Microbiology Spectrum
American Society for Microbiology Press

Potential for Meta-Analysis in the Realm of Preharvest Food Safety

JAN M. SARGEANT[1] and ANNETTE M. O'CONNOR[2]

[1]Center for Public Health and Zoonoses and Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, ON, Canada N1G 2W1; [2]Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University College of Veterinary Medicine, Ames, IA 50100

ABSTRACT Meta-analysis, the statistical combination of results from multiple studies, can be used to summarize all of the available research on an intervention, etiology, descriptive, or diagnostic test accuracy question. Meta-analysis should be conducted as a component of a systematic review, to increase transparency in the selection of studies and to incorporate an evaluation of the risk of bias in the individual studies included in the meta-analysis. The process of meta-analysis may include a forest plot to graphically display the study results and the calculation of a weighted average summary effect size. Heterogeneity (differences in the effect size between studies) can be evaluated using formal statistics and the reasons for heterogeneity can be explored using sub-group analysis or meta-regression. Thus, meta-analysis may be a useful methodology for preharvest food safety research to aid in policy or clinical decision-making or to provide input to quantitative risk assessment or other models.

INTRODUCTION

Meta-analysis refers to the statistical combination (pooling) of data from multiple original research studies. The results from different studies on the same topic can vary, and meta-analysis provides a means of summarizing a parameter or effect across studies to develop a more precise estimate of the outcome of interest (1). The combination of data from multiple studies can be undertaken using two broad approaches: combining individual-level data from multiple studies or combining study-level results (effect sizes) from multiple studies (1). The former requires the meta-analyst to have access to all of the original data from each study subject for each study and is therefore not commonly seen in the preharvest food safety literature (2), although databases of microbial growth and inactivation kinetics for foodborne pathogens are available and growing (3). Therefore, this chapter will focus on meta-analysis in the context of combining effect sizes from multiple studies to calculate a summary effect size.

Meta-analyses should be conducted as the statistical component of a systematic review. The systematic review methodology aims to provide a transparent and comprehensive summary of the scientific evidence for a specific clinical or policy question. Nesting a meta-analysis within a systematic review helps to ensure that all available literature is considered in the summary effect size estimate. Further, the systematic review methodology explicitly includes an assessment of the risk of bias in the included studies, thereby reducing the potential for biased data to influence the results of the meta-analysis.

The advantages of meta-analysis include an increased sample size for estimating a parameter or an effect size and therefore greater precision in the estimate; the confidence intervals around a summary estimate can be calculated and displayed; sensitivity analysis can be conducted to explore the contribution of specific studies to the summary estimate; individual study results may be weighted by the precision of the effect size estimate (often partly a function of sample size); and graphical

Received: 25 August 2014, **Accepted:** 17 December 2015, **Published:** 16 September 2016

Editors: Kalmia Kniel, Department of Animal and Food Science, University of Delaware, Newark, DE; Siddhartha Thakur, North Carolina State University, College of Veterinary Medicine, Raleigh, NC

Citation: Sargeant JM, O'Connor AM. 2016. Potential for meta-analysis in the realm of preharvest food safety. *Microbiol Spectrum* 4(5):PFS-0004-2014. doi:10.1128/microbiolspec.PFS-0004-2014.

Correspondence: Jan M. Sargeant, sargeanj@uoguelph.ca

methods are available for displaying the results of the individual studies contributing to the summary estimate (4). Technically, a meta-analysis can be conducted if at least two studies with the same outcome are included in the review, although larger sample sizes increase the utility of a meta-analysis.

A meta-analysis can address a number of types of questions. These fall into four general categories: descriptive questions (e.g., prevalence or incidence), intervention questions, exposure questions, and questions about diagnostic test accuracy (4). All of these question types are of potential relevance to preharvest food safety. The results of meta-analyses could be used to inform clinical or policy decision-making (5) or could be used as inputs to quantitative risk assessment or other models (3, 6, 7).

Examples of published meta-analyses for intervention or exposure questions related to preharvest food safety include evaluations of the effect of direct-fed microbials on fecal shedding of *Escherichia coli* O157 in cattle (8), of vaccination on fecal shedding of *E. coli* O157 in cattle (9, 10), and of interventions to reduce *Salmonella* in broiler chickens (11–13) and in swine (14) and comparisons between the prevalence of enteric pathogens in organic and traditional production systems (15). These meta-analyses provide a summary of the totality of information that is available to address each review question. Therefore, a policy person, veterinarian, producer, or manufacturer can read one meta-analyses to obtain the scientific information needed to make a decision on implementing a preharvest intervention rather than having to read all of the primary literature. Examples of meta-analyses related to a descriptive question in preharvest food safety include an evaluation of factors influencing the prevalence of *Salmonella* on swine farms (16) and a study to estimate the prevalence of *E. coli* O157 in cattle (17). The latter meta-analysis also explored possible explanations for differences in the prevalence of *E. coli* O157, such as region, cattle type, and diagnostic method. For diagnostic test accuracy questions, meta-analysis has been used to evaluate differences in the diagnostic accuracy of culture versus PCR methods for detection of *Salmonella* in swine (18).

META-ANALYSIS IN THE CONTEXT OF SYSTEMATIC REVIEWS

Ideally, meta-analyses are conducted as the statistical component of a systematic review. Not all systematic reviews include a meta-analysis due to an insufficient quantity of data addressing the same question using the same outcome. However, including a systematic review approach increases the transparency and rigor of a meta-analysis.

