indicates better effectiveness. Then, the lower or upper 95% CI is compared to zero, depending on whether a larger or a smaller effect is of interest, respectively. The conclusion that the new treatment is superior to the active control can be made in the same fashion as described previously for noninferiority hypotheses. This can be conducted posthoc with no inflation in the type I error rate because the null hypotheses of the 2 tests (ie, $\mu_T - \mu_A \geq -\delta$, $\delta > 0$ and $\mu_T - \mu_A \geq 0$) form a nested or closed set.^{15,17} Although the simultaneous testing of noninferiority and superiority in an active control trial is a theoretically sound approach from a statistical perspective, the implementation of this approach in practice is uncommon. Because of the subjective nature of a noninferiority trial and the noninferiority margin, if a study is designed as an active control superiority trial, one cannot test for noninferiority posthoc after failing to demonstrate superiority.^{7,12} In other words, the noninferiority test and noninferiority margin must be prespecified in the protocol.

Intent-To-Treat Principle

For all randomized controlled trials, the intent-to-treat (ITT) principle may be the most fundamental principle that ensures unbiased hypothesis testing.¹⁸ In superiority trials, this principle can lead to lack of power attributable to significant nonadherence to the assigned treatment, but it also prevents false-positive conclusions and preserves Type I error. In noninferiority trials, however, this principle may provide an advantage to the test drug because the effect of nonadherence often dilutes the difference between the test drug and the active control, making them look more similar. It thus can increase Type I error in such cases. Although the fact that the active control is assumed to be effective alters the impact of nonadherence, it is especially important to design and conduct noninferiority trials in a manner to decrease the occurrence of nonadherence.

On the other hand, the per-protocol (PP) analysis includes all patients who completed the full course of assigned treatment and who had no major protocol violations and may be preferable in noninferiority trials.³ Therefore, in a noninferiority trial, the full analysis set based on both the ITT and PP analysis has equal importance and their use should lead to the same conclusion for a robust study interpretation.¹⁹

Challenges to Noninferiority Trials in Veterinary Medicine

The validity and interpretability of a noninferiority trial depends largely on being able to make assumptions such as assay sensitivity, constancy of effect, and

acceptability of the noninferiority margin. These assumptions cannot be verified in the trial itself, but can only be inferred based on historical evidence from previous studies and how closely the study design features adhered to previous studies. Designing appropriate veterinary noninferiority trials is especially challenging because of several factors, including (1) the lack of historical placebo-controlled studies; (2) changing study designs or conditions, such as patient population or standard of care; and (3) subjective and highly variable clinical endpoints.

A major constraint in designing noninferiority trials for veterinary medicine is that potential active controls may be considered effective, but have not been directly compared with placebo in 1 or more well-designed, randomized clinical trials. The active control may be the standard of care, it may be a drug approved in humans, or it may have been demonstrated to have an effect in the target species that is closely related to the clinical condition in a small laboratory or clinical study. The problem with using active controls such as these is not just the question of whether or not they are effective, it is that without a well-designed, placebo-controlled, randomized clinical trial, the magnitude of benefit is unknown and therefore a margin of non-inferiority cannot be determined *a priori*. As discussed above, the noninferiority margin cannot be greater than the smallest reliable effect size, δ_0 , to ensure that the new treatment has at least some effect, but δ_0 cannot be determined without at least 1 well-designed, placebo-controlled study of the active control drug. An alternative source of information on δ_0 may be good historical knowledge of the clinical response in the absence of treatment. This is rarely the case, however, because current clinical practice would not have allowed a serious disease or condition to be left completely untreated. Without the incorporation of smallest reliable effect size into the selection of the margin, it is quite possible to conclude that an ineffective new treatment is noninferior to an active control, regardless of the efficacy of the active control. Obviously, if the active control is not actually effective (because there is no well-designed, placebo-controlled study, it is quite possible), establishing noninferiority of the new treatment is almost a certainty.

