The role of systematic or critical reviews for interventions in veterinary medicine

Paige Baltzell
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd
Part of the Veterinary Medicine Commons

Recommended Citation
Baltzell, Paige, "The role of systematic or critical reviews for interventions in veterinary medicine" (2015). Graduate Theses and Dissertations, 14357.
http://lib.dr.iastate.edu/etd/14357

This Thesis is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
The role of systematic or critical reviews for interventions in veterinary medicine

by

Paige Baltzell

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Veterinary Preventative Medicine

Program of Study Committee:
Annette M. O’Connor, Major Professor
Terry Engelken
Derald J Holtkamp

Iowa State University
Ames, Iowa
2015

Copyright © Paige Baltzell, 2015. All rights reserved.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>NOMENCLATURE</td>
<td>vi</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>References</td>
<td>12</td>
</tr>
<tr>
<td>CHAPTER II A CRITICAL REVIEW AND META-ANALYSIS OF THE EFFICACY OF WHOLE-CELL KILLED TRITRICHOMONAS FOETUS VACCINES IN BEEF CATTLE</td>
<td>14</td>
</tr>
<tr>
<td>Footnotes</td>
<td>29</td>
</tr>
<tr>
<td>References</td>
<td>30</td>
</tr>
<tr>
<td>Figures</td>
<td>33</td>
</tr>
<tr>
<td>Tables</td>
<td>38</td>
</tr>
<tr>
<td>CHAPTER III A CRITICAL REVIEW AND META-ANALYSIS OF THE MAGNITUDE OF THE EFFECT OF ANTHelmINTIC USE ON STOCKER CALF PRODUCTION PARAMETERS IN NORTHERN US STATES</td>
<td>40</td>
</tr>
<tr>
<td>References</td>
<td>57</td>
</tr>
<tr>
<td>Figures</td>
<td>60</td>
</tr>
<tr>
<td>Tables</td>
<td>63</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

## CHAPTER II

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Flow chart of the search and identification of papers relevant to the review</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed <em>T. foetus</em> vaccine for risk of infection in <em>T. foetus</em> susceptible heifers</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed <em>T. foetus</em> vaccine on open risk in <em>T. foetus</em> susceptible heifers</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed <em>T. foetus</em> vaccine on duration of infection in heifers</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed <em>T. foetus</em> vaccine on abortion risk in pregnant heifers</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed <em>T. foetus</em> vaccine for prevention of infection with <em>T. foetus</em> in bulls</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>Funnel plots for risk of infection in <em>T. foetus</em> susceptible heifers (A), open risk in <em>T. foetus</em> susceptible heifers (B), abortion risk in pregnant heifers (C) prevention of infection with <em>T. foetus</em> in bulls (D) and cumulative meta-analysis for risk of <em>T. foetus</em> infection in heifers (plot E)</td>
<td>36</td>
</tr>
</tbody>
</table>

## CHAPTER III

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Database and research report search results</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>Forest plot for mean difference of ADG (kg) in anthelmintic treated stocker cattle vs. untreated controls</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>Funnel plot of studies included in meta-analysis for ADG of stocker calves (mean difference)</td>
<td>62</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

## CHAPTER II
Table 1  Summary of findings for 2-3 subcutaneous doses of whole-cell killed *Tritrichomonas foetus* (*T. foetus*) vaccine .................................................. 38

Table 2  Evidence profile summaries from the GRADE profiler for whole-cell Killed *Tritrichomonas* ........................................................................................................... 39

## CHAPTER III
Table 1  Relevant states included in the review ........................................................................ 63

Table 2  Intervention protocols for manuscripts included in review ................................. 64

Table 3  Average daily gain (kg/day(SEM/SD)) reported in stocker calves .................. 65

Table 4  Weight gain (kg(SEM/SD)) reported for stocker calves............................... 66

Table 5  Risk of bias ........................................................................................................ 67

Table 6  Checklist items for the assessment of comprehensiveness of reporting .... 68
NOMENCLATURE

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>RR</td>
<td>Risk Ratio</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>MTC</td>
<td>Mixed Treatment Comparison</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Annette O’Connor, and my committee members, Dr. Terry Engelken, and Dr. Derald Holtkamp, for their guidance and support throughout the course of this research.

In addition, I would also like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience.

Finally, thanks to my family for their encouragement.
ABSTRACT

This thesis introduces systematic reviews for interventions, why they are used, and how they can benefit veterinary scientific research. Two examples were presented. The first discusses using the tool GRADE for quality assessment of outcomes for systematic reviews and meta-analyses. A critical review and meta-analysis is presented on the topic of vaccine efficacy for the disease Tritrichomonas foetus in beef cattle. This review used GRADE to help determine the quality of the outcomes that were used for this review and meta-analysis. The findings were reported in summary tables. The overall conclusion of this review was that there is a lack of conclusive evidence to support the use of this vaccine in areas where good biosecurity practices are in place, but readers can use the GRADE evidence tables to make their own decision about the results depending on their unique situation. Additionally, this review helped to point out that there were relatively few studies in cows and bulls and therefore the efficacy of the vaccine in these groups of cattle could not be assessed, and may serve as an area of future clinical research.

The second example given in this thesis describes an approach in meta-analysis to compare different treatments for the same disease or problem, indirectly, called a MTC meta-analysis. Only a few MTC meta-analyses have been published in veterinary medicine. An MTC would have been performed for the critical review and meta-analysis that was presented, as there were several treatments utilized that were compared with control (no treatment). This unfortunately could not be done, due to missing information in most of the manuscripts. The overall conclusion of this review was that the results suggest that there is evidence that anthelmintic use has an effect on ADG in beef cattle production systems in a
northern climate of the United States and that no conclusion could be made on weight gain as a meta-analysis could not be conducted due to poor reporting. This conclusion points out that better reporting throughout the current studies would have been needed to fully understand the magnitude of effect anthelmintic interventions for improved ADG and weight gains in beef cattle in northern climates of the United States. Furthermore, additional research would be necessary to draw conclusions regarding the timing anthelmintic interventions as well as a ranking of different anthelmintic products.

The two critical reviews and meta-analyses that are presented in this thesis illustrate the need for quality primary research and comprehensive reporting of primary research. There are key places where veterinary researchers can help to make literature more usable for systematic reviews. Properly reporting measures of precision such as standard errors or standard deviations, transparency in the materials and methods so that extraction of data such as treated and control populations, sex of animals used, or number lost to follow-up, etc. are easily done. Researchers should take these important parameters into consideration prior to starting their research in order to minimize biases and attempt to do so throughout the trial. High quality, and transparent studies are much easier to include in systematic reviews and meta-analyses as well as obtain much more useful information.

Finally, systematic reviews are an important part of medical literature. It seems important for the veterinary medical community to also see these benefits and begin to incorporate systematic reviews more often into veterinary research. There is a movement to ensure that medical treatments are based on the best data available. Therefore, systematic reviews should be used as part of the research process and should play a role in development and design of new research. Many human medical journals are urging or requiring
researchers to perform or utilize an existing systematic review before starting a trial.

Although this may seem like it might add significant time to the research process, this extra step can help to guide research in a more effective way and decrease repetition.
There are two broad categories of review literature for assessing the effect of interventions, traditional or narrative review and systematic review with or without a meta-analysis. Both methods of reviewing literature summarize what is known about a certain topic. However, the difference in the two types for assessing interventions lies in how information is gathered to inform each review. Narrative reviews are different from systematic reviews because of the way information is gathered for the review is not ‘systematic.’ It is up to the author to pick what articles he or she would like to include in the review. In a systematic review as the name implies there is a more regimented way of approaching study selection where all relevant articles are found to be included in the review.

In a traditional review, the author of the review is usually an expert in the field of study and the methods of collecting and information are unknown. The reviewer then summarizes the findings to come to one conclusion. Unfortunately in narrative reviews, a lack of transparency is inherent. This is because each reviewer uses a different set of criteria for deciding what studies should and should not be included in the review and authors don’t usually describe how articles were collected for their review. Additionally, narrative review authors often do not describe the approached used to judge the “weight” that each study is going to be given and this may lead to reviewers overestimating the value of some studies based on personal preference or notions about the subject. Therefore, these types of studies are very prone to bias and are generally not repeatable.
A systematic review is “a review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are included in the review.”¹ Therefore, in contrast to narrative reviews, systematic reviews have a clear set of rules for reviewers to follow. They are less subjective and are transparent, making them repeatable. Increasingly, there are reports that systematic reviews are being reported more often than other types of clinical research including randomized trials.²

Systematic reviews were first started in the 1970’s when a researcher (Glass) who conducted synthesis in the area of psychotherapy, coined the term ‘meta-analysis’.²,³ Around the same time, Archie Cochrane urged health practitioners to practice evidence-based medicine.³ Beginning in the mid 1980’s to 1990’s, researchers moved from narrative reviews to systematic reviews in human healthcare topics.³,⁴ As this was happening, a group called the Cochrane Collaboration opened a center in Oxford in 1992 the aim of this group was to conduct and educate researchers about systematic review methods. The Cochrane Collaboration is now an international network of researchers, academics, and practitioners. Today’s systematic reviews are driven by the evidence-based medicine movement and from methods established by the Cochrane Collaboration.¹

In healthcare, there is a movement to ensure that medical treatments are based on the best data available. Therefore, systematic reviews should be used as part of the research process and should play a role in development and design of new research, veterinary research included. Many human medical journals are urging or requiring researchers to perform or utilize an existing systematic review before starting a trial.⁴ One article urged researchers to begin any randomized control trial with a reference to the systematic review or reviews that prompted the
researchers to start their research in the first place, and then to end their research with a
discussion of an updated systematic review of all the evidence.\textsuperscript{2} Although this may seem like it
might add significant time to the research process, this extra step can help to guide research in a
more effective way and decrease unnecessary repetition.

Systematic reviews and meta-analysis were first adopted by the human medical
community and are now being used in veterinary medicine. Quality systematic reviews are
needed in veterinary medicine to help guide better primary research. One study found that animal
studies on the topic of emergency medicine that did not randomize or blind were more likely to
have reported a difference between study groups than were studies that used these methods to
control bias.\textsuperscript{5} They reported that of the studies that met the inclusion criteria, 32.4\% (95\% CI
27.1\% to 38.1\%) reported randomization. Blinding was reported in 10\% (95\% CI 7.4\% to
14.8\%) and both randomization and blinding was reported in 9.7\% (95\% CI 6.5 to 13.6\%).
Additionally, the studies that did not report randomization or blinding were more likely to be
outcome-positive than studies that used either or both of these methods with an OR of 3.3 (95\%
CI 1.6 to 6.5).\textsuperscript{5} Although this study only reports on a small section of veterinary research, it still
outlines the importance of a critical step in primary veterinary research synthesis that appears to
be commonly left out.

This thesis includes two critical reviews and meta-analysis on topics in veterinary
science. Chapter II is a critical review and meta-analysis on the vaccine efficacy for the venereal
disease \textit{Tritrichomonas foetus} in beef cattle. Chapter III is a critical review and meta-analysis on
anthelmintic use in stocker beef cattle in the northern United States. The apparent benefit from
using the systematic review methodology in these projects associated with the chapters was
intended to highlight a unique aspect of the research synthesis. The first project, documented the
use of the GRADE decision making approach in addition to conducting a systematic review and the second project aimed to conduct a type of meta-analysis called a mixed treatment comparison (MTC) meta-analysis; however, due to issues with the reported data, this was not possible.

The subsequent sections of the introduction addresses 3 topics 1) basic steps of a systematic review, 2) short summary of the GRADE process, and 3) a short discussion about the application of MTC meta-analysis.

The goal of a systematic review is to attempt to take all empirical evidence that fits pre-specified eligibility criteria and answer a specific research question. Procedures have to be defined in advance to make sure that the review is transparent and repeatable and is designed to minimize bias. Usually systematic reviews will contain a meta-analysis, which is used to summarize the results.

In 2010, the EFSA or European Food Safety Authority published in its journal guidelines for carrying out systematic reviews. This publication described in detail how systematic reviews should be conducted. Systematic reviews should start by defining the review question. For assessing intervention questions, the acronym PICO(s) is commonly used. This stands for population, intervention, comparison, outcome, and sometimes study design. In veterinary science, the population of interest helps to define the extent of the review and can include things like the species of animals studied, age of animals studied, or production system that animals belong to, etc. Interventions usually describe therapies such as vaccines or medication that are being used or studied. The comparison should be defined prior to starting and could be a negative control, an alternative treatment, or both. Finally, the outcome is a quantifiable characteristic that should be meaningful to the end user. In the systematic reviews presented in this thesis, the outcomes were either continuous or binary outcomes. As an example, the review
question from the systematic review in chapter II is: what is the magnitude of reduction of infection risk, open risk, abortion risk and duration of infection in heifers, bulls or both that received a whole-cell killed T. foetus vaccine compared with no vaccine?