A number of resources are available that describe the process of systematic review and meta-analysis in the context of preharvest food safety (2, 4, 19–23). The steps of a systematic review include defining the systematic review question, searching for studies, selecting relevant studies based on eligibility criteria, extracting data from relevant studies, assessing the risk of bias in the selected studies, synthesizing the results qualitatively or quantitatively (meta-analysis), and presenting and interpreting the results (4).

A systematic review question is built around key elements, which should be defined in the review question (4). The definition of the key elements, and associated eligibility criteria, set the scope of the systematic review and, subsequently, the meta-analysis. The broad example of *Salmonella* in swine is used for illustration. For a descriptive question, the key elements are the population and the outcome. Thus, a systematic review question might be "What is the prevalence of *Salmonella* (outcome) in market-weight swine (population)?" For an intervention question, the key elements are the population, intervention, comparator, and outcome (PICO). An example would be "What is the effect of vaccination (intervention) on fecal shedding of *Salmonella* (outcome) compared to no vaccination (comparator) in market-weight swine (population)?" The key elements of a diagnostic accuracy question are the population, index test, and target organism. An example would be "What is the sensitivity and specificity of PCR-based methods (index test) to identify *Salmonella* spp. (target organism) from the feces of market-weight swine (population)?" The key elements of an exposure question are the population, exposure, comparator, and outcome (PECO). Exposure questions are common with toxin exposures. An example would be "Is there a dose-response relationship between high dose exposure to toxin X (exposure) and liver toxicity (outcome) in humans (population) compared to low dose exposure to toxin X (comparator)?"

The key elements of the question can then be used to create eligibility criteria for inclusion of original research studies in the review. The type of review question also will determine which study designs are appropriate to address that question, and study design may be used as an eligibility criterion. For instance, a systematic review on the efficacy of a therapeutic intervention may restrict eligibility of original research studies to only randomized controlled trials (RCTs), because this design has the

highest evidentiary value for questions about intervention efficacy (21).

In preharvest food safety, it is possible to conduct clinical trials in the species of interest (animal or plant) with an induced disease outcome ("challenge trials"). For instance, if researchers wished to evaluate a vaccine for *E. coli* O157 in cattle, they could randomly allocate cattle to receive or not receive the vaccine and then wait for the animals to be naturally exposed to *E. coli* O157 (natural exposure trial), or they could deliberately expose all of the cattle to *E. coli* O157, perhaps by inoculating the bacteria into the water or food source (challenge trial). The advantage of a deliberate disease exposure is that the researchers can be sure that exposure will take place (in the RCT example, above, it is possible that the experimental animals would not end up being exposed to *E. coli* O157 during the course of the trial). In a challenge trial, all of the animals in the trial are exposed to the pathogen of interest, meaning that a smaller sample size will be needed in a challenge trial compared to an RCT. However, while challenge trials often are an efficient way to evaluate proof-of-concept for an intervention in the species of interest, they are of lower evidentiary value compared to natural disease exposure trials (21). The disease challenge may not represent natural exposure to disease; the challenge may involve a higher dose of the infectious disease agent or exposure by a different route than is typical in a natural exposure. Additionally, with many foodborne agents, some degree of biocontainment may be needed for challenge trials, meaning that, for logistical reasons, the experimental animal population may differ in a meaningful way (i.e., smaller or younger) from the animals for which the intervention would be used in the field.

There is some empirical evidence that the results of challenge trials differ from the results of natural exposure trials for preharvest food safety interventions (24). However, for some interventions, data from challenge trials may be the only form of evidence available; as an example, 63 of 66 studies included in a meta-analysis of competitive exclusion products and *Salmonella* in broiler chickens were challenge trials (11). Thus, researchers conducting a meta-analysis will need to consider how, or whether, to include challenge trials. If there is a sufficient body of literature using natural exposure trials, then it might be appropriate to exclude challenge trials, However, if this is not the case, then the researcher may decide to include both types of trials. If challenge trials are included in a review, the researcher should evaluate whether the results differ between natural exposure trials and challenge trials as a component of the meta-analysis.

The next step of a systematic review is to conduct a comprehensive and documented search of the peer-reviewed and "gray" (non-peer-reviewed) literature to identify all possible publications providing information to address the review question. The validity of the findings of a systematic review is directly related to the comprehensiveness of the search and to the reproducibility of the search protocol (4, 19). Search terms are created based on the key elements of the question and the eligibility criteria. These search terms are entered into electronic databases or used to search other sources. Although no guidelines have been published on methods of searching the literature specifically for preharvest food safety, guidelines are available for searching the medical literature (25–27) and the veterinary literature (28, 29).

The intention is to create a highly sensitive search. Thus, the specificity of the search may be low, and it is necessary to screen the titles and abstracts identified by the search to select the original research studies that are relevant to the review question (30). This is done using a small number of questions designed to rapidly identify nonrelevant articles. For instance, if the purpose of the review is to compare the prevalence of *Campylobacter* spp. between organic and traditional swine farms, the review questions might include the following: (i) Does the title/abstract describe a primary research study? (ii) Is *Campylobacter* in swine a measured outcome? and (iii) Does the study include both traditional and organic farms? If the answer to any of these questions is "no," then the publication would be excluded from further stages of the review. If the answer to any of the questions is "yes" or "unclear," then the full publication would be acquired and the publication would be included in subsequent steps of the review. The titles and abstracts of all citations identified by the search are independently assessed by at least two reviewers using the screening questions, with any disagreements between reviewers resolved by consensus. Abstracts that are not relevant are excluded from the review at this stage.