Even in cases in which an active control has demonstrated effectiveness against placebo in a randomized clinical trial, departures from the original design may hinder the determination of a robust effect size (Figs 1 and 2) and undermine the assumption of assay sensitivity. Trial design changes that can have unintended consequences include, but are not limited to, changes in the enrollment criteria (eg, disease severity, age), clinical endpoint (eg, continuous response versus binary treatment failure), and treatment duration. In these cases, it will be difficult to determine the appropriate margin based on historical information, resulting in an overreliance on clinical judgment in its selection. In addition, there is no assurance that the modified design will still maintain assay sensitivity.

Lastly, one of the constraints in veterinary clinical trials is the difficulty of establishing valid and repeat-

able measures of clinical outcomes. In most cases, the outcome of treatment is subjectively measured and may differ from 1 study to another. The constancy assumption may not hold in the case of subjective and changing endpoints because the magnitude of the effect may change in unexpected ways. Even when the null hypothesis is rejected (Equations 3 or 4) and noninferiority is concluded based on statistical analysis, conclusively stating that the new treatment is effective can be problematic. For example, if the effect of the active control in the noninferiority trial turns out to be substantially less than the effect of active control in the placebo-controlled study(ies) used to establish the smallest reliable effect size (δ_0), doubts can occur about both the appropriateness of the *a priori* selected noninferiority margin and the weight of evidence that the new treatment has any effect at all (Fig 3). It is possible that the δ_0 based on previous trials may have been an overestimate of the true reliable effect size and thus the selected noninferiority margin δ is too wide and cannot be used to rule out inferiority or lack of any effect at all.

Unfortunately, none of the identified noninferiority trials conducted in veterinary medicine justified the selected noninferiority margin adequately (Table 1), particularly with regard to establishing the smallest reliable effect size (δ_0). Some of these studies may have selected the noninferiority margin appropriately, but just failed to report the justification. In general, noninferiority studies used either differences in success rates or a ratio of a score/scale as the primary endpoint. Noninferiority margins generally were declared *a priori*, with a margin of 15 percentage points (15%) for rates and .20 for ratios being most common. In addition, all veterinary noninferiority studies reported in freedom of information summaries by the FDA used rates differences with noninferiority margins of $\leq 15\%$. Analgesics and antibiotics were the most common therapeutic class studied in veterinary noninferiority trials, and the majority of noninferiority trials were conducted in dogs or cats. Most likely, the ethical concerns of withholding an established treatment in these types of situations in dogs and cats led to the decision to conduct a noninferiority study instead of a placebo-controlled superiority study.

Example

Choosing the Active Control

Consider the design of a clinical trial for a novel NSAID intended to demonstrate that it is effective for the control of postoperative pain associated with soft tissue surgery in dogs. A placebo-controlled trial would markedly decrease the likelihood of concluding that the novel NSAID is effective when it is not, but it is determined that a placebo-controlled superiority trial is not ethical given that there are NSAIDs and other classes of drugs that have been demonstrated to control postoperative pain. The selection of another NSAID as a positive control is justified because