Once the review question is established then a comprehensive literature search is done to identify all potentially relevant literature. This is done by establishing search terms that attempts to capture all relevant literature. Identical search terms are used on several electronic databases. Additionally, search terms can be used to search for non-published data as well, as was done in the systematic review in chapter III.

All studies’ titles and abstracts are compiled so that review authors can select relevant studies from the search. This is done by defining inclusion criteria and having 2 or more reviewers read each title individually to either include or exclude the studies. Then abstracts of included titles are reviewed for relevance. Finally, full articles are gathered and are screened with inclusion criteria and a final group of studies is complied. This is always done by two or more people to ensure agreement between reviewers for inclusion of all relevant studies found for the review. For the systematic review in chapter III, this process was completed in Distiller SR, a systematic review software program.

Data is then extracted from each study usually by completing identical forms that are made prior to the extraction process. The risk of bias is assessed, results are synthesized either in a qualitative or quantitative (meta-analysis) way, and then results are interpreted. Many systematic reviews contain a meta-analysis as a way to quantitatively synthesize results, as was done in both chapters II and III. The meta-analysis helps to combine all the relevant studies and provide a more precise estimate of the effect than can be done from the individual study.6
While systematic reviews summarize the effect of an intervention on a specific population, another step is required to make decisions about what to do with that information i.e. to recommend use of the intervention or not. This section discusses the quality assessment of outcomes reported and its use is demonstrated in chapter II. The tool used in this example was GRADE or grading of recommendations assessment, development, and evaluation. This tool was developed by the GRADE working group that began in 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present systems in healthcare research. GRADE is an approach to using the evidence in systematic reviews for making decisions. It offers a coordinated approach to developing and presenting summaries of evidence for systematic reviews. Most importantly, GRADE can help clinicians and policy makers to make clearer decisions when making recommendations or trying to change policy.

GRADE assesses the quality of each outcome separately within a study. This is due to the fact that quality may differ from one outcome to another within a single study. Quality is defined in the context of systematic reviews as our confidence that the estimates of the effect are correct and if that evidence is adequate to support a decision or recommendation. High quality evidence is usually associated with more confident recommendations than lower quality; however, this can vary depending on the outcome.

GRADE assesses quality by making judgments to a body of evidence using 5 separate factors. These include: risk of bias, inconsistency, indirectness, imprecision, and publication bias. These five categories are used because they bring together all the issues that have an affect on the quality of the evidence.

There are several contributing factors that are assessed when determining the risk of bias. Lack of concealment to allocation, or when those enrolling patients are aware of the group a
patient will be allocated. Lack of blinding of patients, caregivers, those recording outcomes, those judging outcomes, or data analysts are aware of the arm to which patients are allocated all contribute to the risk of bias.\textsuperscript{11} In veterinary science, lack of blinding patients is not generally an issue; however, blinding owners/caretakers may be more of a consideration.

Additionally, Loss to follow-up is the incomplete accounting of patients and outcome events and can contribute to the risk of bias.\textsuperscript{11} In veterinary studies this is usually due to mortality but can be due to other factors, such as owners not returning for follow up care. GRADE working group states that this issue becomes more significant depending on the rates of loss to follow-up to the number of events. The higher the proportion lost to follow-up in relation to intervention and control event rates, and differences between intervention and control groups, the greater the threat to bias.\textsuperscript{11}

Finally, selective reporting bias or incomplete or absent reporting of some outcomes and not others on the basis of results can have an effect on the overall risk of bias.\textsuperscript{11} This becomes an issue when authors or study sponsors selectively report positive outcomes and analysis within a trial. Selective reporting likely will produce overestimates of the intervention effects.

Other limitations that may have an effect on risk of bias include stopping early for benefit, the use of un-validated outcome measures and carryover effects in crossover trials.\textsuperscript{11} All aspects of the risk of bias need to be taken into consideration when making a judgment about the overall risk of bias.

Inconsistency is another category that GRADE uses to assess quality of evidence. Inconsistency is assessed by making a judgment to the extent of heterogeneity, or any amount of variability among studies\textsuperscript{6}, based on similarity of point estimates, extent of overlap of CIs, and
statistical criteria including tests of heterogeneity and $I^2$. The $I^2$ quantifies the proportion of the variation in the point estimates due to among-study differences. There are four criteria for assessing inconsistency in results. These include: when point estimates vary widely across studies, when CIs show minimal to no overlap, when statistical tests for heterogeneity (which tests the null hypothesis that all studies in the meta-analysis have the same underlying magnitude of effect), and when the $I^2$ is large. It is generally accepted that an $I^2$ of less than 40% is low, 30-60% is moderate, 50-90% is substantial, and 75-100% is considerable. GRADE suggests that reviewers should rate down the quality of evidence if large inconsistency in study results remains after exploration of why that might be the case.

Direct evidence is research that compares the interventions of interest for the population of interest and measures outcomes important to those patients, directly. Therefore, indirectness is when the study population, intervention or outcome is different from the target population, intervention or outcome. Looking at these differences can aid in assessing indirectness. Intervention is not generally an issue in systematic reviews as they clearly specify the intervention of interest. GRADE does not down rate for indirectness unless the biology in the population of interest is so different from that of the population tested that the magnitude of effect will differ substantially. For example, applying a human study to animals or vice versa.

Imprecision is judged based on the 95% confidence interval (CI). The CI represents the uncertainty that is inherent in the estimate and describes a range of values that one can be reasonably sure that the true estimate lies. GRADE uses the CI to describe the impact of random error on evidence quality. When considering quality, GRADE looks at whether the CI around the point estimate is sufficiently narrow. There are reasons described by GRADE to consider downgrading quality on the basis of imprecision. This includes, when the CI around the
estimate include both benefit and harm of an intervention, or if a recommendation would differ if the upper vs. lower boundary of the CI represented the truth. The final criterion that GRADE uses to assess quality is publication bias. GRADE working group states that there is empirical evidence to show that studies with statistically significant results are more likely to be published than studies without statistically significant results. This leaves the least statistically significant studies to remain unpublished or obscurely published, making them difficult to impossible to find. These unidentified studies may have an impact (either towards or away from the null) on the overall effect than those identified.

Publication bias is assessed differently than the other criteria, and is done by looking at the outcomes from a group of studies and not each outcome from each study. The pattern of study results in a funnel plot can be used as a criterion for judging publication bias. A reviewer’s suspicion for publication bias may increase if visual inspection demonstrates an asymmetrical rather than a symmetrical funnel plot, or if statistical tests for asymmetry are positive. Review authors should consider rating down for likelihood of publication bias when evidence consists of a number of small studies. The inclination to rate down for publication bias should increase if most of those small studies are industry supported or likely industry supported.

Chapter II is a published critical review and meta-analysis, on vaccine efficacy for a venereal disease in cattle, which uses GRADE. The use of GRADE in this critical review helps to define the quality of the studies included and therefore helps the end user make judgments about the true efficacy of this vaccine.

The systematic review and meta-analysis of randomized control trials are becoming main sources of evidence in health sciences. Most systematic reviews focus on pair-
wise, direct comparison of treatments vs. control. However, there are many clinical questions with several treatments that already exist and compete with each other. Most end users would benefit from ranking their benefits (and harms) to choose the best options. Ideally, having direct head-to-head comparisons of alternative interventions made within randomized studies would be the best option for comparing multiple treatments but, such studies are often not available. The absence of large, high quality, randomized control trials comparing all eligible treatments clinicians are left to rely on indirect comparisons of multiple treatments. For example, finding the estimate of A over B by comparing trials of A v C and B v C, with C being the same control.

The need for making indirect comparisons has led to the development of mixed treatment comparison (MTC) meta-analysis or also called multiple-treatment meta-analysis or network meta-analysis. These statistical methods are used mostly to analyze studies with various intervention groups and to synthesize studies making different comparisons of interventions as well as ranking possible alternatives. One interesting perspective suggested that ranking benefits (and harms) to choose the best treatment options could be particularly important for policy makers from developing countries. These policy makers want the best treatment available but this option may not be affordable in their situation. Therefore, knowing the second or third best would be helpful in making the best selection.

Some drawbacks of indirect evidence have been identified regarding how valid these comparisons are. Some critics worry that MTCs systematically overestimate the effects of treatments. Evidence that is gained from an MTC should always be considered as retrospective, observational investigations, even when they address high-quality randomized control trials.
Heterogeneity in MTC meta-analysis as in any meta-analysis can point to either genuine diversity between studies or bias within studies, and is very important in the case of MTCs. MTCs make the strong assumption that studies of different interventions or comparisons are similar in all other ways except the interventions that are being compared. Subtle to prominent differences in the exact treatments used (dose, timing, and schedule), type of population studied, setting, or the use of related treatments may exist.\textsuperscript{17}

Additionally, all types of bias affect comparisons in a MTC meta-analysis just as they would in a traditional meta-analysis. Due to the fact that MTCs make indirect comparisons, and therefore are not randomized, they suffer the same biases as observational studies, such as confounding.\textsuperscript{6} Study quality and risk of bias has been recognized as an important part of conventional meta-analysis and, in an MTC meta-analysis it’s equally important. When mixtures of studies with different quality and different risks of bias are used together, it may lead to questionable validity and interpretability of the results.\textsuperscript{17} Therefore, quality of studies included a MTC meta-analysis should be taken into consideration as well before drawing conclusions.

One additional possible source of bias that needs to be examined before pursuing an MTC is the impact of private company sponsorship. Even in simple pairwise comparisons, there are suggestions that the magnitude and direction of observed effect may occasionally be influenced by the sponsorship of the trial.\textsuperscript{17} As an example, comparison of A vs. B favors A when sponsored by the company that makes A. Comparison of A vs. B favors B when sponsored by company that makes B. Usually this is not just publication bias, or selective reporting. It usually reflects more subtle or less subtle differences in study design or the way the study was conducted to help the preferred drug come out on top in that particular study.\textsuperscript{17} As one might imagine, when this is then put into a network comparison it can make the impact of sponsorship
bias more complicated. One author admits that clinical or epidemiological evaluation of inconsistency is hard but suggests that having a plan before starting research to investigate sources of inconsistency is advised. These reports should distinguish planned analysis from post-hoc analysis.\textsuperscript{17}

Chapter III is a critical review and meta-analysis that reviews the use of anthelmintics in stocker calves. The initial goal for this study was to use a MTC meta-analysis to rank the available anthelmintics that are available on the market today. Unfortunately, this was not possible for this group of manuscripts. There is a discussion of why this was the case in that section.

References

CHAPTER II

A CRITICAL REVIEW AND META-ANALYSIS OF THE EFFICACY OF WHOLE-CELL KILLED TRITRICHOMONAS FOETUS VACCINES IN BEEF CATTLE

Published in the Journal of Veterinary Internal Medicine
P. Baltzell, H. Newton, A. O’Connor

Introduction

A critical review and meta-analysis was performed to estimate the efficacy of killed, whole-cell *T. foetus* vaccine with regard to incidence of *T. foetus* infection in heifers and bulls, duration of *T. foetus* infection in heifers and bulls, pregnancy percentage and abortion risk in heifers, and the ability of the vaccine to clear *T*. foetus-infected bulls of the infection. The motivation for conducting the review was the reemergence of *T. foetus* infection as a cause of reproductive failure in US Midwest cow-calf herds.\textsuperscript{16-18}

Bovine Trichomoniasis is a sexually transmitted disease caused by *Trichomonas foetus* (*T. foetus*).\textsuperscript{1-3} In bulls, *T. foetus* lives in the smegma of the epithelial lining of the penis, prepuce, and distal urethra, and is transmitted to females through infected preputial secretions.\textsuperscript{4-6} Infected bulls older than 3-4 years of age often are chronically infected. In cows and heifers, the most common sequela to infection is reproductive failure, but overt clinical signs of infection can include endometritis, salpingitis, placentitis, abortion, and potential subsequent pyometra.\textsuperscript{11-13} It may take months for cattle to regain fertility\textsuperscript{12}. Based on a simulation model, a herd with a 20 to 40 percent prevalence of *T. foetus* infection in the breeding bulls might expect a 14 to 50 percent reduction in annual calf crop size, a growing period decreased by 12 to 30 days, and weaning weights decreased by 22 to 53 pounds.\textsuperscript{14} The result is wide variability of weaning weights, forcing producers to sell calves at lower weights or incur higher feeding costs.\textsuperscript{7}
Common approaches to prevention of *T. foetus* introduction into a herd include biosecurity practices: limiting the potential for bulls from neighboring properties to mate with the herd, purchasing only virgin bulls, purchasing older bulls confirmed to be *T. foetus* negative, and artificial insemination. When biosecurity measures are not practiced, the efficacy of these measures cannot be ensured, or if further assurance of a *T. foetus* free herd is desired, vaccination may also be considered. Currently, 1 *T. foetus* vaccine is available on the US market, a killed, whole-cell protozoan vaccine indicated for vaccination of healthy cattle as an aid in the prevention of disease caused by *T. foetus*.