Once relevant studies have been identified, full articles are obtained for the relevant studies and data are extracted. These data include information on study characteristics such as the population (including animal/plant characteristics and sample sizes in each intervention group), the study setting and details on the intervention, and information on the outcome and results (23). The data are extracted using structured forms, with at least two reviewers independently extracting data from each article. Any disagreements are resolved by discussion or reference to an additional reviewer.

The risk of bias also is assessed for each of the relevant studies included in the review. The important sources of bias, and the methods used to reduce them, vary between study designs and review question type. For intervention questions using RCTs, a common instrument used to assess bias is the Cochrane risk of bias tool (31), which focuses on five domains of bias: selection bias, performance bias, detection bias, attrition bias, and reporting bias. To determine the potential for bias in an original research study, the reviewer considers sources of bias, including random sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting, and any other potential sources of bias specific to the context of the review question. A risk of bias graph can then be generated to demonstrate the proportion of studies with each judgment or the individual assessments for each study (31). Examples using hypothetical data are shown in Fig. 1 and Fig. 2.

Risk of bias tools developed for human intervention studies may not address all of the important issues in preharvest food safety studies (23). However, the Cochrane risk of bias tool could be modified for use in evaluating preharvest food safety trials. For instance, reviewers may want to include a consideration of whether the exposure to a pathogen of interest was due to natural exposure or whether there was a deliberate pathogen challenge. If the unit of intervention allocation was at the group level (for instance, a pen of animals), the reviewer may want to include a description of whether or not the potential for clustering of data was included in the analysis.

For questions that relate to diagnostic test evaluation, a commonly used risk of bias tool that can be modified for food safety is the QUADAS-2 tool (32) and the Cochrane website devoted to systematic reviews of diagnostic tests (http://methods.cochrane.org/sdt/). Authors of reviews about prevalence will likely need to design a risk of bias tool specific to the review topic. This tool should consider the representativeness of the sample (potential for sampling bias), the reliability of the test used to assess the outcome and the potential for information bias, and if applicable, the loss to follow-up (attrition bias) (33, 34). Risk of bias tools for nonrandomized studies used for questions about etiology are being developed. None are specific to preharvest food safety.

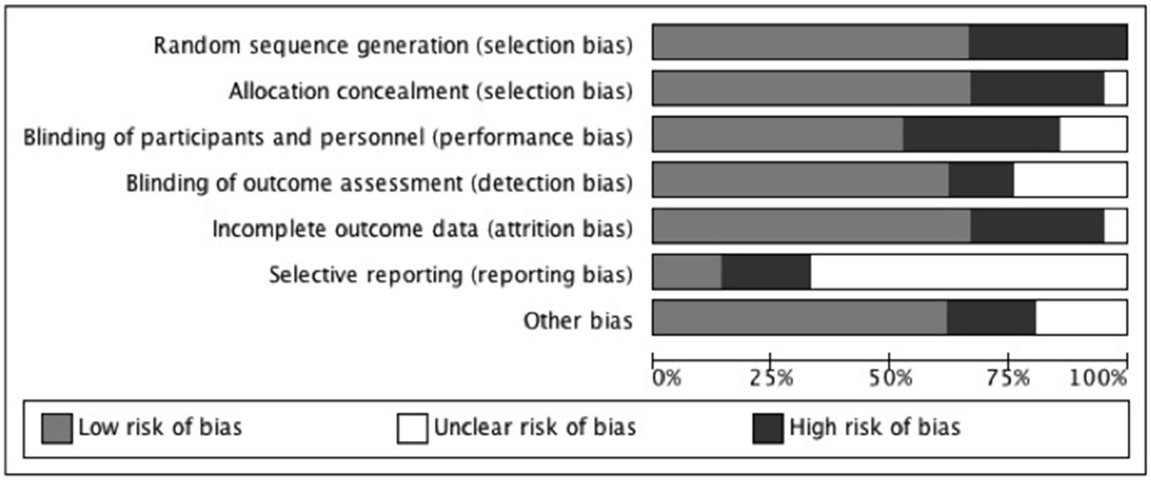
The risk of bias assessment should be conducted by at least two reviewers working independently, with any disagreements resolved by consensus. Information on the risk of bias in individual studies can be used in meta-analysis as a potential explanatory variable in meta-regression or as a factor in subgroup analysis (see sections below). Once the data on study characteristics, study outcomes, and risk of bias are collected, a meta-analysis can be undertaken.

THE PROCESS OF META-ANALYSIS

Visualizing the Results from Individual Studies

Meta-analysis involves the calculation of a summary effect size as a weighted average of the results from individual studies. Describing the terminology surround-

FIGURE 1 Example of a risk of bias graph using hypothetical data (created in Revman version 5.2). Each study included in the review has been evaluated for the risk of bias based on the domains shown in this figure. Each row of the figure summarizes the proportion of studies classified as low risk of bias, high risk of bias, or unclear risk of bias for that domain.