NSAIDs differ in their safety rather than qualitative aspects of analgesia. Moreover, selecting another class of drugs such as opioids may introduce confounding issues attributable to known behavioral effects of opioids as an extension of their pharmacology. As an active control, there are 5 NSAIDs (carprofen, firocoxib, deracoxib, meloxicam, and robenacoxib) that have been approved by the EMA, FDA, or both for control of postoperative pain in dogs. Large (>100 dogs), randomized, placebo-controlled clinical trials for control of postoperative pain associated with soft tissue surgery have been conducted for only 2 of these NSAIDs (carprofen and firocoxib).^{20,21} A single placebo-controlled study of meloxicam for postoperative pain after soft tissue surgery was small, with questionable effectiveness demonstrated.²² The approval of meloxicam for postoperative pain in dogs was based on a small active control study using 2 different positive control drugs (ketoprofen and butorphanol), where it was demonstrated to be superior to butorphanol alone given before surgery and immediately postoperatively.^{23,24} Robenacoxib has been demonstrated to be noninferior to meloxicam,^{25,26} but no placebo-controlled trial has been published. And finally, deracoxib has not been approved for the control of pain associated with soft tissue surgery in the United States, and it is limited to the control of pain associated with orthopedic surgery as demonstrated in a placebo-controlled trial.²⁷ The large placebo-controlled clinical trials of carprofen and firocoxib for postoperative pain report different effectiveness endpoints, a repeatedly measured visual analog scale (VAS) and treatment failure rate based on reaching a threshold pain score of ≥ 8 using Glasgow Composite Pain Scale (GCPS), respectively. As it is generally desired to have a single, easily interpretable effectiveness endpoint, firocoxib was selected as the active control to allow noninferiority to be established using treatment failure rate. Despite a relatively large number of NSAIDs indicated for control of postoperative pain in dogs, it is evident that the options for conducting a well-designed, noninferiority study for a novel theoretical NSAID for the control of postoperative pain associated with soft tissue surgery are limited.

Determining the Margins

Based on the large placebo-controlled clinical trial of firocoxib in the United States for postoperative pain associated with soft tissue surgery, the treatment failure rates for placebo and firocoxib treatment were 24.2% (31 of 128 dogs) and 6.3% (8 of 126 dogs), respectively.²¹ Thus, the point estimate of the absolute value of the difference in treatment failure rates was 17.9 percentage points. A 2-sided 95% CI was calculated to determine the smallest reliable effect size, δ_0 . The lower bound of 95% CI for difference in failure rates between placebo and firocoxib was calculated as 9.3 percentage points. To incorporate clinical judgment into the margin of the difference, it was determined that maintaining at least half of the point estimate of

the effect of firocoxib compared with placebo would be sufficient for establishment of noninferiority. At least half of 17.9% result in a clinically relevant margin of 8.9%. As 8.9% is less than that of δ_0 (ie, 9.3%), 8.9% was set as δ , incorporating both statistical reasoning and clinical judgment into the *a priori* selected noninferiority margin. With a noninferiority margin of 8.9% and an assumed failure rate of 6.3% for both firocoxib and the novel NSAID, ~120 dogs per treatment group are needed to maintain at least 80% power for the study design.

Interpreting Study Results

As an illustration, we consider 2 possible hypothetical outcomes to the noninferiority trial design above comparing the novel NSAID to the active control firocoxib with 120 dogs in each treatment group. First, consider the outcome where the treatment failure rates are 6.7% (8 of 120 dogs) and 5.0% (6 of 120 dogs) for the novel NSAID and firocoxib, respectively. The point estimate of the treatment differences (novel NSAID—active, as given by Equation 4 when a smaller effect indicates better effectiveness) is 1.7% points and the 2-sided 95% CI is -4.3% to 7.6% . As the upper bound of the 95% CI of 7.6% is $\leq \delta$ of 8.9% (ie, Case A in Fig 4), we reject the null hypothesis given by Equation 4 and conclude that the novel NSAID is noninferior to the active control firocoxib.

Alternatively, consider the outcome where the treatment failure rates are 10.0% (12 of 120 dogs) and 5.8% (7 of 120 dogs) for the novel NSAID and firocoxib, respectively. The point estimate of the treatment differences is 4.2% points and the 2-sided 95% CI is -2.6% to 11.0% . Because the upper 95% CI of 11.0% is $> \delta$ of 8.9% (ie, Case B and C in Fig 4), we fail to reject the null hypothesis given by Equation 4 and conclude that there is insufficient evidence that the novel NSAID is noninferior to the active control firocoxib, even though the point estimate is within the noninferiority margin. In addition, because the upper 95% CI is greater than the smallest reliable effect (ie, δ_0) of 9.3%, there is insufficient evidence from the trial that the novel NSAID has any effect on the treatment failure rate at all.