Given the emergence of *T. foetus*, the review question was, “What is the magnitude of reduction of infection risk, open risk, abortion risk and duration of infection in heifers, bulls or both that received a whole-cell killed *T. foetus* vaccine compared to no vaccine?” Furthermore, we used a slightly modified GRADE Summary of Findings and Evidence Profile Table to present the results in a format previously not used in veterinary science.

**Materials and methods**

**Literature selection**

The approach to the review was designed by 1 of the authors (AOC). No protocol was registered or externally reviewed. One author (PB) had experience conducting 1 systematic review and another author (AOC) had conducted numerous systematic reviews and meta-analyses. One author had no prior experience with systematic reviews. The review question followed the PICO format for systematic reviews: the relevant population (heifers or bulls), intervention (a whole-cell killed *T. foetus* vaccine), comparator (no vaccine) and outcomes (magnitude of reduction of infection risk, pregnancy loss, and duration of infection). The review was limited to whole-cell *T. foetus* vaccines, because the only commercially available product is
a whole-cell vaccine. Any study that described the use of a whole-cell vaccine was used, (i.e., both commercially available and in-house laboratory based whole-cell vaccines were considered relevant to the review).

To identify relevant primary research, the citation indexes PubMed, CAB Abstracts, and Agricola were searched in the first 2 weeks of June 2012. The PubMed search terms were “(Cattle OR Bovine) AND (Trichomonas foetus) AND (Vaccine* OR Vaccinate* OR Immunization OR Control OR Prevention.)” Analogous search terms were used in CAB Abstracts and Agricola. No language or date restrictions were imposed during the search. Retrieved citations were imported into Endnote Web®. Within Endnote Web® duplicates were removed based on title match only, manuscripts published in languages other than English were removed, as there were no funds for translation of articles, and articles published before the 1950’s were excluded as prior experience suggests few of these studies include control groups. We also hand-searched the reference lists of previously published reviews about T. foetus for relevant studies.8-10 We contacted the manufacturer of the only commercially registered whole-cell killed T. foetus vaccine in the US to request studies about vaccine efficacy.

To determine the outcomes considered in the review, a list of possible outcomes was extracted from the several articles considered likely to be included in the review. Three beef production experts at Iowa State University College of Veterinary Medicine were asked to rank the outcomes by relative importance, highest to lowest. The experts ranked titers to T. foetus and other immunological measures as non-important outcomes. Therefore, data from antibody tests were not extracted from any study, although many studies reported this outcome.
Screening of the remaining retrieved citations was conducted independently by 2 authors (PB and HN), DVM students working in a summer research program. Both authors screened all retrieved citations based on the following eligibility criteria:

1. Did the study describe primary research?
2. Was the research conducted in cattle associated with beef production?
3. Did the study assess the efficacy of killed whole cell *T. foetus* vaccine for the prevention or control of Trichomoniasis?
4. Did the study utilize a treatment and control group, the latter of which did not receive the vaccine?

If the response to each question was “yes”, the study was included in the review. When it was not possible to determine relevance based on the abstract and title, the full text articles were obtained and evaluated based on the same 4 questions.

Data extraction

From the full text of relevant publications data on whether trials reported randomization, the intervention, and relevant outcomes were extracted by 2 authors (PB and HN). We did not extract information on study populations demographics or other design features because the inclusion criteria were very narrow. The unit of concern for data extraction was the study. Some manuscripts had more than 1 study. When clarification about an outcome or data were needed, the third author was consulted. Data were extracted for the following outcomes of interest: infection risk in heifers and bulls, duration of infection in heifers and bulls, open risk and abortion risk in heifers, and the ability of the vaccine to clear infected bulls of infection. The open risk refers to the number of heifers not pregnant divided by those bred.
When extracting data for days of infection, sometimes these data were reported as group means and standard error of the mean (SEM). Other times the individual animal tests results over time were reported and we hand calculated the mean days of infection and standard deviation. Furthermore, when studies reported the SEM for the days infected, we used Review Manager (RevMan) software to back-calculate the standard deviation for each treatment group based on the number of animals in each group and SEM.

Data analysis

For outcomes with more than 1 study, the extracted data were entered into and analyzed using RevMan. For dichotomous outcomes, infection risk, open risk and abortion risk, a risk ratio with a 95% CI for each study was determined. The summary effect measure comparing the risks was the Mantel-Haenszel summary risk ratio from a random effect model. For the average number of days infected, a continuous outcome, the summary effect measure was the mean difference in days infected. The hypothesis was that the overall summary effect was equal to the null value (risk ratio = 1 or mean difference = 0). For all outcomes, it was expected that vaccination would decrease adverse events; therefore, the risk ratio should be < 1.0 and the mean difference < 0 if the vaccine is effective. A subgroup analysis based on the length of time for follow up was conducted for 2 outcomes, infection risk and open risk in heifers. This aim of the subgroup analysis was to assess if length of follow-up was a source of heterogeneity.

Heterogeneity was assessed using the chi-square test for heterogeneity overall and within the subgroups. The null hypothesis was that heterogeneity was not present. If the p value for heterogeneity between the subgroup effects was significant (P<0.1), we concluded the subgroups were different and only then assessed and interpreted the p value for heterogeneity within the subgroup. If the p value for subgroup heterogeneity was > 0.1, we concluded that the subgroup
was not a source of heterogeneity. The subgroups were collapsed, and heterogeneity then was assessed across the entire population. We also reported the I^2 which describes the percentage of variation across studies due to heterogeneity rather than chance.\textsuperscript{24,25} The data were also used to create a forest plot and funnel plot of each outcome with greater than 1 study as well as a cumulative meta-analysis plots, in meta and rmeta, packages in R.\textsuperscript{d}

After meta-analysis, the data were exported from RevMan into GRADEpro,\textsuperscript{e} a software package designed to guide reviewers through the process of assessing the quality of the scientific literature contributing to evidence base for each outcome, and to generate a Summary of Findings and Evidence Profile table. The GRADE systems stands for “The Grading of Recommendations, Assessment, Development, and Evaluation”\textsuperscript{21}

The GRADE system is based on the presence of inconsistency,\textsuperscript{26} indirectness,\textsuperscript{27} imprecision,\textsuperscript{28} and risk of bias\textsuperscript{29,30} existing in evidence base contributing to each outcome. The evidence base is considered to show evidence of inconsistency if there is a wide variation in point estimates, lack of overlap in CI, or evidence of heterogeneity among studies.\textsuperscript{26} The evidence base is considered to have evidence of indirectness if the study populations, interventions, or outcomes are different from those of interest.\textsuperscript{27} The evidence base is considered to have evidence of imprecision if the studies have wide CI. This could result from a sample size that is smaller than the number generated by a conventional sample size calculation for an adequately powered trial.\textsuperscript{28} The evidence base is considered to have evidence of risk of bias if the studies in the review fail to report concealment of allocation, blinding, have incomplete accounting of subjects, large loss to follow up, show selective outcome reporting, or other factors such as recruitment bias, stopping early, or using unvalidated outcome measures.\textsuperscript{30}
The presence of inconsistency,\textsuperscript{26} indirectness,\textsuperscript{27} imprecision,\textsuperscript{28} or risk of bias\textsuperscript{29,30} decreases the quality of evidence grade assigned to the evidence base. The quality of the reviewed body of work can be increased if the observed effects are large, if the only possible confounding bias is towards the null and if evidence of a dose response is available. The GRADE quality scales and interpretations that GRADE attaches to these terms are as follows:

- **Very low quality** - indicates that we have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

- **Low quality** - indicates that our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.

- **Moderate quality** - indicates that we are modestly confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

- **High quality** - indicates that we are very confident that the true effect lies close to that of the estimate of the effect.

For this review, the outcomes reported were not subjective, (i.e. infection detected by culture or pregnancy or abortion), and thus the impact of failure to blind on the risk of bias was considered minimal. Failure to randomize was considered important but, its impact was most likely to be seen as publication bias. The impact of multiplicity was also considered important but, it also was thought that it would manifest itself as selective reporting of outcomes with p values < 0.05 or publication bias. Suspicion of risk of publication bias was considered present if there was evidence that studies showing positive results were more likely to be published, whereas negative results were more likely to be excluded from the literature.\textsuperscript{29} Publication bias may be inferred using a funnel plot and cumulative meta-analyses.
The rankings of the outcomes obtained from the 3 production specialists at the start of the review process were used to categorize the outcomes as critical or important in GRADE ranking system. Critical outcomes included the prevention of *T. foetus* infection in heifers and bulls. Important outcomes included clearance of infection in bulls after therapeutic use of the vaccine, and all other outcomes in heifers.

After ranking the outcomes and grading the body of evidence, the Summary of Findings table and Evidence Profile were created from the GRADEpro work. GRADE has 2 components, 1 that refers to the grading and presentation of evidence using the Summary of Findings table and Evidence Profile table, employed here. The second component of GRADE is formal approach to making recommendations based on the evidence, which we did not employ. The GRADE Summary of Findings tables were augmented with the forest plot and meta-analysis from RevMan to create a comprehensive summary of the evidence base for each outcome. When the p-value for the hypothesis testing that the subgroup was a source of heterogeneity was > 0.1, we collapsed the data and only assigned a GRADE to the combined body of work, not the individual sub-groups. However, when sub-groups were available, we still included these in the forest plot as a means of enabling end-users to visually assess the evidence for heterogeneity by subgroup.

The information in the tables should be used by end users to determine if they would recommend vaccine use. Based on local value and preferences, resource availability, end users may reach different conclusions about the value of the vaccine.

**Results**
The results of the search are reported in Figure 1. For all outcomes, the GRADE Summary of Findings is presented in Table 1. The GRADE Evidence Profiles are provided in Table 2. Meta-analyses are provided Figures 2 through 6. Funnel plots are presented in Figure 7. Eight manuscripts reported risk (cumulative incidence) of *T. foetus* infection in heifers as an outcome, 32-39 2 of which reported 2 separate studies for a total of 10 studies.36,38 The quality of evidence was considered low due to the presence of inconsistency based on the overall heterogeneity and possible publication bias (Table 1 and Table 2). The funnel plot (Figure 7 A) suggested that larger studies were more likely to report effects closer to the null value (no effect). The cumulative forest plot suggested that more recently conducted studies also were likely to demonstrate decreased efficacy compared to older smaller studies (Figure 7 E). We assessed length of follow up as a source of subgroup heterogeneity. The test for subgroup heterogeneity was not significant (chi-square p value for subgroup heterogeneity = 0.28), therefore the data were combined into a single summary effect measure. The Mantel-Haenszel summary risk ratio was 0.89 (95% CI; 0.76 to 1.05), suggesting a 10% decrease in the risk of infection in vaccinated exposed heifers compared to non-vaccinates (p value for overall effect = 0.16). The uncertainty about this effect extends from protective to not protective (95% CI; 0.76 to 1.05).

Seven manuscripts evaluated the open risk as an outcome, 32-37,39 4 of which involved 2 separate studies33,35-37 (i.e., 11 studies with relevant outcomes). The quality of the evidence in the 7 studies was considered moderate. The factor that downgraded the evidence base was the potential for selective reporting bias (Table 1 and Table 2) The test for subgroup heterogeneity was not significant (chi-square p value for subgroup heterogeneity = 0.93). Therefore, the data were combined into a single meta-analysis. The Mantel-Haenszel summary risk ratio was 0.80 (95% CI; 0.63 to 1.01), suggesting a 20% decrease in the open risk in vaccinated exposed heifers.
compared to non-vaccinates (p value for overall effect = 0.06). The uncertainty about this effect extends from protective to barely un-protective (95% CI; 0.63 to 1.01).

Five manuscripts reported 4 studies evaluating the effect of whole-cell *T. foetus* vaccine on the duration of infection in heifers,\(^{34-37,39}\) 1 of which involved 2 different studies\(^{36}\) (i.e., 6 studies with relevant outcomes). The quality of evidence was considered moderate (Table 1 and Table 2). The reason this body of evidence was not given a high rating was associated with our concerns about the meta-analysis approach rather than the analyzed studies. For this outcome, the data were poorly reported, and it was often unclear if the measures of variation reported were SEM or standard deviations. Furthermore, it was unclear that such data were normally distributed. Therefore, the validity of these as measures of variation was unclear. Some studies did not report summary data but reported individual data. Therefore, we calculated average days infected from these data. When we calculated the data we only included animals that were infected in calculations, but it was not clear if others had done so. Finally, although the average mean differences in days infected was normally distributed, a survival analysis could have been used to better capture the distribution of the outcome. Consequently, the body of evidence was downgraded from high to moderate. The forest plot is presented in Figure 4. The data suggested the mean difference of duration of infection was decreased by almost 23 days (95% CI; -38.36 to -7.85) in vaccinates compared to non-vaccinates (p value for overall effect <0.00001). There was no evidence of heterogeneity among studies (p value for overall heterogeneity = 0.11).