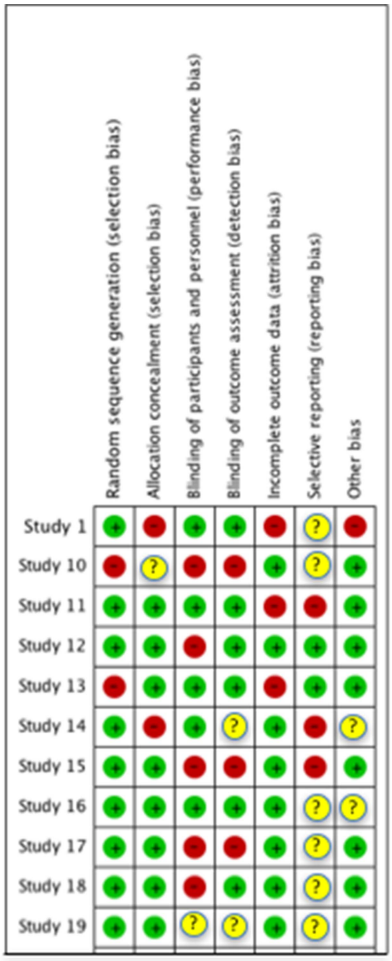


FIGURE 2 Example of a risk of bias summary using hypothetical data (created in Revman version 5.2). The results of the risk of bias assessment for each study for each risk of bias domain are shown, where "+" (green circles) corresponds to a low risk of bias in a specific study for that domain, "-" (red circles) corresponds to a high risk of bias, and "?" (yellow circles) corresponds to an unclear risk of bias.

ing study results can be confusing because studies can have several layers of outcome. A study subject may have an outcome (e.g., the organism was detected: yes or no), those individual results can be compiled to create an outcome for within an intervention group (e.g., the proportion positive), and the comparison of the intervention among groups could correspond to the outcome (e.g., the ratio of the group proportions). For meta-analysis, the outcome of interest is the comparison level outcome, i.e., the one number that describes the result of the comparison.

The actual outcome that is used will vary depending on the question being addressed by the meta-analysis. For instance, if the intent is to estimate the prevalence of a foodborne pathogen, then the outcome will be the proportion positive (number positive / number at risk). If the intent of the meta-analysis is to estimate the efficacy of an intervention or to evaluate the impact of an exposure, the outcome will be a comparison of two groups, such as a ratio for categorical outcomes (risk ratio, odds ratio) or the difference in a continuous outcome (mean difference or standardized mean difference). For questions about diagnostic tests, the outcome will be sensitivity and specificity. The discussion in this section will focus on intervention questions, although the principles apply to any question type.

Data on the outcome are extracted from each of the individual studies. This may consist of data for each intervention group (e.g., proportion positive in each group) or intervention effects at the comparison level (e.g., relative risk of an event in the treated group versus the control group). A measure of variability also is extracted or calculated. For preharvest studies where there is the potential for nonindependence between intervention groups (e.g., clustering of observations due to animal grouping), the intervention effect and measure of variability corrected for nonindependence should be used.

Intervention effects from multiple studies often are displayed using a forest plot (35) (Fig. 3). Each row in the forest plot represents an intervention comparison (e.g., intervention group versus control group). The x axis represents the summary outcome of interest. In Fig. 3, the outcome of interest is a relative risk (risk ratio). However, the x axis could be another relative measure such as odds ratio or a parameter such as prevalence, depending on the specific question being addressed by the meta-analysis. For each intervention comparison, the center of the solid square box corresponds to the estimated risk ratio obtained for that comparison, and the size of the box corresponds to the weight given to that comparison in the meta-analysis. The horizontal line on either side of the box corresponds to the precision of the estimate (generally the 95% confidence interval).

When using risk ratios or odds ratios, the effects are usually plotted on a log-scale to produce symmetrical confidence intervals. A vertical line drawn at a risk ratio or odds ratio value of 1.0 represents no difference in the effect between the intervention groups. For desirable outcomes, a risk ratio or odds ratio value >1 indicates that the intervention was effective in improving the outcome compared to the control group. For example, if the outcome is being "pathogen negative," and the intervention group is 80% pathogen negative and the control group is 20% pathogen negative, the risk ratio is 4 (0.8/0.2), and the desirable outcome occurs at a 4-fold

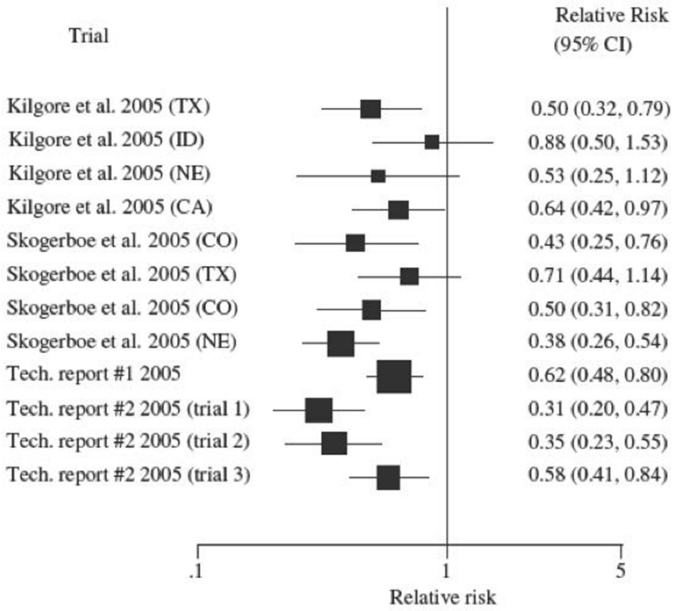


FIGURE 3 Forest plot illustrating relative risk of retreatment for bovine respiratory disease following treatment with tulathromycin compared to other available antibiotic treatments (61). Each row corresponds to treatment comparison, with the box representing the relative risk estimate for the comparison and the line corresponding to the 95% confidence interval around that estimate. The size of the box is representative of the relative amount of information contributed for that comparison (study weighting). The vertical line represents the null effect (relative risk of 1).