Compare the study inferences described above for an appropriately designed and interpreted noninferiority trial to a study design and interpretations where a treatment is compared to an active control as a superiority trial with the hypotheses set up in the familiar fashion given by Equation 1. When the null hypothesis is not rejected, the appropriate conclusion in a superiority trial design is that there was insufficient evidence to demonstrate that the treatments were different. However, all too often in veterinary medicine it is concluded (incorrectly) that the new treatment is effective and equivalent to the active control despite insufficient evidence to support the claim. As illustrated in the second hypothetical NSAID outcome above, there was insufficient evidence that the novel NSAID had any effect. If the study was conducted exactly the same but

with the hypotheses set up as in Equation 1, the study would have been an active control superiority trial and the null hypothesis would fail to be rejected at the $\alpha = 5\%$ significance level because the 95% CI included zero. It would be incorrect in the case of the superiority trial to conclude that the novel NSAID is as effective as firocoxib. As is often quoted, “Absence of evidence (of a difference) is not evidence of absence.”²⁸ When the correct conclusion from the superiority trial is made, it does not contradict appropriately interpreted results of the noninferiority trial. The effectiveness of a new treatment cannot be demonstrated in a superiority trial (active or placebo-controlled) unless the null hypothesis is rejected.

Alternatives to Noninferiority Studies

Given the limitations of the noninferiority trials, particularly in the absence of well-designed, contemporary, placebo-controlled trials with an active treatment, other alternative study designs may need to be considered. One option that is sometimes offered is a 3-treatment group noninferiority trial that includes a placebo control, an active control, and the new test treatment. Although this option addresses all of the limitations of the noninferiority trial by allowing for demonstration of assay sensitivity within the trial (which is a substantial improvement to the study design), it is only a viable option if the reason for conducting the trial is to determine the comparative safety and effectiveness of 2 products. Usually, a noninferiority trial is conducted because of ethical concerns of administering a placebo in a study when an established effective treatment exists. Thus, a 3-treatment noninferiority trial does not address the ethical concern of administering a placebo, the very reason the noninferiority design was likely pursued in the first place.

Another possible option involving untreated patients is to conduct a placebo-controlled superiority trial with early escape and administration of a rescue medication attributable to treatment failure. Because many treatment failures are likely to occur in the placebo group, missing data would be a major issue for the completed trial if any endpoint other than treatment failure is used to demonstrate effectiveness, severely limiting the clinical endpoints that can be used in the trial design. The other problem in the design is that it may not sufficiently address the ethical concerns of administering a placebo to patients, depending on the disease or condition being treated (eg, if death is major risk of being untreated). A related but not mutually exclusive option is to identify a subpopulation of patients in which it is ethically appropriate to conduct a placebo-controlled trial. Possible subpopulations may include patients that have failed to respond to the active control or cannot tolerate its adverse effects. A limitation to this study design is that the entire target population of the new treatment is not studied, only a subpopulation, which decreases the inferential value of the trial.

When the new treatment and the established treatment have distinct mechanisms of action, a viable

alternative to a noninferiority trial is an add-on superiority trial. In this study design, all patients get the established treatment and also are randomized to also get either placebo or the new treatment. Thus, with this study design, there is no ethical concern of withholding the established standard of care. In addition, the statistical analysis of an add-on superiority trial is the same as that for a placebo-controlled trial with the hypotheses set up in the familiar fashion (ie, Equation 1), greatly simplifying the statistical considerations of the trial design and providing inherent assay sensitivity in the event that the new add-on treatment demonstrates a benefit over the established treatment alone. However, the major limitation of an add-on trial design is that the effects of the new treatment are not studied independent of the established standard of care. Therefore, regardless of whether or not the new add-on treatment demonstrates a benefit within the trial, it remains unknown if the new treatment would be effective in the absence of the established treatment. An additional limitation of the add-on study design arises if the established treatment has a large effect, because the dynamic range of the response to the add-on new treatment may be limited, potentially decreasing the power of the trial design compared to having an untreated placebo control group.