Five manuscripts reported 5 studies evaluating the effect of whole-cell *T. foetus* vaccine on the incidence of abortion in pregnant heifers.\(^{32,34,35,37,39}\) The quality of the evidence was rated as low due to imprecision and possible selective reporting bias (Table 1 and Table 2). The forest plot is presented in Figure 5 and the Mantel-Haenszel summary risk ratio was 0.57 (95% CI;
0.42-0.78), suggesting a 40% reduction in the incidence of abortion among vaccinated pregnant heifers compared to non-vaccinated pregnant heifers (p value for overall effect = 0.0003). There was no evidence of heterogeneity among studies (chi-square p value for overall heterogeneity = 0.58).

Two manuscripts evaluating risk of infections in bulls were identified. The quality grade for this evidence was low due to imprecision (3 small studies with total n = 68) and possible publication bias because only 3 studies were available, and the smaller studies had larger effects (Figure 7. D). The magnitude of the Mantel-Haenszel summary risk ratio was 0.41 (95% CI; 0.17 to 0.99), suggesting an approximately 64% decrease in risk of infection among vaccinates (p value for overall effect = 0.05). There was little evidence of statistical heterogeneity across these studies (chi-square p value for overall heterogeneity = 0.20, I² = 39%).

Only 1 manuscript assessed therapeutic use of whole-cell killed T. foetus vaccine in infected bulls. The quality grade for the evidence was very low due to very serious imprecision (a single study with 12 animals per group) and a strong possibility of publication bias (Table 1 and Table 2) (i.e., it seems unusual that this finding has not been replicated since its original publication in 1983). The reported magnitude of the Mantel-Haenszel risk ratio for this single study was 0.36 (95% CI; 0.16 to 0.77), implying an approximately 60% decrease in infection, (i.e. fewer remaining infections in vaccinates; chi-square p value for overall effect = 0.009).

Discussion

This review summarized the body of literature describing the efficacy of killed, whole-cell Tritrichomonas foetus vaccine and its effects on incidence of infection, duration of infection, open risk, and abortion risk. With respect to the quality of the body of work for the outcomes related to females, none were considered high quality bodies of evidence. Consistent issues
identified included small studies, non-randomized studies, and possible selective reporting and publication bias. Consequently, the conclusion is that there is limited or no evidence that vaccines decrease infection or open risk in heifers as the magnitude of the effect observed was small and the body of work of moderate quality. There was some evidence that vaccination decreases abortion as the magnitude of the effect was large however the body of work was of low quality. Our reservations about this finding stem mainly from the potential for selective reporting. Most of the studies assessed short term pregnancy percentage, but only a few reported final cumulative incidence of pregnancy. The concern is that the studies that observed, but did not report, long term pregnancy percentages or abortion risks did so because nothing of interest was observed (i.e., the vaccine had no effect on the abortion risk). If available, such results would have made the vaccine look less effective.

The quality of the work describing the outcomes in bulls was even poorer than those for female-associated outcomes. For infection risk in bulls, the number of studies identified was only 3, 2 of which were not randomized. The only randomized study had 4 vaccinated animals and 8 unvaccinated animals. It is very concerning that veterinarians should be expected to conclude that vaccines are associated with an approximate 60% reduction in infection risk based only on 1 randomized study with 4 vaccinates. Information about clearance of *T. foetus* infection in bulls was only available from 1 study, and therefore although the study was reasonably large by veterinary standards, this is still only 1 observation, and it is not possible to judge if the findings from that study are representative of the true vaccine effect. Again, it is difficult to believe that veterinarians should be expected to rely upon data from a single non-randomized study for reliable decision-making.
With respect to the execution of the review, we did not call this a systematic review preferring instead to use the term critical because several features of a systematic review are missing, in particular external review of a protocol and a panel with diverse expertise. The review however does have many aspects of a systematic review to enhance transparency. We conducted a comprehensive electronic search but did not manually search journals for articles that may have been overlooked. We did, however, verify that no relevant citations from the prior reviews were missed by the electronic search. Therefore, the potential for bias was minimized. We did not hand search conference proceedings of the World Buiatrics Annual Meeting, and relevant studies may have been missed in that publication. However, because no relevant studies from that conference were identified by prior reviews, it seems unlikely any were overlooked. We relied upon indexing in CAB abstracts to capture articles published in Bovine Practitioner and the conference proceedings of the Annual Meeting of the American Association of Bovine Practitioners.

Studies that were non-randomized were included in summary of the evidence. This was done because there were few studies that randomized (Table 2) and also met our inclusion criteria. Also consistent with GRADE, our aim was to provide a comprehensive summary of the literature for others to evaluate. The likely direction of bias by failing to exclude these non-randomized studies is away from the null, as the studies that identified a protective effect of the vaccine because of confounding would have been more likely to be published due to publication bias. An alternative strategy, exclusion of non-randomized articles, would have limited the information available to make meaningful conclusions. Failure to blind also was not used as an exclusion criterion. The rationale was that the observed outcomes were objective, and therefore
most likely unaffected by blinding. Veterinarians should consider their interpretation of the impact of randomization on vaccine efficacy when considering whether to recommend vaccination to producers.

The use of the GRADEpro system and the GRADE approach to grading the evidence base is novel. GRADE has 2 components, 1 that refers to the grading and presentation of evidence, employed here. The second component of GRADE is a formal approach to making recommendations based on the evidence, and was not employed in this project. Our ability to extend the project to include recommendation-making was limited by the fact that the project was a summer project for a student, whereas recommendation-making needs engagement of stakeholders. The value of the GRADE tables is that they enable others to develop recommendations based on this evidence. GRADE refers to this feature as globalization of evidence but localization of recommendations.

It is not possible here to review all of the publications about GRADE tables and recommendation-making process, and the reader is referred to other publications. With GRADE, assessments of the quality level of each evidence base were partitioned into 4 separate assessments about the presence of inconsistency, indirectness, imprecision, and risk of bias. In doing so, these judgments become transparent to the reader. Similar judgments are made in all reviews, but rarely communicated to the end-user. The recommendation-making process not included in this review includes considerations of evidence of efficacy and values and preferences and more judgments.

The GRADE working group encourages the use of the Summary of Findings and Evidence Profile tables (Table 1 and Table 2). End-users need short and concise information,
but reviewers also need to convey the nuances of interpretation that have gone into the quality judgment; these tables aim to balance these goals. Our decision to use GRADE was largely based on the use of absolute risk measures to convey the impact of the interventions.\textsuperscript{22,44}

Absolute risk measures are considered easier to interpret than relative measures of effect such as the risk or odds ratio.\textsuperscript{45,46} By presenting populations of different risk, producers and veterinarians should be able to better evaluate the decision to use the vaccine. For example, in a population with 10\% risk of infection, it is expected that 4 cases would be prevented by vaccination, but this number could be as low as 2 or as high as 10, given the uncertainty of the estimate (Table 1). Producers may decide not to use the intervention because the cost of 100 doses of vaccine may exceed the savings from 6 cases given the uncertainty of the estimate. Alternatively, in a high-risk population, the reduction from 80 to 34 cases (95\% CI; 14 to 79 cases) due to vaccination may be considered highly economical.

Another concern about this body of work is that all studies used the individual as the unit of treatment allocation, rather than the herd. This raises the possibility of interference (i.e., herd immunity).\textsuperscript{47-49} Such interference usually biases the body of work towards the null. We hypothesize that this effect would be minimal in this review. Herd immunity occurs because transmission of the organism is decreased. However, for this disease, infected bulls are the major source of transmission and are not vaccinated in the heifer studies. Therefore, if the vaccine was 100\% effective in preventing Trichomoniasis in heifers, the incidence of infection in non-vaccinated heifers would be unaffected by proximity to large numbers of vaccinated heifers. Of course, if the major mode of transmission was from infected heifer to non-infected heifer, with bulls acting only as vectors of the organism without becoming infected themselves, this
may result in a bias towards the null, (i.e. herd immunity). Our understanding is that the organism lives in the prepuce of the bull, and it is the bull that is the main source of organism.

For all outcomes, the evidence bases were designated as moderate to very low quality. Based on this review, our conclusion is that there is a lack of conclusive evidence to support the use of this vaccine in areas where good biosecurity practices are in place. Others may use the GRADE tables to reach similar or different conclusions, depending upon local circumstances. For example, some veterinarians still may recommend the vaccine in settings where biosecurity is difficult may elect to use vaccine, as the circumstances differ. Such a judgment could only be reached if the veterinarian placed lower value on the small study sizes, quality of the evidence and sources of bias than we do, and placed higher value of the estimates of effect. Although there has been research on the effect of the vaccine in heifers, the effect in cows has not yet been assessed. Also, only 3 unconvincing studies have been published assessing the efficacy of this T. foetus vaccine in bulls. To improve the current body of research, herd level studies of multiple year duration would need to be conducted in production systems at equal risk of infection due to certain management practices. This would greatly add to the understanding of the efficacy of whole-cell killed T. foetus vaccine in a susceptible cattle population.

Footnotes

a TrichGuard and TrichGuard V5L, Boehringer Ingelheim Vetmedica, Inc., St. Joseph, MO
References


FIGURES

Search Terms

- “cattle OR bovine”
- “Tritrichomonas foetus”
- “vaccine* OR vaccinate* OR immunization OR control OR prevention”

Total articles retrieved (n=334)
- CAB abstracts (n=232)
- PubMed (n=64)

Duplicates excluded (n=72)

Abstracts reviewed for assessment for inclusion or exclusion (n=262)

Abstracts excluded – did not meet criteria (n=222)

Full text articles retrieved for further assessment for inclusion or exclusion criteria (n=40)

Full text articles excluded – did not meet criteria (n= 23)

Abstracts excluded – did not meet criteria for vaccine type (n=7)

Articles that qualified to be included in review (n=10)

Figure 1. Flow chart of the search and identification of papers relevant to the review
Fig 2. Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed \textit{T. foetus} vaccine for risk of infection in \textit{T. foetus} susceptible heifers

Fig 3. Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed \textit{T. foetus} vaccine on open risk in \textit{T. foetus} susceptible heifers

Fig 4. Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed \textit{T. foetus} vaccine on duration of infection in heifers
Fig 5. Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed *T. foetus* vaccine on abortion risk in pregnant heifers.

Fig 6. Meta-analysis and forest plot for effect of 2-3 subcutaneous doses of whole-cell killed *T. foetus* vaccine for prevention of infection with *T. foetus* in bulls.
A. Outcome - Risk of infection following exposure to *Trichomonas foetus* based on culture of cervical mucus samples.

B. Outcome - Open risk in heifers after exposure to *Trichomonas foetus* based on number exposed to bulls

C. Outcome: Abortion risk - Pregnant heifers suffering reproductive losses following exposure to *Trichomonas foetus* based on abortion of pregnancy

D. Outcome - Risk of infection in bulls following exposure to *Trichomonas foetus* based on culture of preputial samples.

E. Cumulative meta-analyses forest plot for cumulative incidence of infection following exposure to *T. foetus* based on culture of cervical mucus.
Fig 7. Funnel plots for risk of infection in *T. foetus* susceptible heifers (A), open risk in *T. foetus* susceptible heifers (B), abortion risk in pregnant heifers (C), prevention of infection with *T. foetus* in bulls (D) and cumulative meta-analysis for risk of *T. foetus* infection in heifers (plot E)
### Tables

#### Table 1 Summary of findings for 2-3 subcutaneous doses of whole-cell killed Tritrichomonas foetus (T. foetus) vaccine

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Illustrative comparative risks* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No vaccine</td>
<td>Whole-cell killed T. foetus vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of T. foetus infection in heifers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range of follow-up after challenge: 1-18 weeks</td>
<td>Study population</td>
<td>82 infections per 100 susceptible heifers</td>
<td>73 infections per 100 susceptible heifers (63 to 87)</td>
<td>RR 0.89 (0.76 to 1.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate risk population</td>
<td></td>
<td></td>
<td>50 per 100</td>
</tr>
<tr>
<td>Open risk</td>
<td>Range of follow-up after challenge: 35 days- calving</td>
<td>Study population</td>
<td>39 open heifers per 100 bred heifers</td>
<td>31open heifers per 100 bred heifers (25 to 39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate risk population</td>
<td></td>
<td></td>
<td>50 per 100</td>
</tr>
<tr>
<td>Duration of infection in heifers in days</td>
<td>The mean duration of infection in days when exposed to T. foetus was 83 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abortion risk</td>
<td>Study population</td>
<td>68 abortions per 100 susceptible pregnant heifers</td>
<td>38 abortions per 100 susceptible pregnant heifers (28 to 53)</td>
<td>RR 0.57 (0.42 to 0.78)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low risk population</td>
<td></td>
<td></td>
<td>10 per 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate risk population</td>
<td></td>
<td></td>
<td>50 per 100</td>
</tr>
<tr>
<td>Risk of T. foetus infection in bulls</td>
<td>Study population</td>
<td>80 infections per 100 T. foetus susceptible bulls</td>
<td>33 infections per 100 T. foetus susceptible bulls (14 to 79)</td>
<td>RR 0.41 (0.17 to 0.99)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low risk population</td>
<td></td>
<td></td>
<td>10 per 100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate risk population</td>
<td></td>
<td></td>
<td>50 per 100</td>
</tr>
<tr>
<td>Clearance of T. foetus infection in bulls</td>
<td>Study population</td>
<td>88 infections remain at 14 weeks per 100 infected bulls</td>
<td>32 infections remain at 14 weeks per 100 infected bulls (14 to 67)</td>
<td>RR 0.36 (0.16 to 0.77)</td>
</tr>
</tbody>
</table>

---

*a Some studies were very small.