higher level in the intervention group compared to the control group. However, in many food safety studies, the outcome is undesirable (e.g., shedding of a foodborne pathogen in feces). If that is the case, then relative risk or odds ratio values <1 indicate a beneficial intervention (i.e., a reduction in shedding). For example, if the outcome is being "pathogen positive," and the intervention group is 20% pathogen positive and the control group is 80% pathogen positive, the risk ratio is 0.25 (0.2/0.8), and the undesirable outcome occurs in the intervention group at one-quarter of the rate it does in the control group. The confidence interval describes our certainty about the magnitude of the effect size. A confidence interval overlapping the vertical line of "no effect" represents the lack of a statistically significant association at the 5% alpha level for that comparison.

If the outcome is continuous, such as plant yield, average daily gain, or log colony forming units of a pathogen, then the summary effect may be the difference in means between intervention groups. The vertical line in a forest plot will then present a 0 difference in means, the null value. The interpretation of the negative and positive values is based on how the mean difference is calculated. For example, if the mean difference is calculated as A – B, then a negative value implies that group B has a higher mean and a positive value means that group A has a higher mean.

While a graphic display of the results from individual studies provides a useful visual representation of intervention effectiveness, it is inappropriate to generate inferences based on a graphical display. It is necessary to conduct a formal meta-analysis to statistically combine the results from the individual studies to make inferences regarding the statistical significance of the summary effect size.

Calculating a Summary Effect

Meta-analysis involves the calculation of a weighted mean of the results from the individual studies; thus, each observation in the meta-analysis is the result from an original research study. A simple arithmetic mean of the results across all trials would give misleading results, because small studies are more subject to random error and therefore should be given less weight (36). The larger the weight given to a specific study, the more the results of that study will contribute to the weighted average summary effect measure. The weights are therefore chosen to reflect the amount of information that each study contains.

Meta-analysis is based on one of two basic statistical models, either a fixed effect or a random effects model (1, 31). The underlying assumption of a fixed effect model is that there exists a "true effect size" (intervention efficacy), and therefore any difference between studies is the result of sampling error (chance). In contrast, the random effects approach assumes that there may be different effect sizes underlying different studies (for example, the effect size may vary based on characteristics of the study populations) (1). Random effects models generally produce a point estimate of the summary effect size that is similar to that obtained from a fixed-effects model, but with wider confidence intervals. Random effects models are generally more appropriate and provide a more conservative estimate. In the absence of heterogeneity (differences among studies), a fixed effect model will produce the same results as a random effects model.

In a fixed effect meta-analysis, it is assumed that the observed intervention effect varies between studies only because of the random error inherent in each study. Therefore, the weight assigned to each study is commonly based on the inverse of the study's variance (1). With this approach, larger studies (which have smaller

standard errors/variance) are given more weight than smaller studies (which have larger standard errors). There are three common approaches to estimating the weights: the inverse variance method, the Mantel-Haenszel method, and the Peto's odds ratio method (31, 37).

In the inverse variance method, the weight assigned to each study is the reciprocal of the squared standard error. The inverse variance method has wide applicability, because this approach can be used for either dichotomous or continuous data and requires that each study provide only an intervention effect estimate and a standard error. In some instances, an intervention effect such as the odds ratio (for categorical data) can be calculated from the data provided in a publication. Again, for categorical data, when zero cells (no events in one arm of a study) are present in a study, it is necessary to add a small quantity (generally 0.5) to that cell to allow calculation of the odds ratio (37, 38). When data are sparse, either because of low event rates or small trials, estimates of the standard error are poor. In these instances, the Mantel-Haenszel or Peto method may be preferred.

In a random effects meta-analysis, it is necessary to compute both the within study variance and the between study variance (the variance of the intervention effect size across the population of studies). A common method is the DerSimonian and Laird model (39), in which the study effects are assumed to follow a normal distribution, with the variance of that distribution estimated from the data.

The random effects method tends to be more conservative (wider confidence intervals on the summary effect size) compared to the fixed effect method and gives relatively more weight to smaller studies. The DerSimonian and Laird random effects method has the same wide applicability as fixed effect models, in that it can be used for studies with any type of outcome data as long as an intervention effect and standard error are provided in the individual studies or can be calculated from the data presented (37).

Evaluating Heterogeneity

While meta-analysis produces an estimate of the summary effect size, it is equally important to understand the consistency of that effect size across studies. For instance, the applicability of an intervention that consistently reduced fecal shedding of a foodborne pathogen by 50% across all studies would differ from an intervention that resulted in an average reduction in fecal shedding of 50% but ranged from 10% to 90% in the individual studies that contributed to that average.