Another option is to demonstrate superiority of the new treatment compared with the active control. An active control superiority study is only feasible when the new treatment is expected to be substantially more effective than the active control; otherwise, the large size of the study necessary to adequately power it often is not feasible. A final alternative to a noninferiority trial is to conduct a single treatment group-historical control superiority trial. As with a noninferiority trial, good historical knowledge of the disease response in the absence of treatment is needed. However, knowledge is not needed on the disease response with the current standard of care, decreasing the prior study information needed to design the trial. Because there are no concurrent controls within the study, the assay sensitivity of such a study design is not known. Thus, a historical control trial still suffers the same major design limitation as a noninferiority trial and provides no direct information on comparative effectiveness and safety of the 2 treatments.

Ethical Considerations Compared to Placebo Control

The ethics of using placebo controls have been debated elsewhere²⁹ and such arguments have been used to favor an active control noninferiority trial, when available, so as to do no harm in the conduct of the trial. However, understanding the effect of an investigational drug is often best fully achieved in comparison with a negative control.¹² Some therapies that are known to be effective are no better than placebo in particular trials because of variable responses to drugs in particular populations, unpredictable and small effects, and high rates of spontaneous improvement in

patients.^{30,31} In these instances, without a placebo group to ensure validity, the finding that there is no difference between the investigational and standard treatments can be misleading. Thus, all too often it is assumed in these debates that the quality of the evidence from a noninferiority trial is the same as that from a placebo-controlled trial, such that all that is being debated is the ethics of not treating individual patients within the trial itself. In reality, however, noninferiority trials have a large number of caveats and uncertainties associated with them that result in a lower quality of evidence of effectiveness compared with placebo-controlled trials. Therefore, the complete ethical consideration involves weighing the ethics of not treating individual patients within the placebo trial itself, and the ethics of the increased risk of a noninferiority trial incorrectly concluding that an investigational treatment is effective and then subsequently treating a much larger patient population with this ineffective treatment.

Summary

Noninferiority trials usually are pursued because of ethical concerns of administering a placebo to veterinary patients when an established effective treatment exists. Selecting the appropriate active control and an *a priori* noninferiority margin between the new treatment and the control are unique trial design aspects compared to “classical” placebo-controlled superiority studies. However, without good historical knowledge of the disease response with and without treatment, proper design and interpretation of a noninferiority trial are not possible, and a trial of this design should not be conducted. Despite the appeal of conducting noninferiority trials attributable to eliminating the concerns of administering animals a placebo in a superiority trial, there are limitations and questions that arise with the use of such trials. Alternative trial designs to noninferiority trials do exist when ethical concerns prevent administration of a placebo to animals and good historical knowledge of the disease response is lacking, but these trial designs have their own limitations and must be considered carefully.

Acknowledgments

Disclaimer: This article reflects the views of the authors and should not be construed to represent FDA’s views or policies.

Conflicts of interest: Authors disclose no conflict of interest.

References

1. Wellek S. Testing Statistical Hypothesis of Equivalence and Noninferiority, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC; 2010.
2. ICH. ICH Harmonised Tripartite Guideline Choice of Control Group and Related Issues in Clinical Trials E10. International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use; 2000.