*b A moderate effect in small studies.

*c Single study with n=24,31

*d Studies of shorter duration showed less difference between groups which would be expected given the study design.
The funnel plot suggests larger studies showed effects closer to the null. Further, studies that reported more protective effects tended to be older. This provides evidence of regression towards the mean.

Studies that reported short-term pregnancy percentages would most likely also have access to calving percentages and abortion risks. However fewer studies reported calving percentages and abortion risks and all that did report, reported an outcome favorable to the vaccine. Selective reporting of vaccine favorable outcomes is therefore a possibility. (Note: several studies that reported pregnancy percentages euthanized animals before calving.) Other studies slaughtered animals prior to 180 days.

Downgrade based on concerns about the validity of the data form-i.e SEM and standard deviations derived from time to event data.

The basis for the assumed risk is the median of control group risk across studies. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). CI: Confidence interval; RR: Risk ratio.

Table 2 Evidence profile summaries from GRADE profiler for whole-cell killed Tritrichomonas foetus vaccine

<table>
<thead>
<tr>
<th>No. of studies</th>
<th>Design</th>
<th>Risk of bias</th>
<th>Inconsistency</th>
<th>Indirectness</th>
<th>Imprecision</th>
<th>Other Considerations</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome: Risk of infection in susceptible heifers (follow-up 18 weeks), a critical outcome</td>
<td>6 randomized and 4 non-randomized challenge studies</td>
<td>no serious risk of bias</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>no serious imprecision</td>
<td>publication bias</td>
<td>LOW</td>
</tr>
<tr>
<td>Outcome: Open risk in susceptible heifers: an important outcome</td>
<td>6 randomized and 5 non-randomized challenge studies</td>
<td>no serious risk of bias</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>no serious imprecision</td>
<td>selective reporting bias</td>
<td>MODERATE</td>
</tr>
<tr>
<td>Outcome: Duration of infection in days in susceptible heifers: an important outcome</td>
<td>2 randomized and 3 non-randomized challenge studies</td>
<td>serious risk</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>no serious imprecision</td>
<td>no concerns</td>
<td>MODERATE</td>
</tr>
<tr>
<td>Outcome: Abortion risk - pregnant heifers suffering reproductive losses: an important outcome</td>
<td>3 randomized and 2 non-randomized challenge studies</td>
<td>no serious risk of bias</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>serious</td>
<td>selective reporting bias</td>
<td>LOW</td>
</tr>
<tr>
<td>Outcome: Risk of infection in susceptible bulls (follow-up 34 months), a critical outcome</td>
<td>2 randomized and 1 non-randomized challenge studies</td>
<td>serious risk of bias</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>serious</td>
<td>publication bias</td>
<td>LOW</td>
</tr>
<tr>
<td>Outcome: Clearance of infection from infected bulls (follow-up 36 months); an important outcome</td>
<td>1 non-randomized challenge studies</td>
<td>no serious risk of bias</td>
<td>no serious inconsistency</td>
<td>no serious indirectness</td>
<td>very serious</td>
<td>publication bias</td>
<td>VERY LOW</td>
</tr>
</tbody>
</table>

a Some studies were very small
b Single study with n=24.31
c Studies of shorter duration showed less difference between groups which would be expected given the study design.
d The funnel plot suggests larger studies showed effects closer to the null. Further, studies that reported more protective effects tended to be older. This provides evidence of regression towards the mean.
e Studies that reported short-term pregnancy percentages would most likely also have access to calving percentages and abortion risks. However fewer studies reported calving percentages and abortion risks and all that did report, reported an outcome favorable to the vaccine. Selective reporting of vaccine favorable outcomes is therefore a possibility. (Note: several studies that reported pregnancy percentages euthanized animals before calving.) Other studies slaughtered animals prior to 180 days.
f Downgrade based on concerns about the validity of the data form-i.e SEM and standard deviations derived from time to event data.

** As the outcomes reported where not subjective i.e. infection detected by culture or pregnancy, the impact of failure to blind was considered minimal. The studies reported minimal loss-to-follow up. Failure to randomized was considered important however its impact was most likely to be seen as publication bias. The impact of multiplicity was thought also be manifest as selective reporting or publication bias.
CHAPTER III

A CRITICAL REVIEW AND META-ANALYSIS OF THE MAGNITUDE OF THE EFFECT OF ANTHELMINTIC USE ON STOCKER CALF PRODUCTION PARAMETERS IN NORTHERN US STATES

Submitted to the Journal of Veterinary Parasitology
P. Baltzell, T. Engleken, A. O’Connor

Introduction

Gastrointestinal parasitism is a leading cause of diminished health and productivity of grazing livestock in North America.\(^1\) Treatment with anthelmintics to control parasitism is currently recommended to reduce the impacts. However, the expected increase in performance from treatment, to control parasitism, is difficult to estimate. Although results from primary research are available, it is often difficult for veterinarians or producers to combine the results reported in multiple studies to help gauge the benefits and variation that could be expected. Knowing the expected magnitude of gain from a product, will enable producers and veterinarians the ability to compare different products’, as well as determine if the intervention worked as expected. Without knowledge of magnitude of gain and variation, producers and veterinarians are given only partial information upon which to base product selection decisions.

The family Trichostrongylidae, including the genera *Trichostrongylus, Haemoncus, Osterteragi, Cooperia,* and *Nematoderia,* cause most subclinical and clinical infections in beef cattle production systems in the United States.\(^2\) Most nematodes of importance to cattle production have six life stages; the egg, four larval stages, an immature adult stage, and a mature adult stage.\(^3\) A unique aspect of the nematode lifecycle is the ability to arrest larval development (hypobiosis) and therefore prolong stages of the life cycle. Nematodes can survive with arrested larval development of the L3 on the pasture or L4 larval stage in the animal.\(^2\) As hypobiosis is under the influence of season, the ecology of nematode infections and the impact on production
differs based on local climate. In northern climates of the United States, one study found that the L3 larvae of *Nematodirus spp.* survive as long as 23 months on pasture and *Ostertagia, and Cooperia spp.* survive for as long as 12 months on pasture, and is termed overwintering.\(^4\) In contrast, in southern climates of the United States, high temperatures and low rain fall in the summer reduce the survival of the L3 population and persistence of the parasite on pasture does not occur.\(^5\) It is known that anthelmintic effects are modified by climate; therefore, this review is limited to northern climates of the United States.

To reduce the impact of nematodes on cattle production, it is necessary to have control programs that aim to reduce worm burdens within cattle, and to prevent reinfection by reduction of parasites on pastures. Anthelmintic products are available to reduce burdens within cattle, and the timing of application impacts the burden of the pasture.\(^6\) A parasite control program is therefore composed of the product, the dose and the timing of administration.

The purpose of this critical review was to evaluate the magnitude of change in production outcomes associated with the use of anthelmintics in stocker cattle in northern climates of the United States. Further, the aim was to assess if season of administration or product type modified those outcomes. The review question was defined by four question components, as is standard for systematic reviews, using the PICO acronym: the population, the intervention, the comparison and the outcome.\(^7,8\) Populations of interest were stocker calves in the United States, where overwintering of parasite populations was likely, i.e., northern climates. The intervention and comparator of interest were any anthelmintic that is approved for use in the United States used at the labeled dose approved by the United States Food and Drug Administration. The outcomes of interest were any production outcomes for stocker calves such as, but not limited to, average daily gain and weight gain.
Methods and materials

Protocol and registration

A review protocol was generated before the start of the project. The protocol was not registered or published on a website as mechanisms for registrations did not exist at the time. No major modifications were made to the protocol other than those noted in the ancillary analysis. The protocol was established for all cattle in beef production systems. There were two classifications of cattle production systems that were studied separately. One group was stocker cattle and the second group was cow-calf pairs. This studies reports the conduct and results of the review for stocker cattle; however, the search criteria were the same for both groups. The review team had expertise in beef production and epidemiology and research synthesis. All members of the review team, except the expert in beef production, had been involved in systematic reviews previously.

Eligibility criteria

The population of interest, weaned beef calves (greater than 6 mo. of age or 400lbs), were considered for this review. In the United States this unique class of animal is referred to as stocker calves, and represents an animal that is intended for beef production but does not enter the feedlot directly after weaning. Instead this class of animals is raised for 6-9 months on pasture with or without supplementation. At the end of that period, these calves are transferred to the feedlot and receive a grain ration prior to slaughter. The population of interest was further limited to cattle naturally infected with nematodes and living on pasture, in a relevant state in the United States, for the duration of the study.

Any anthelmintic product and dose approved for use in cattle in the United States were eligible as intervention and/or comparator; comparators could also be non-treated controls.
Studies published before 1970 were not included in the review as anthelmintic products used prior 1970 differ from the products used after 1970; therefore, the results from such products were considered of little relevance to today’s producers.

Predetermined outcomes of interest were measures of daily gain (ADG), and measures of total weight gain. Although, the approach to measuring these could differ by study, the approach to measurement was not an exclusion criterion. Measures of nematode burden, such as eggs per gram (EPG), were not of interest as these do not directly measure a production outcome. Studies were only included in the review if published in English. Given the country of interest was the United States this was an unlikely an issue. Experimental design was not used as a criterion for inclusion other than the requirement for natural exposure to parasites.

Information sources

Databases were searched on March 23, 2013: CAB Abstracts (Thomson Reuters®, 1970-2013); BIOSIS Previews (Thomson Reuters®, 1970-2013); PubMed (1970-2013), and Agricola (EBSCO®, 1970-2013). WorldCat (FirstSearch, 1970-2013) was used to identify additional research in beef research reports. This database is the world's largest network of library content and it allows users to search library catalogues all over the world, including university libraries. United States based universities, especially land-grant institutions, commonly publish research reports, on a regular basis, to highlight research that has been done at the university but has not been published in a peer reviewed journal. These reports can be found in print, and some are available online.

Search

The search was designed with assistance from and information scientist with specialization in veterinary science. As stated previously, the original search was design to
identify two populations, grazing stocker cattle and grazing cows-calf pairs. The search contained terms that described the populations (animal type and state), products of interest, and outcomes of interest. The CAB Abstracts search was conducted March 23rd, 2013 using the following search string: “(Calf or calves or cow* or cattle) AND (anthelmintic* OR deworm* OR fenbendazole OR oxfendazole OR albendazole OR ivermectin OR eprinomectin OR doramectin OR moxidectin OR levamisole) AND (“weight gain” or “body weight” or “liveweight gain” or growth or “weaning weight” or weight* or gain* or conception or pregnancy or performance or “cost benefit analysis” or economics or “body condition” or liveweight or “feed conversion efficiency” OR “BCS” OR “body condition score” OR seasons OR “seasonal variation”) AND (USA OR “United States” OR Connecticut OR Delaware OR Illinois OR Indiana OR Iowa OR Kansas OR Kentucky OR Maryland OR Massachusetts OR Michigan OR Minnesota OR Missouri OR Nebraska OR New Jersey OR New York OR North Dakota OR Ohio OR Pennsylvania OR Rhode Island OR South Dakota OR Vermont OR Virginia OR West Virginia OR Wisconsin). Limits used in this search included publication dates limited to in or after 1970 and language limited to English. Analogous search terms and limits were used in the remaining three databases.

The search in WorldCat (FirstSearch, 1970-2013) was conducted on March 28th, 2013 using the following terms: report and (“experiment station” or extension) and (cattle or beef) and (state name or university name). As an example, the state name Indiana or university name Purdue would be put into the search. Indiana is a relevant state and Purdue University is the land grant university in that state. The search was repeated for every relevant state and land grant universities in the state listed in Table 1. A list of beef research reports was made and, due to time constraints, only research reports indexed online were evaluated for eligibility. The search
was also limited to research reports written in or after 1970. Titles from each online publication’s table of contents were examined for every year indexed.