Heterogeneity refers to the differences in effect sizes among studies (31). Variations in effect size between studies may be due to random error (chance) or "true" heterogeneity. True heterogeneity may occur due to differences between studies in characteristics of the populations, due to interventions and outcomes ("clinical heterogeneity"), and/or due to differences between studies in study design or risk of bias ("methodological heterogeneity") (40). For instance, if an intervention had a different effect in preweaned animals compared to adult animals, and studies with different ages of animals were included in the meta-analysis, then "age" would be a source of clinical heterogeneity in the results. Alternatively, if an antibiotic had a different efficacy when administered intravenously compared to intramuscularly, and if studies evaluating both methods of administration were included in the meta-analysis, then "route of administration" would be a source of clinical heterogeneity. If studies that employed blinding of the outcome assessor tended to have a different effect size compared to studies that did not use blinding, and if both types of studies were included in the meta-analysis, then "blinding of outcome assessor" would be a source of methodological heterogeneity. If substantive heterogeneity exists, it may not be meaningful or appropriate to calculate a summary effect size.

Heterogeneity should be evaluated as a component of all meta-analyses; the evaluation may be used to provide context on the degree of difference among studies in the review, to determine whether pooling of results should be undertaken, or as a first step for evaluating possible sources of heterogeneity. A variety of formal statistical tests are available to evaluate heterogeneity among a group of studies (41). One commonly used test to determine whether or not heterogeneity is present (at a predetermined P value cut-point) is Cochran's chi-square test of homogeneity (sometimes known as Cochran's Q or simply Q). This test measures deviation of observed effect sizes from an underlying overall effect size. The null hypothesis for the Q test is that there is homogeneity of effect sizes among the studies included in the analysis; therefore, rejection of the null hypothesis implies that heterogeneity is present. However, the Q test may be a poor indicator of true heterogeneity among studies if a small number of studies are included in the calculation, resulting in low power to detect heterogeneity. It is common that a P value cut-off of 0.10 or larger be used as a significance level, although this approach carries the risk that when many large studies are available, a clinically unimportant difference could be significantly heterogeneous.

In addition to testing for the presence or absence of heterogeneity, it is possible to quantify the amount of heterogeneity. The I^2 test provides an estimate of the proportion of total variability that can be attributed to heterogeneity beyond chance. I^2 lies between 0% and 100%, with a value of 0% indicating no observed heterogeneity and larger values corresponding to increasing heterogeneity beyond that expected by chance. The I^2 is less affected by the number of studies in the analysis compared to the Q test (41), although this has been disputed (42). Given that the Q test and I^2 statistic both contribute to understanding heterogeneity in a meta-analysis, both should be calculated and reported.

One of the most difficult questions in meta-analysis is How much heterogeneity is too much for calculating a summary effect size? Heterogeneity will always exist, and the question is actually What is meaningful heterogeneity (43)? As a general guideline, I^2 values of 0 to 40% are likely unimportant, 30 to 60% represents moderate heterogeneity, 50 to 90% represents substantial heterogeneity, and 75 to 100% represents considerable heterogeneity (31). As an example of the range of heterogeneity estimates that have been reported in the food safety literature, Totton et al. (13) conducted 21 random-effects meta-analyses on different feed and water additives to evaluate their effect on *Salmonella* concentration or prevalence in the ceca of broiler chickens. The I^2 estimates for the meta-analyses ranged from 0 to over 95%; the authors did not report summary estimates for meta-analyses where heterogeneity was statistically significant.

Exploring Causes of Heterogeneity: Meta-Regression and Subgroup Analysis

Testing and reporting the magnitude of heterogeneity does not help to explain the factors associated with heterogeneity. Subgroup analysis and meta-regression are tools in meta-analysis that are used to understand factors associated with heterogeneity.

Subgroup analysis involves splitting the original studies into subgroups to make comparisons between them (for example, based on geographic location or other population characteristics or based on indicators of study quality such as randomized versus nonrandomized trials). Subgroup analysis may be conducted to explore heterogeneity or to address specific questions about populations or types of studies (31). For instance, researchers might be interested in evaluating the efficacy of vaccination for a foodborne pathogen separately for dairy cattle and for beef cattle, or in calves versus adult cattle. Snedeker et al. (9) used subgroup analysis to evaluate two vaccine types for reducing fecal shedding of *E. coli* O157 and found that reduction in fecal shedding was similar between the two vaccine types. Subgroup analysis is limited to comparisons of a single variable that is categorical. However, numerous subgroup analyses can be conducted. Subgroup analysis can be a useful tool for exploring data but should be used with caution, particularly if the subgroups were not identified *a priori*, due to the increased probability of false-positive significance tests as multiple tests are performed.

If it is of interest to explore multiple potential explanatory variables or a continuous covariate as a source of heterogeneity, then meta-regression can be used. Meta-regression is a weighted regression where the unit of concern is the study, and the outcome of interest is the effect of the intervention at the study level. The weights in meta-regression are frequently the inverse variance of each study's result. The regression coefficient describes how the outcome variable changes with a unit increase in the explanatory variable. If the summary effect is a ratio measure, such as a risk ratio or odds ratio, the log-transformed value of the intervention effect is used, and the exponent of the regression coefficient gives an estimate of the relative change in intervention effect size with a unit increase in the explanatory variable. Meta-regression allows the exploration of potential sources of heterogeneity or potential biases (such as indicators of study quality). Examples of meta-regression in the food safety literature include an evaluation of study-level predictors on the prevalence of *Salmonella* on swine farms (16) and an investigation of both study-level variables and methodological criteria as explanatory variables in a meta-regression on the effect of competitive exclusion products on *Salmonella* in broiler chickens (11). Wisener et al. (24) used meta-regression and subgroup analyses to evaluate whether intervention effect sizes differed between challenge trials and natural exposure trials for three food safety pathogen–commodity group combinations.