3. D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: Design concepts and issues – The encounters of academic consultants in statistics. *Stat Med* 2003;22:169–186.
4. Chow SC, Shao J. On non-inferiority margin and statistical tests in active control trials. *Stat Med* 2006;25:1101–1113.
5. Slingsby L. Considerations for prospective studies in animal analgesia. *Vet Anaesth Analg* 2010;37:303–305.
6. CHMP. Guideline on the Choice of the Non-inferiority Margin. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf: European Medicines Agency (EMA); 2006.
7. CDER, CBER. Guidance for Industry Non-Inferiority Clinical Trials Draft Guidance. <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM202140.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA); 2010.
8. CVM. Guidance for Industry Active Controls in Studies to Demonstrate Effectiveness of a New Animal Drug for Use in Companion Animals Draft Guidance. <http://www.fda.gov/downloads/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/UCM308680.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2012.
9. Lange S, Freitag G. Choice of delta: Requirements and reality—Results of a systematic review. *Biom J* 2005;47:12–27; discussion 99–107.
10. Blackwelder WC. “Proving the null hypothesis” in clinical trials. *Control Clin Trials* 1982;3:345–353.
11. Fleming TR. Design and interpretation of equivalence trials. *Am Heart J* 2000;139:171–176.
12. Snapinn SM. Noninferiority trials. *Curr Control Trials Cardiovasc Med* 2000;1:19–21.
13. Gao P, Ware JH. Assessing non-inferiority: A combination approach. *Stat Med* 2008;27:392–406.
14. Hung HM, Wang SJ, O'Neill R. A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biom J* 2005;47:28–36; discussion 99–107.
15. Senn S. *Statistical Issues in Drug Development*, 2nd ed. West Sussex, UK: John Wiley & Sons Ltd; 2007.
16. ICH. ICH Harmonised Tripartite Guideline Statistical Principles for Clinical Trials E9. International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use; 1998.
17. Morikawa T, Yoshida M. A useful testing strategy in phase III trials: Combined test of superiority and test of equivalence. *J Biopharm Stat* 1995;5:297–306.
18. Cook TD, DeMets DL. *Introduction to Statistical Methods for Clinical Trials*. Boca Raton, FL: Chapman & Hall/CRC; 2007.
19. CPMP. *Points to Consider on Switching between Superiority and Non-inferiority*. London: The European Agency for the Evaluation of Medicinal Products Evaluation of Medicines for Human Use; 2000.
20. Pfizer Inc. Freedom of Information Summary, Supplemental NADA 141-199, Rimadyl® Injectable (carprofen). <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/ucm117619.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2003.
21. Merial Ltd. Freedom of Information (FOI) Summary, Supplemental NADA 141-230, Previcox® Firocoxib Chewable Tablets for Dogs. <http://www.fda.gov/Downloads/Animalveterinary/Products/Approvedanimaldrugproducts/Foiadrugsummaries/Ucm118041.Pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2007.
22. Leece EA, Brearley JC, Harding EF. Comparison of carprofen and meloxicam for 72 hours following ovariohysterectomy in dogs. *Vet Anaesth Analg* 2005;32:184–192.
23. Mathews KA, Pettifer G, Foster R, et al. Safety and efficacy of preoperative administration of meloxicam, compared with that of ketoprofen and butorphanol in dogs undergoing abdominal surgery. *Am J Vet Res* 2001;62:882–888.
24. CVMP. Metacam: European Public Assessment Report (EPAR) - Scientific Discussion. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/veterinary/000033/WC500065773.pdf: European Medicines Agency (EMA); 2010.
25. Gruet P, Seewald W, King JN. Evaluation of subcutaneous and oral administration of robenacoxib and meloxicam for the treatment of acute pain and inflammation associated with orthopedic surgery in dogs. *Am J Vet Res* 2011;72:184–193.
26. CVMP. Onsiar: European Public Assessment Report (EPAR) - Scientific Discussion. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Summary_for_the_public/veterinary/000127/WC500067754.pdf: European Medicines Agency (EMA); 2008.
27. Novartis Animal Health US Inc. Freedom of Information Summary, Supplemental NADA 141-203, Deramaxx™ Chewable Tablest (deracoxib). <http://www.fda.gov/downloads/.../FOIA DrugSummaries/ucm117645.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2002.
28. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
29. Parving HH, Lehnert H, Brochner-Mortensen J, et al. The effect of irbesartan on the development of diabetic nephropathy in patients with type 2 diabetes. *N Engl J Med* 2001;345:870–878.
30. Ellenberg SS, Temple R. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: Practical issues and specific cases. *Ann Intern Med* 2000;133:464–470.
31. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Ann Intern Med* 2000;133:455–463.
32. Hirsch AC, Philipp H, Kleemann R. Investigation on the efficacy of meloxicam in sows with mastitis-metritis-agalactia syndrome. *J Vet Pharmacol Ther* 2003;26:355–360.
33. Boehringer Ingelheim Vetmedica Inc. Freedom of Information (FOI) Summary, Supplemental NADA 141-219, METACAM. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/ucm118027.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2004.
34. Pharmacia & Upjohn Co. Freedom of Information (FOI) Summary, NADA 141-232, SIMPLICEF Tablets. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/ucm118043.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2004.
35. Heine J, Krieger K, Dumont P, et al. Evaluation of the efficacy and safety of imidacloprid 10% plus moxidectin 2.5% spot-on in the treatment of generalized demodicosis in dogs: Results of a European field study. *Parasitol Res* 2005;97(Suppl 1):S89–S96.
36. Krieger K, Heine J, Dumont P, et al. Efficacy and safety of imidacloprid 10% plus moxidectin 2.5% spot-on in the treatment of sarcoptic mange and otocariosis in dogs: Results of a European field study. *Parasitol Res* 2005;97(Suppl 1):S81–S88.
37. Rupprecht CE, Hanlon CA, Blanton J, et al. Oral vaccination of dogs with recombinant rabies virus vaccines. *Virus Res* 2005;111:101–105.