All retrieved citations from both the databases and relevant research reports were imported into Endnote X6® (Thomson Reuters©, 2012) and duplicate citations based on title and author were removed. All citations from EndNote X6® were uploaded to Distiller SR® (Evidence Partners, Canada), a systematic review software program.

Study selection

Distiller SR® was used for manuscript screening and data extraction. Within Distiller SR® eligible screening forms were produced before manuscript screening took place. All forms were evaluated by reviewers prior to use, using example abstracts, to verify agreement about eligibility. Each title and abstract were screened (April 2013) by two reviewers independently, based on the predetermined eligibility criteria questions.

1. Does the study describe an assessment of anthelminthic product designed to control gastrointestinal parasites?
2. Is the study population grazing cattle with naturally occurring exposure to parasites?
3. Does the study describe primary research that compared at least one production outcome?
4. Does the study include two or more comparison groups?
5. Was the study conducted on grazing beef cattle in a relevant state based on title or text?

Possible responses to each question were, yes, no, or unclear. If an answer to a question was “no,” for both reviewers then the manuscript was excluded from further consideration and the remaining questions were not answered. When the response to all of the questions by both reviewers was either “yes” or “unclear” the manuscript was passed to the second screening of full text articles. Conflicts in answers to any question were resolved between the reviewers.
during a meeting. Full text screening for relevance was conducted (July 2013) by two reviewers independently. Additional eligibility screening questions at this 2nd level were used to include or exclude studies from the review.

1. Is the product used labeled for use in the United States?
2. Is the dose of anthelmintic recommended by the manufacturer?

As the full texts were available, possible responses were “yes” or “no”. Again, each question was answered in order on an identical form for each manuscript. If an answer to a question was “no,” for both reviewers the manuscript was excluded and the remaining questions were not answered. When the responses to all of these questions were “yes” the manuscript was included in the review. Any conflicts in answers to any question were resolved between the reviewers during a telephone meeting.

Data collection process

Data extraction was conducted, using Distiller SR®, on all relevant publications. The unit for data extraction was the study. If a manuscript contained more than one relevant study, a new form was used for each study. The form was piloted and amended several times before the final form was approved for use and data extraction began. Initial data extraction was conducted by the same reviewer for all relevant papers. The accuracy of the extracted data was verified by a second reviewer. Half of the studies were assigned to and verified by one reviewer and the other half was assigned to and verified by another reviewer.

Data items

Extracted information related to three levels: study level information, intervention group information and outcome level information. Study level information extracted included, the US state the study was conducted based on text, production group studied (i.e., stocker or weaned
calves, unweaned calves (included in a different review), cow calf pairs (included in a different review), sex, breed, age, weight, baseline worm burden, and the commingling of treatment groups on the same pasture. Intervention group information extracted was intervention used, route and frequency of administration, month of administration, interval between treatments, number of animals receiving intervention, age receiving intervention, weight of animals receiving intervention, and reported loss to follow-up for intervention. Outcome level information extracted included, the outcome reported, interval of time outcome was measured, number of animals assessed, and length of time post enrollment the outcome was measured. Intervention level point estimates for each outcome were gathered along with any measure of variation (SE or SD.) If a measure of variation was not reported, the p-value was extracted if reported. When standard error (SE) was reported for a group, it was converted to the standard deviation (SD) for use in the meta-analysis\textsuperscript{10} If weight was reported in pounds (lbs.) it was converted to kilograms (kg) for analysis and presentation.

Conflicts that arose in data verification were documented. The first reviewer was notified, and reviewed the conflict. If a change needed to be made the change was made then the second reviewer would verify the entry again. An example of the data extraction form is listed in Table 6.

Risk of bias in individual studies

The risk of bias in individual studies was assessed at the time of data extraction. The risk domains were selection bias, performance bias, detection bias, attrition bias, and other biases.\textsuperscript{10} Selection bias was assessed by noting if animals were placed in groups by a system of random sequence generation and if there was concealment of allocation of animals to group. Performance bias was determined by assessing if personnel were blinded to treatment groups and allocation of
treatments. Detection bias was assessed by observing if outcome assessors were blinded to
treatment. Attrition bias was evaluated by assessing loss to follow-up for each studied outcome.
Other biases were assessed in the “note” section where reviewers could add comments about any
other bias that was noted but not covered by other biases stated above.

Summary measures

The mean difference (MD) was used as the summary effect measure for continuous
outcomes. If ADG were greater in the intervention arm than the control arm it was expected that
the point estimate of the study mean difference to be positive (mean difference = treatment group
mean – control group mean).

Synthesis of results

For descriptive purposes, mean difference observed in studies that reported an
appropriate measure of variation was summarized using a forest plot where available.
For the outcome measure, mean difference, meta-analysis was performed using the statistical
software R\textsuperscript{11} using the “Meta” package\textsuperscript{12}. A random effects model was used. The null
hypothesis of no association means a value of zero for the summary mean difference.

Tests of heterogeneity (chi-square Q test) and measures of heterogeneity (I-squared
percentage) were calculated for each outcome. The Cochran Handbook for Systematic Reviews
guidelines\textsuperscript{10} were used to described the interpretation of the I-squared percentage as follows:

- 0% to 40%: might not be important;
- 30% to 60%: may represent moderate heterogeneity;
- 50% to 90%: may represent substantial heterogeneity;
- 75% to 100%: considerable heterogeneity.
Pre-planned subgroup analyses were based on a categorical variable that grouped the studies by the intervention used. Another pre-planned sub-group analysis was based on a categorical variable that grouped the studies by protocol used, i.e., given in the fall vs. given in the fall and spring. A summary effect measure was calculated for each subgroup as well as an overall effect using a random effects model.

After the meta-analysis was complete, to assess the influence of particular studies we conducted a sensitivity analysis. Studies that were believed to be influential were removed and the meta-analysis was re-run to determine if the heterogeneity changed. Furthermore, in our original meta-analysis, all studies used in the meta-analysis, regardless of if they came from the same manuscript or not, were treated as independent. To test if this assumption was true, ‘manuscript’ was added as a random effect in the model.

Risk of bias across study

Risk of bias across study was investigated using a funnel plot. The funnel plot was used to assess the risk of small study bias amongst the collected data. Asymmetry in the funnel plot was judged subjectively.

Additional analysis

After data analysis it was noticed that one study had a very different baseline and therefore the absolute difference may not be an appropriate effect measure. Therefore post hoc, we decided to repeat the analysis using the ratio of the means as the effect size. If ADG were greater in the intervention arm than the control arm it was expected that the point estimate of the study ratio of means would be greater than 1 (ROM = treatment group mean / control group mean). This meta-analysis was conducted in REVMAN. The aim of this post hoc analysis was
to determine if the effect measure changed the inference. This comparison would only be subjective, as there are no formal methods to compare different summary effect measures.

A further additional analysis was performed to describe the comprehensiveness of reporting, a checklist of items based on the REFLECT statement\(^8\) was used and modified for anthelmintic studies (Table 6). The REFLECT statement is a checklist designed to help authors comprehensively report clinical trials in food safety and animal health. Items in the checklist are designed to assess external and internal validity of each manuscript individually. A single reviewer completed the checklist for each article included the systematic review. Each item in the checklist was answered with a “yes” or a “no” depending on if the manuscript reported that item. A final tally or score of “yes” answers was given to each manuscript individually.

**Results**

**Study Selection**

The results of the search process are presented in Figure 1. From the four citation databases, 630 articles were gathered from the initial search. An additional five articles were identified in WorldCAT. Of the 635 total articles, 123 were duplicates; therefore, 512 citations were reviewed for eligibly, this included both cow/calf and stocker manuscripts. 459 did not meet the predetermined eligibly criteria and were excluded. The full texts of 53 articles were retrieved for further assessment. After evaluation of the full text, only 9 manuscripts and 23 studies were relevant to the review of production parameters for stocker calves.\(^{13-21}\) Of the 23 studies, only 14 were suitable for meta-analysis; however, all studies are summarized in tables. As shown in the flow chart, the stocker papers and the cow/calf manuscripts were collected together and they were split into categories after the final assessment. Assessment in this paper was only done for the stocker calf data.
Study characteristics

The characteristics of the nine relevant manuscripts are presented in Table 2. The data presented in Table 2 includes, state of conduct, number of farms, the interventions assessed, and duration of study. ADG was reported in seven manuscripts \(^{13-16,18,20,21}\) with 20 separate studies. Weight gain was reported in three manuscripts with eight studies. \(^{16,17,19}\) All three manuscripts failed to report measures of variation; therefore, a meta-analysis could not be performed using weight gain as an outcome.

Risk of bias

The detailed results of risk of bias assessment are provided in Table 5. For most studies included in this review, the risk of bias was mostly unclear. All studies reported random sequence generation with most studies except two using a random sequence generation scheme. Otherwise, reporting was poor; allocation concealment was not described in all studies with one exception, blinding of personnel was not described in any study, blinding of outcome assessment was not described in any study with the exception of one study and incomplete outcome data was only reported in two manuscripts. Poor reporting of study details hindered the assessment of risk.

Risk of bias across the study is shown in the funnel plot (figure 3). Risk of bias across study or publication bias subjectively does not seem to be a problem in this dataset, as the funnel appears symmetrical with both positive and negative results being reported. There are some extreme observations. This occurred for studies that had decreased precision of their study, likely smaller sample sizes, but they are symmetrically distributed.

Descriptive Outcomes and Meta-analysis

Information about observed mean difference for ADG for each study is found in Table 3 and Figure 2. Four of the seven manuscripts reporting ADG did not report any measures of
precision\textsuperscript{13,16,20,21}. These four manuscripts included six studies that could not be included in the meta-analysis. The overall magnitude of the mean difference of ADG on anthelmintic use in northern climates of the United States was an advantage of 0.05 kg per day with a 95% CI of 0.03-0.07, with an overall p-value of $<0.0001$. Slight heterogeneity was observed (Figure 2, $I$-squared = 28.8% and $\tau$-squared = 0.0004, p-value = 0.1477) with these data.

Sub-group analysis by anthelmintic product was performed for ADG. The subgroups analyzed were ivermectin and long acting moxidectin. Other reported products were doramectin and thiabendazole but those were not included in the meta-analysis because the authors did not report measures of precision such as standard errors or standard deviations. The overall mean difference for the subgroup ivermectin was 0.05 kg/day with a 95% confidence interval (CI) of 0.03 to 0.06. No evidence of heterogeneity within the subgroup was found ($I$-squared = 0.0% and $\tau$-squared = 0.0, p-value = 0.64). The overall mean difference for the subgroup long-acting moxidectin was 0.13 kg/day with a 95% confidence interval (CI) of 0.04 to 0.23. Moderate to substantial subgroup heterogeneity was noted ($I$-squared = 62.4% and $\tau$-squared = 0.0031, p-value = 0.1031). Based on the sensitivity analysis, the heterogeneity noted can be attributed to one manuscript.\textsuperscript{14} The test for subgroup differences had moderate evidence of an effect with a p-value of 0.08. Sub-group analysis by protocol could not be performed as all manuscripts in this review used similar protocols, deworming one time in the spring.

Additional analysis

When the meta-analysis was performed using the ratio of means, there was no difference in conclusion as compared to using mean difference. Therefore, we have only included results of the mean difference analysis and will use that to draw conclusions. Similarly, when the meta-
analysis was conducted using manuscript as a random variable, there was no difference in the summary effect measure.

Comprehensiveness of Reporting

The results of the assessment of comprehensiveness of reporting are presented in Table 6. As is obvious from the outcome of meta-analysis reporting in this body of work was often not comprehensive. Missing information was often found in the materials and methods. Studies often omitted settings and locations that data was collected, stating primary and secondary objectives of the study, how sample size was determined, who implemented random allocation, who generated the allocation sequence, who enrolled study units, and who assigned study units to their groups, and whether or not those administering the interventions, caregivers and those assessing the outcomes were blinded to group assignment. In the results, missing information that was most detrimental to meta-analysis was, for each group reporting the number of study units that were randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. No manuscripts included were able to report all of these criteria.

Discussion

One goal of this review was to summarize the magnitude of gain associated with use of anthelmintic products and to determine if there was empirical evidence that treating beef cattle in northern climates of the United States, with different anthelmintic protocols (including season of administration) had an effect on production outcomes such as ADG or weight gain. Our rationale for assessing these outcomes was that improved weight gain or ADG is often reported as an outcome in anthelmintic studies without discussion of seasonal timing of interventions, which we believe may have an impact on the magnitude of effect. We were unable to perform the sub-
group meta-analysis for differing timing of protocols because the studies included in the meta-analysis had similarly timed interventions. With respect to this aim, we were unable to find studies that would provide evidence of the magnitude of effect that differently timed protocols have on ADG and weight gain in the northern climates evaluated.