An excellent review of the advantages and disadvantages of meta-regression is available (41). If the number of studies is limited, factors might be investigated one at a time in univariate regressions, or if there are sufficient data, a multivariable regression model could be built. However, investigation of multiple factors is often not possible due to inadequate numbers of studies. Meta-regression should generally not be considered when there are few studies; a rough guideline is that more than 10 studies should be available before considering meta-regression as an approach to exploring heterogeneity (31).

Limitations of Meta-Analysis

A meta-analysis can only assess the literature that is available to the investigators. A large amount of the actual research conducted for many interventions may not be available in the public domain, and this may serve as a source of bias in the estimation of the summary effect measure if availability is associated with the effect size. This type of bias is called reporting bias and has several subcategories: publication bias, time lag bias, duplicate publication bias, location bias, citation bias, language bias, and outcome reporting bias (31, 39, 44).

Publication bias refers to the tendency for studies that do not show a significant effect not to be published, particularly if the sample size was small (either because they are not submitted for publication or because journals are more reluctant to publish them) (45).

Time lag bias is related to the tendency for trials showing a positive intervention effect to be published faster than trials that report nonstatistically significant results. Thus, recent nonsignificant trials may not be available in the public domain as quickly as trials showing a significant difference. Trials showing a significant difference may be more likely to be available in multiple forums (e.g., conference proceeding, report, and peer-reviewed publication), thereby resulting in duplicate publication bias. Trials showing a significant intervention effect also tend to be published in higher-profile journals, which are more likely to be indexed in electronic databases and are more likely to be cited. This results in easier identification for inclusion in a meta-analysis using standard literature search methods, resulting in location bias and citation bias. Literature published in the English language also is easier to identify in searches, and if non-English articles are identified, there may not be the resources to translate the articles for inclusion in the meta-analysis. There is some evidence in the human health care literature that published non-English trials tend to be smaller, of lower quality, and more likely to report a statistically significant difference, although the effect on the overall intervention estimate may not be large (46). Finally, there is evidence that, even within published trials, statistically significant outcomes are more likely to be reported than nonsignificant outcomes, leading to outcome reporting bias (47). A study of conference proceeding abstracts reporting on food safety intervention studies at the preharvest and abattoir level reported that less than half of the research reported in conference abstracts was published in the peer-reviewed literature within 4 years (48). The same study found that conference abstracts reporting at least one positive outcome were more likely to be subsequently published and were published faster than those reporting nonsignificant findings.

One approach to evaluating publication bias in a meta-analysis is to create a funnel plot, which can be visually examined for signs of asymmetry (47, 49). Funnel plots are simple scatter plots of the intervention effect sizes estimated from individual studies (on the x axis) against some measure of each study's precision (y axis), usually the standard error. The name "funnel" is used because precision in the estimation of the true intervention effect size should increase as the sample size of the studies included in the meta-analysis increases. Therefore, the expectation would be that effect sizes from small studies would scatter more widely at the bottom of the graph, with smaller differences in the estimated effect size among larger studies. Theoretically, in the absence of bias, the plot should resemble a symmetrical inverted funnel. Therefore an assessment of the symmetry of a funnel plot is actually an assessment of small study effects. Although funnel plots are often said to assess publication bias, they actually assess whether small studies have different effect sizes compared to larger studies. Publication bias may be one cause of small study effects. Other possible sources of asymmetry in funnel plots can include (i) selection bias, location bias, language bias, citation bias, or multiple publication bias; (ii) poor methodological quality of smaller studies; (iii) true heterogeneity where the magnitude of the intervention effect size differs according to study size (for example, due to differences in the intensity of interventions); (iv) artifactual (sampling variation can lead to an association between the intervention and its standard error); or (v) chance. There are a number of statistical tests to formally evaluate asymmetry in a funnel plot rather than relying on a visual assessment for details (see 45, 50, 51).

Reporting the Results of a Meta-Analysis

As with any type of study, it is important that meta-analyses (and systematic reviews) are reported in sufficient detail to enable the reader to determine how the meta-analysis was conducted and to evaluate the potential for biased results. The PRISMA statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; www.prisma-statement.org) provides guidelines for reporting the results of systematic reviews, including the meta-analysis component (52). An accompanying elaboration document provides explanations for each of the recommended items and examples of good reporting (53).

CHALLENGES

The main challenges associated with meta-analysis are publication bias (as discussed in the previous section), insufficient quantity of research available to address the question of interest, lack of comprehensive reporting of the effect sizes, and high risk of bias in the primary studies. Theoretically, a meta-analysis can be done if two relevant studies are available. However, meta-analysis becomes more informative when there are more studies for inclusion. When the number of studies does not warrant a meta-analysis, a narrative summary may still be useful and can provide insights and directions for further research.

The challenge with poor-quality studies in meta-analysis is the "garbage in, garbage out" paradigm. There is evidence in the human health care literature that trials that do not report important methodological features aimed at reducing bias have exaggerated intervention effects (54–57), and there is some evidence that the same is true in preharvest food safety trials (58). Guidelines are available to describe the items that should be reported in preharvest food safety clinical trials (59, 60; www.reflect-statement.org). While it is possible that a well-reported trial was conducted poorly, it is essential that the reader of a trial know what was done in the trial in order to make an informed decision about the likelihood of bias. A major advantage of conducting meta-analysis as a component of a systematic review is that risk of bias assessment will have been conducted on each of the studies. This information can be included in a meta-analysis and in the interpretation of the results of the meta-analysis.