38. Merial Ltd. Freedom of Information (FOI) Summary, NADA 141-253, EQUIOXX Oral Paste. <http://www.fda.gov/downloads/animalveterinary/products/approvedanimaldrugproducts/foiadrugsummaries/ucm051453.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2005.
39. Hanson PD, Brooks KC, Case J, et al. Efficacy and safety of firocoxib in the management of canine osteoarthritis under field conditions. *Vet Ther* 2006;7:127–140.
40. Merial Ltd. Freedom of Information (FOI) Summary, NADA 141-230, Previcox[®] Chewable Tablets (firocoxib). <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/ucm118042.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2004.
41. Bayer HealthCare LLC. Freedom of Information (FOI) Summary, NADA 141-275, PROFENDER Topical Solution. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/ucm062332.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2007.
42. Orion Corporation. Freedom of Information (FOI) Summary, Supplemental NADA 141-267, DEXDOMITOR. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/ucm062318.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2007.
43. Hillerton JE, Cooper J, Morelli J. Preventing bovine mastitis by a postmilking teat disinfectant containing acidified sodium chlorite. *J Dairy Sci* 2007;90:1201–1208.
44. Stegemann MR, Coati N, Passmore CA, et al. Clinical efficacy and safety of cefovecin in the treatment of canine pyoderma and wound infections. *J Small Animal Pract* 2007;48:378–386.
45. Besche B, Chetboul V, Lachaud Lefay MP, et al. Clinical evaluation of imidapril in congestive heart failure in dogs: Results of the EFFIC study. *J Small Animal Pract* 2007;48:265–270.
46. Boehringer Ingelheim Vetmedica Inc. Freedom of Information (FOI) Summary, NADA 141-273, VETMEDIN. <http://www.fda.gov/downloads/animalveterinary/products/approvedanimaldrugproducts/foiadrugsummaries/ucm062328.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2007.
47. Passmore CA, Sherington J, Stegemann MR. Efficacy and safety of cefovecin for the treatment of urinary tract infections in cats. *J Small Animal Pract* 2008;49:295–301.
48. Pfizer Inc. Freedom of Information (FOI) Summary, NADA 141-285 CONVENIA. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/ucm062340.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2008.
49. Intervet Inc. Freedom of Information (FOI) Summary, NADA 141-299 RESFLOR GOLD. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/UCM231147.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2009.
50. Altreuther G, Radeloff I, LeSueur C, et al. Field evaluation of the efficacy and safety of emodepside plus praziquantel tablets (Profender tablets for dogs) against naturally acquired nematode and cestode infections in dogs. *Parasitol Res* 2009;105 (Suppl 1):S23–S29.
51. Janssen Pharmaceutica NV. Freedom of Information (FOI) Summary, NADA 141-298, SUROLAN. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/UCM202424.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2009.
52. Giraudel JM, Gruet P, Alexander DG, et al. Evaluation of orally administered robenacoxib versus ketoprofen for treatment of acute pain and inflammation associated with musculoskeletal disorders in cats. *Am J Vet Res* 2010;71:710–719.