A second purpose was to determine if there was evidence of differing interventions (i.e., anthelmintic product used), in the northern climates, affecting weight gains or ADG. This analysis was performed for the outcome ADG only. The only two products that were available for meta-analysis were ivermectin and long-acting moxidectin. There was moderate to substantial heterogeneity found in the sub-group long-acting moxidectin, from a single manuscript with two studies. Further, there was a difference in the sex of the animals used for the two studies. The study with the larger magnitude of effect (MD=0.19 kg/day), included only steers whereas the other study with a smaller magnitude of effect (MD=0.9 kg/day) included both steers and heifers. The manuscript stated that researchers were aware that both heifers and steers were included at the site but sex was not considered during assignment of animals to treatment groups. This finding motivated the assessment of ratio of means as the outcome. RoM is used when the outcome measured is measured on a physical scale such as weight and when that scale is unlikely to be zero as is true in this case. Additionally, RoM can be helpful for interpretation of clinical data.

We hypothesized that, as mean difference as the absolute difference in outcome might be misleading to compare heifers to steers, as steers are known have high gains. Perhaps, the proportion of gain for treated animals relative to the control group would be the same for heifers and steers, even if absolute gain was not. The results of the analysis did not bear this out. Using the ratio of means, the effect of the intervention was greater in the steers versus the effect in
heifers. It was not possible to formally assess sex as a source of heterogeneity because most studies used mixed populations.

Conducting a network meta-analysis was discussed prior to the review process as a goal for this particular body of work. MTC meta-analyses answer clinical questions in situations where many treatment regimens already exist, and compare each treatment to rank their benefits to help choose the best option. In respect to this review, it was not possible to perform an MTC for the production outcome ADG. One of the two interventions (long-acting moxidectin), with a single manuscript and two studies, had moderate to substantial heterogeneity in the sub-group analysis (I-squared 62.4%). Therefore, it was not advisable to perform indirect comparisons with the other intervention (ivermectin) in an MTC meta-analysis format.

One major limitation common throughout this review was the ability to extract the effect size and measures of precision from the studies for a meta-analysis. When this is the case, it may be tempting to use “p-value vote counting” as a method of combining the data. Vote counting involves counting studies that did or did not find a significant effect and reaching a conclusion based on the majority vote. Such an approach to data synthesis is deeply flawed. Vote counting ignores the power of the study to detect an outcome, as the p-value is a confounded variable that is a function of both sample size and the magnitude of effect. Further, vote counting is a method that does not take into account the different weights given to each study, as is seen with a meta-analysis. Consequently, the only remaining approach to addressing the review question is a narrative summary of individual studies that were conducted for the comparison of interest. All studies that failed to report SEM for their findings for both outcomes of ADG and weight gain reported either no difference in gain or increased gains for treated animals when compared to
control animals. Some of these studies reported statistical significance at the level of 0.05, others did not.

The comprehensiveness of reporting was poor from this body of work, a consistent finding with other areas in veterinary and animal science.\textsuperscript{27-29} For many of the REFLECT checklist items, it is perhaps not surprising that publications fail to include the information about implementation of randomization, blinding and sample size justification. Although the importance of including this information has long been known, scrutiny of the reporting of studies is a more recent phenomenon. However, it was surprising how many papers failed to provide basic information such as the number of animals in each group and the measures of precision. Of course, these absences limited the ability to conduct meta-analyses, and the absence of this information limits the ability of any end-user to assess the validity of the studies. Factors that affect external validity, such as housing and the location and timing of the study were well reported. Given the clear importance of climate on the value of a deworming intervention, this means that producers and veterinarians should be able to assess the relevance of primary research to their setting, even if the internal validity cannot be assessed.

To conclude, the results of the review suggest that there is evidence that anthelmintic use has an effect on ADG in beef cattle production systems in a northern climate of the United States. We were able to conclude this based upon the data in the forest plot (figure 2); the overall mean difference for the intervention remained greater than zero, indicating a positive difference with respect to average daily gain. The use of increased ADG as a production outcome to decide to deworm cattle in this setting is appropriate. A definitive conclusion cannot be made when using the production outcome of increased weight gain as a parameter to decide to deworm cattle; as a meta-analysis could not be performed. From the data given, it is presumed that weight
gain is affected to some degree by anthelmintic use but, the magnitude of effect cannot be
determined. This conclusion points out that better reporting throughout the current studies would
have been needed to fully understand the magnitude of effect anthelmintic interventions for
improved ADG and weight gains in beef cattle in northern climates of the United States.
Furthermore, additional research would be necessary to draw conclusions regarding the timing
anthelmintic interventions as well as a ranking of different anthelmintic products.

References

   Practice. 2006;22(2):451+-.
2. Yazwinski TA, Tucker CA. A sampling of factors relative to the epidemiology of
gastrointestinal nematode parasites of cattle in the United States. Veterinary Clinics of
North America-Food Animal Practice. 2006;22(3):501+-.
4. Gibbs HC. PERSISTENCE ON PASTURE OF THE INFECTIVE LARVAE OF
   NEMATODES PARASITIZING MAINE DAIRY-CATTLE. American Journal of
   Veterinary Research. 1980;41(10):1694-1695.
5. Williams JC, Bilkovich FR. Development and Survival of Infective Larvae of the Cattle
7. Authority EFS. Application of Systematic Review Methodology to Food and Feed Safety
   2010;8(6):90.
8. O'Connor AM, Sargeant JM, Gardner IA, Dickson JS, Torrence ME. The REFLECT
   Statement: Methods and Processes of Creating Reporting Guidelines for Randomized
   Controlled Trials for Livestock and Food Safety. Journal of Veterinary Internal
   2014.
11. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria:


Figures

Figure 1. Database and research report search results
Figure 2. Forest plot for mean difference of ADG (kg) in anthelmintic treated stocker cattle vs. untreated controls
Figure 3. Funnel plot of studies included in meta-analysis for ADG of stocker calves (mean difference)
### Tables

**Table 1. Relevant states included in the review**

<table>
<thead>
<tr>
<th>Connecticut</th>
<th>Delaware</th>
<th>Illinois</th>
<th>Indiana</th>
<th>Iowa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kansas</td>
<td>Kentucky</td>
<td>Maryland</td>
<td>Massachusetts</td>
<td>Michigan</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Missouri</td>
<td>Nebraska</td>
<td>New Jersey</td>
<td>New York</td>
</tr>
<tr>
<td>North Dakota</td>
<td>Ohio</td>
<td>Pennsylvania</td>
<td>Rhode Island</td>
<td>South Dakota</td>
</tr>
<tr>
<td>Vermont</td>
<td>Virginia</td>
<td>West Virginia</td>
<td>Wisconsin</td>
<td></td>
</tr>
<tr>
<td>Manuscript</td>
<td>State</td>
<td># of Farms</td>
<td>Intervention A</td>
<td>Intervention B</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>------------</td>
<td>----------------------------------------------------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>Ballweber et al 1997</td>
<td>WI</td>
<td>1</td>
<td>doramectin at 200mcg/kg body weight SQ</td>
<td>ivermectin at 200mcg/kg body weight SQ</td>
</tr>
<tr>
<td>Cleale et al 2004</td>
<td>IL, &amp; WI</td>
<td>2</td>
<td>10% Long-Acting Injectable moxidectin at 0.5mL/50kg body weight SQ</td>
<td>N/A</td>
</tr>
<tr>
<td>Epperson et al 2001</td>
<td>SD</td>
<td>1</td>
<td>Ivomec® SR bolus at 1 bolus/275-600Lbs</td>
<td>N/A</td>
</tr>
<tr>
<td>Ferguson et al 1971</td>
<td>NE</td>
<td>3</td>
<td>thiabendazole bolus at 3-5g/100Lbs</td>
<td>thiabendazole bolus at 3-5g/100Lbs; 2 times ~5 months apart (d –170, day 0) – only study 2</td>
</tr>
<tr>
<td>Kunkle et al 2013</td>
<td>MO, MN, WI</td>
<td></td>
<td>5% eprinomectin ERI at 1mL/50kg SQ</td>
<td>N/A</td>
</tr>
<tr>
<td>Mertz et al 2005</td>
<td>SD</td>
<td>11</td>
<td>Ivomec® SR bolus at 1 bolus/275-600Lbs</td>
<td>N/A</td>
</tr>
<tr>
<td>Rehbien et al 2013</td>
<td>MO</td>
<td>1</td>
<td>5% eprinomectin ERI at 1mL/50kg SQ</td>
<td>N/A</td>
</tr>
<tr>
<td>Skogerboe et al 2000</td>
<td>WI</td>
<td>1</td>
<td>0.5% doramectin pour-on solution at 500mcg/kg</td>
<td>N/A</td>
</tr>
<tr>
<td>Smith et al 1973</td>
<td>KS</td>
<td>1</td>
<td>thiabendazole bolus at 3-5g/100Lbs</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 3. Average daily gain (kg/day (SEM/SD)) reported for stocker calves

<table>
<thead>
<tr>
<th>Manuscript and Farm (where applicable)</th>
<th>Intervention A*</th>
<th>Intervention B*</th>
<th>Intervention C*</th>
<th>Control group*</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballweber et al 1997</td>
<td>0.93 §</td>
<td>0.92 §</td>
<td>0.90 §</td>
<td>0.77 §</td>
<td>&lt;0.05*</td>
</tr>
<tr>
<td>Cleale et al 2004</td>
<td>0.41 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.32 (0.03)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Illinois</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleale et al 2004</td>
<td>0.94 (0.03)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.75 (0.04)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epperson et al 2001</td>
<td>0.60 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.54 (0.02)</td>
<td>0.02</td>
</tr>
<tr>
<td>Ferguson et al 1971</td>
<td>0.18 §</td>
<td>N/A</td>
<td>N/A</td>
<td>0.13 §</td>
<td>not reported</td>
</tr>
<tr>
<td>Trial 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferguson et al 1971</td>
<td>1.15 §</td>
<td>1.17 §</td>
<td>N/A</td>
<td>1.15 §</td>
<td>not reported</td>
</tr>
<tr>
<td>Trial 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferguson et al 1971</td>
<td>0.60 §</td>
<td>N/A</td>
<td>N/A</td>
<td>0.59 §</td>
<td>not reported</td>
</tr>
<tr>
<td>Trial 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mertz et al 2005 Site A 1999</td>
<td>0.75 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.72 (0.02)</td>
<td>0.26</td>
</tr>
<tr>
<td>Mertz et al 2005 Site B 1999</td>
<td>0.85 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.82 (0.03)</td>
<td>0.36</td>
</tr>
<tr>
<td>Mertz et al 2005 Site C 1999</td>
<td>0.67 (0.03)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.64 (0.03)</td>
<td>0.40</td>
</tr>
<tr>
<td>Mertz et al 2005 Site A 2000</td>
<td>0.72 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.64 (0.01)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mertz et al 2005 Site D 2000</td>
<td>0.5 (0.03)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.44 (0.02)</td>
<td>0.07</td>
</tr>
<tr>
<td>Mertz et al 2005 Site E 2000</td>
<td>1.01 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.99 (0.02)</td>
<td>0.39</td>
</tr>
<tr>
<td>Mertz et al 2005 Site F 2000</td>
<td>0.88 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.80 (0.02)</td>
<td>0.004</td>
</tr>
<tr>
<td>Mertz et al 2005 Site G 2000</td>
<td>0.52 (0.05)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.55 (0.03)</td>
<td>0.60</td>
</tr>
<tr>
<td>Mertz et al 2005 Site H 2000</td>
<td>0.56 (0.04)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.53 (0.02)</td>
<td>0.41</td>
</tr>
<tr>
<td>Mertz et al 2005 Site I 2000</td>
<td>0.63 (0.04)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.58 (0.02)</td>
<td>0.25</td>
</tr>
<tr>
<td>Mertz et al 2005 Site A 2001</td>
<td>0.53 (0.02)</td>
<td>N/A</td>
<td>N/A</td>
<td>0.51 (0.01)</td>
<td>0.38</td>
</tr>
<tr>
<td>Skogerboe et al 2000</td>
<td>1.00 §</td>
<td>N/A</td>
<td>N/A</td>
<td>0.89 §</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Smith et al 1973</td>
<td>0.51 §</td>
<td>N/A</td>
<td>N/A</td>
<td>0.51 §</td>
<td>not reported</td>
</tr>
</tbody>
</table>

* For description of intervention A, B, C, and Control group see Table 2

≠ P-value <0.05 for treated vs. control

§ SEM/SD not reported
Table 4. Weight gain (kg (SEM/SD)) reported for stocker calves

<table>
<thead>
<tr>
<th>Manuscript and Farm (where applicable)</th>
<th>Intervention A*</th>
<th>Intervention B*</th>
<th>Control group*</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferguson et al 1971 Trial 1</td>
<td>25.5 §</td>
<td>N/A</td>
<td>18.6 §</td>
<td>not reported</td>
</tr>
<tr>
<td>Ferguson et al 1971 Trial 3</td>
<td>116.8 §</td>
<td>118.6 §</td>
<td>116.4 §</td>
<td>not reported</td>
</tr>
<tr>
<td>Ferguson et al 1971 Trial 4</td>
<td>75.5 §</td>
<td>N/A</td>
<td>75.0 §</td>
<td>not reported</td>
</tr>
<tr>
<td>Kunkle et al 2013 Missouri</td>
<td>72.1 §</td>
<td>N/A</td>
<td>48.5 §</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Kunkle et al 2013 Minnesota</td>
<td>112.2 §</td>
<td>N/A</td>
<td>90.2 §</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Kunkle et al 2013 Wisconsin</td>
<td>79.9 §</td>
<td>N/A</td>
<td>73.4 §</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Rehbein 2013</td>
<td>60.6 §</td>
<td>N/A</td>
<td>37.6 §</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>

* For description of intervention A, B, and Control group see Table
§ SEM/SD not reported
Table 5. Risk of Bias

<table>
<thead>
<tr>
<th>Manuscript</th>
<th>Random sequence generation (selection bias)</th>
<th>Allocation concealment (selection bias)</th>
<th>Blinding of personnel (performance bias)</th>
<th>Blinding of outcome assessment (detection bias)</th>
<th>Incomplete outcome data (attrition bias)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballweber et al 1997</td>
<td>Described – Yes</td>
<td>Not described ?</td>
<td>Not described ?</td>
<td>Not described ?</td>
<td>Not reported ?</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleale et al 2004</td>
<td>Described-Yes</td>
<td>Not described ?</td>
<td>Not described ?</td>
<td>Described – Yes</td>
<td>Reported – No</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Epperson et al 2001</td>
<td>Described – Yes</td>
<td>Described – Yes</td>
<td>Not Described</td>
<td>Not Described</td>
<td>Reported – Yes</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
<td>–</td>
</tr>
<tr>
<td>Ferguson et al 1971</td>
<td>Described – Not random</td>
<td>Not Described</td>
<td>Not Described</td>
<td>Not Described</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mertz et al 2005</td>
<td>Described – Yes</td>
<td>Not Described</td>
<td>Not described</td>
<td>Not described</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skogerboe et al 2000</td>
<td>Described – Yes</td>
<td>Not described</td>
<td>Not described</td>
<td>Not described</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith et al 1973</td>
<td>Described – not random</td>
<td>Not described</td>
<td>Not described</td>
<td>Not described</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

+ low risk of bias   ?unclear risk of bias   – high risk of bias
Table 6. Checklist items for the assessment of comprehensiveness of reporting

<table>
<thead>
<tr>
<th>Rationale or explanation</th>
<th>Frequency Yes/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title and Abstract</strong></td>
<td></td>
</tr>
<tr>
<td>1. How study units were allocated to interventions. Clearly state whether the outcome</td>
<td>4/9</td>
</tr>
<tr>
<td>was the result of natural exposure or was the result of a deliberate agent challenge.</td>
<td></td>
</tr>
<tr>
<td><strong>Methods and Materials</strong></td>
<td></td>
</tr>
<tr>
<td>2. Eligibility criteria for owner/managers and study units at each level of the</td>
<td>1/9</td>
</tr>
<tr>
<td>organizational structure. The settings and locations where the data were collected.</td>
<td></td>
</tr>
<tr>
<td>3. Precise details of the interventions intended for each group, the level at which the</td>
<td>9/9</td>
</tr>
<tr>
<td>intervention was allocated, and how and when interventions were actually administered.</td>
<td></td>
</tr>
<tr>
<td>4. Specific objectives and hypotheses. Clearly state primary and secondary objectives.</td>
<td>2/9</td>
</tr>
<tr>
<td>5. Clearly defined primary and secondary outcome measures and the levels at which they</td>
<td>3/9</td>
</tr>
<tr>
<td>were measured, and, when applicable, any methods used to enhance the quality of</td>
<td></td>
</tr>
<tr>
<td>measurements.</td>
<td></td>
</tr>
<tr>
<td>6. How sample size was determined and, when applicable, explanation of any interim</td>
<td>0/9</td>
</tr>
<tr>
<td>analyses and stopping rules. Sample-size considerations should include sample-size</td>
<td></td>
</tr>
<tr>
<td>determinations at each level of the organizational structure and the assumptions used</td>
<td></td>
</tr>
<tr>
<td>to account for any non-independence among groups or individuals within a group.</td>
<td></td>
</tr>
<tr>
<td>7. Method used to generate the random allocation sequence at the relevant level of the</td>
<td>7/9</td>
</tr>
<tr>
<td>organizational structure, including details of any restrictions (eg, blocking,</td>
<td></td>
</tr>
<tr>
<td>stratification)</td>
<td></td>
</tr>
<tr>
<td>8. Method used to implement the random allocation sequence at the relevant level of the</td>
<td>0/9</td>
</tr>
<tr>
<td>organizational structure, (i.e., numbered containers), clarifying whether the</td>
<td></td>
</tr>
<tr>
<td>sequence was concealed until interventions were assigned.</td>
<td></td>
</tr>
<tr>
<td>9. State who generated the allocation sequence, who enrolled study units, and who</td>
<td>0/9</td>
</tr>
<tr>
<td>assigned study units to their groups at the relevant level of the organizational</td>
<td></td>
</tr>
<tr>
<td>structure.</td>
<td></td>
</tr>
<tr>
<td>10. Whether or not those administering the interventions, caregivers and those</td>
<td>1/9</td>
</tr>
<tr>
<td>assessing the outcomes were blinded to group assignment. If done, how the success</td>
<td></td>
</tr>
<tr>
<td>of blinding was evaluated. Provide justification for not using blinding if it was</td>
<td></td>
</tr>
<tr>
<td>not used.</td>
<td></td>
</tr>
<tr>
<td>11. Statistical methods used to compare groups for all outcome(s); Clearly state the</td>
<td>6/9</td>
</tr>
<tr>
<td>level of statistical analysis and methods used to account for the organizational</td>
<td></td>
</tr>
<tr>
<td>structure, where applicable; methods for additional analyses, such as subgroup</td>
<td></td>
</tr>
<tr>
<td>analyses and adjusted analyses.</td>
<td></td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
</tr>
<tr>
<td>12. Flow of study units through each stage for each level of the organization structure</td>
<td>0/9</td>
</tr>
<tr>
<td>of the study (a diagram is strongly recommended). Specifically, for each group,</td>
<td></td>
</tr>
<tr>
<td>report the numbers of study units randomly assigned, receiving intended treatment,</td>
<td></td>
</tr>
<tr>
<td>completing the study protocol, and analyzed for the primary outcome. Describe</td>
<td></td>
</tr>
<tr>
<td>protocol deviations from study as planned, together with reasons.</td>
<td></td>
</tr>
<tr>
<td>13. Dates defining the periods of recruitment and follow-up.</td>
<td>5/9</td>
</tr>
<tr>
<td>14. Baseline demographic and clinical characteristics of each group, explicitly</td>
<td>2/9</td>
</tr>
<tr>
<td>providing information for each relevant level of the organizational structure. Data</td>
<td></td>
</tr>
<tr>
<td>should be reported in such a way that secondary analysis, such as risk assessment,</td>
<td></td>
</tr>
<tr>
<td>is possible.</td>
<td></td>
</tr>
<tr>
<td>15. Number of study units (denominator) in each group included in each analysis and</td>
<td>7/9</td>
</tr>
<tr>
<td>whether the analysis was by &quot;intention-to-treat.&quot; State the results in absolute</td>
<td></td>
</tr>
<tr>
<td>numbers when feasible (i.e., 10/20, not 50%).</td>
<td></td>
</tr>
<tr>
<td>16. For each primary and secondary outcome, a summary of results for each group,</td>
<td>4/9</td>
</tr>
<tr>
<td>accounting for each relevant level of the organizational structure, and the</td>
<td></td>
</tr>
<tr>
<td>estimated effect size and its precision (i.e., 95% confidence interval)</td>
<td></td>
</tr>
<tr>
<td>17. Address multiplicity by reporting any other analyses performed, including subgroup</td>
<td>0/9</td>
</tr>
<tr>
<td>analyses and adjusted analyses, indicating those pre-specified and those exploratory.</td>
<td></td>
</tr>
<tr>
<td>18. All important adverse events or side effects in each intervention group.</td>
<td>3/9</td>
</tr>
</tbody>
</table>
CHAPTER IV
SUMMARY AND CONCLUSION

This thesis introduced systematic reviews for interventions, why they are used, and how they can benefit veterinary scientific research. Two examples were presented. The first discussed using GRADE for quality assessment of outcomes for systematic reviews and meta-analyses. A critical review and meta-analysis was presented on the topic of vaccine efficacy for the disease Tritrichomonas foetus in beef cattle. This review used GRADE to help determine the quality of the outcomes that were used for this review and meta-analysis. The findings were reported in summary tables. The overall conclusion of this review was that based on this review, there is a lack of conclusive evidence to support the use of this vaccine in areas where good biosecurity practices are in place. However, the review directed readers to the evidence tables so that they might make their own decision about its use depending on their unique situation. For example, some veterinarians still may recommend the vaccine in settings where biosecurity is difficult and may elect to use vaccine, as the circumstances differ. Such a judgment could only be reached if the veterinarian placed lower value on the small study sizes, quality of the evidence and sources of bias than we did, and placed higher value of the estimates of effect. Additionally, this review helped to point out that there were relatively few studies in cows and bulls and therefore the efficacy of the vaccine in these groups of cattle could not be assessed, and may serve as an area of future clinical research.

The second example given in this thesis describes a novel approach in meta-analysis to comparing different treatments for the same disease or problem, indirectly, called a MTC meta-analysis. Only a few MTC meta-analysis have been published in veterinary medicine. An MTC
would have been performed for the critical review and meta-analysis that was presented, as there were several treatments utilized that were only compared with control (no treatment). This unfortunately could not be done, due to missing information in most of the manuscripts. Without this information these studies were not included in the meta-analysis and therefore left only two different treatments to compare. The overall conclusion of this review was that the results suggest that there is evidence that anthelmintic use has an effect on ADG in beef cattle production systems in a northern climate of the United States and that no conclusion could be made on weight gain as a meta-analysis could not be conducted due to poor reporting. This conclusion points out that better reporting throughout the current studies would have been needed to fully understand the magnitude of effect anthelmintic interventions for improved ADG and weight gains in beef cattle in northern climates of the United States. Furthermore, additional research would be necessary to draw conclusions regarding the timing anthelmintic interventions as well as a ranking of different anthelmintic products.

The two critical reviews and meta-analyses that were presented in the above sections of this thesis illustrate the need for quality primary research and comprehensive reporting of primary research. In both critical reviews and meta-analyses presented, the issues with the data included very poor quality reporting, missing information and an overall chaos that made it difficult at best to conduct this type of research synthesis. In Chapter II, the use of GRADE assessed study quality of the included literature. Quality ranged from moderate to very low with most of the studies included being low or very low quality. In Chapter III, this review had planned on performing a MTC meta-analysis and was unable to do so due to a lack of reporting measures of precision such as standard error or standard deviation as well as substantial heterogeneity within one intervention. These are only two examples, however, it could be
speculated that this would not be the only area of veterinary research that would have these same issues.

Veterinary journals should follow suit with human medical journals in encouraging and/or requiring systematic reviews as part of any primary research about interventions that is submitted. As was mentioned previously this can help with focusing research into the areas where it is needed most, to help reduce wasted time repeating studies that have already been done.

Additionally, there are a few key places where veterinary researchers can help to make literature more usable for systematic reviews. For example, veterinary researchers should be taking the time to properly report measures of precision such as standard errors or standard deviations. This information greatly helps with the ability of the end-user to evaluate the point estimate given for an outcome and subsequently is necessary for meta-analysis. Also, transparency in the materials and methods so that extraction of data such as treated and control populations, sex of animals used, or number lost to follow-up, etc. are easily done. Finally, researchers should take into consideration prior to starting their research how they could minimize biases and attempt to do so throughout the trial. High quality, and transparent studies are much easier to include in systematic reviews and meta-analyses as well as obtain much more useful information.

As the human medical community has already found out, systematic reviews are an important part of medical literature. It seems important for the veterinary medical community to also see these benefits and begin to incorporate systematic reviews more often into veterinary research and to look to improve quality of research over quantity of research.