53. McDougall S. A randomised, non-inferiority trial of a new cephalonium dry-cow therapy. *N Z Vet J* 2010;58:45–58.
54. Sickafoose L, Hosgood G, Snook T, et al. A noninferiority clinical trial comparing fluconazole and ketoconazole in combination with cephalexin for the treatment of dogs with Malassezia dermatitis. *Vet Ther* 2010;11:E1–E13.
55. Riesen SC, Schober KE, Cervenec RM, et al. Comparison of the effects of ivabradine and atenolol on heart rate and echocardiographic variables of left heart function in healthy cats. *J Vet Intern Med* 2011;25:469–476.
56. Altreuther G, Gasda N, Adler K, et al. Field evaluations of the efficacy and safety of Emodepside plus toltrazuril (Procox (R) oral suspension for dogs) against naturally acquired nematode and *Isospora* spp. infections in dogs. *Parasitol Res* 2011;109 (Suppl 1):S21–S28.
57. Kamata M, King JN, Seewald W, et al. Comparison of injectable robenacoxib versus meloxicam for peri-operative use in cats: Results of a randomised clinical trial. *Vet J* 2012;193:114–118.
58. Sano T, King JN, Seewald W, et al. Comparison of oral robenacoxib and ketoprofen for the treatment of acute pain and inflammation associated with musculoskeletal disorders in cats: A randomised clinical trial. *Vet J* 2012;193:397–403.
59. Nexcyon Pharmaceuticals Inc. Freedom of Information (FOI) Summary, NADA 141-337, RECUVYRA. <http://www.fda.gov/downloads/AnimalVeterinary/Products/ApprovedAnimalDrugProducts/FOIADrugSummaries/UCM314828.pdf>: U.S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Veterinary Medicine (CVM); 2012.
60. Linton DD, Wilson MG, Newbound GC, et al. The effectiveness of a long-acting transdermal fentanyl solution compared to buprenorphine for the control of postoperative pain in dogs in a randomized, multicentered clinical study. *J Vet Pharmacol Ther* 2012;35(Suppl 2):53–64.
61. Edamura K, King JN, Seewald W, et al. Comparison of oral robenacoxib and carprofen for the treatment of osteoarthritis in dogs: A randomized clinical trial. *J Vet Med Sci* 2012;74:1121–1131.
62. Raymond N, Speranza C, Gruet P, et al. Robenacoxib vs. carprofen for the treatment of canine osteoarthritis: a randomized, noninferiority clinical trial. *J Vet Pharmacol Ther* 2012;35:175–183.
63. Westropp JL, Sykes JE, Irom S, et al. Evaluation of the efficacy and safety of high dose short duration enrofloxacin treatment regimen for uncomplicated urinary tract infections in dogs. *J Vet Intern Med* 2012;26:506–512.
64. Arruda AG, Godden S, Rapnicki P, et al. Randomized noninferiority clinical trial evaluating 3 commercial dry cow mastitis preparations: I. Quarter-level outcomes. *J Dairy Sci* 2013; 96(7):4419–4435.
65. Gruet P, Seewald W, King JN. Robenacoxib versus meloxicam for the management of pain and inflammation associated with soft tissue surgery in dogs: A randomized, non-inferiority clinical trial. *BMC Vet Res* 2013;9:92.
66. Grandemange E, Fournel S, Woehrlé F. Efficacy and safety of cimicoxib in the control of perioperative pain in dogs. *J Small Animal Pract* 2013;54:304–312.