SUMMARY

Meta-analysis is the statistical pooling of data from multiple studies. Including meta-analysis as the analytical component of a systematic review increases the transparency and rigor of the analysis. The steps to a meta-analysis include extracting the data from the original research studies and displaying the data using a forest plot, combining the data using either a fixed effect or a random effects models, and evaluating heterogeneity. If the amount of heterogeneity suggests that a single summary effect size does not well represent the data, it may not be appropriate to report a summary effect size. Reasons for heterogeneity can be explored using meta-regression or subgroup meta-analysis. Meta-analysis may be a useful methodology for preharvest food safety research to aid in policy or clinical decision-making or to provide inputs to quantitative risk assessment or other models.

REFERENCES

1. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. 2009. *Introduction to Meta-Analysis*. John Wiley and Sons, West Sussex, United Kingdom.

2. O'Connor AM, Sargeant JM, Wang C. 2014. Conducting systematic reviews of intervention questions. III. Synthesizing data from intervention studies using quantitative approaches (meta-analysis). *Zoonoses Public Health* **61**(Suppl. 1)**:**52–63.

3. den Besten HMW, Zwietering MH. 2012. Meta-analysis for quantitative microbiological risk assessments and benchmarking data. *Trends Food Sci Technol* **25:**34–39.

4. European Food Safety Authority. 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. *Eur Food Saf Auth J* **8:**1637–1727.

5. Sargeant JM, Rajic A, Read S, Ohlsson A. 2006. The process of systematic review and its application in agri-food public-health. *Prev Vet Med* **75:**141–151.

6. Gonzales-Barron U, Butler F. 2011. The use of meta-analytical tools in risk assessment for food safety. *Food Microbiol* **28:**823–827.

7. Vialette M, Pinon A, Leporq B, Dervin C, Membré JM. 2005. Meta-analysis of food safety information based on a combination of a relational database and a predictive modeling tool. *Risk Anal* **25:**75–83.

8. Wisener LV, Sargeant JM, O'Connor AM, Faires MC, Glass-Kaastra SK. 2015. The use of direct-fed microbials to reduce shedding of *Escherichia coli* O157 in beef cattle: a systematic review and meta-analysis. *Zoonoses Public Health* **62:**75–89.

9. Snedeker KG, Campbell M, Sargeant JM. 2012. A systematic review of vaccinations to reduce the shedding of *Escherichia coli* O157 in the faeces of domestic ruminants. *Zoonoses Public Health* **59:**126–138.

10. Varela NP, Dick P, Wilson J. 2013. Assessing the existing information on the efficacy of bovine vaccination against *Escherichia coli* O157:H7: a systematic review and meta-analysis. *Zoonoses Public Health* **60:**253–268.

11. Kerr AK, Farrar AM, Waddell LA, Wilkins W, Wilhelm BJ, Bucher O, Wills RW, Bailey RH, Varga C, McEwen SA, Rajić A. 2013. A systematic review-meta-analysis and meta-regression on the effect of selected competitive exclusion products on *Salmonella* spp. prevalence and concentration in broiler chickens. *Prev Vet Med* **111:**112–125.

12. Totton SC, Farrar AM, Wilkins W, Bucher O, Waddell LA, Wilhelm BJ, McEwen SA, Rajic A. 2012. A systematic review and meta-analysis of the effectiveness of biosecurity and vaccination in reducing *Salmonella* spp. in broiler chickens. *Food Res Int* **45:**617–627.

13. Totton SC, Farrar AM, Wilkins W, Bucher O, Waddell LA, Wilhelm BJ, McEwen SA, Rajić A. 2012. The effectiveness of selected feed and water additives for reducing *Salmonella* spp. of public health importance in broiler chickens: a systematic review, meta-analysis, and meta-regression approach. *Prev Vet Med* **106:**197–213.

14. Wilhelm B, Rajić A, Parker S, Waddell L, Sanchez J, Fazil A, Wilkins W, McEwen SA. 2012. Assessment of the efficacy and quality of evidence for five on-farm interventions for *Salmonella* reduction in grow-finish swine: a systematic review and meta-analysis. *Prev Vet Med* **107:**1–20.

15. Young I, Rajić A, Wilhelm BJ, Waddell L, Parker S, McEwen SA. 2009. Comparison of the prevalence of bacterial enteropathogens, potentially zoonotic bacteria and bacterial resistance to antimicrobials in organic and conventional poultry, swine and beef production: a systematic review and meta-analysis. *Epidemiol Infect* **137:**1217–1232.

16. Sanchez J, Dohoo IR, Christensen J, Rajic A. 2007. Factors influencing the prevalence of *Salmonella* spp. in swine farms: a meta-analysis approach. *Prev Vet Med* **81:**148–177.

17. Islam MZ, Musekiwa A, Islam K, Ahmed S, Chowdhury S, Ahad A, Biswas PK. 2014. Regional variation in the prevalence of *E. coli* O157 in cattle: a meta-analysis and meta-regression. *PLoS One* **9:**e93299. doi:10.1371/journal.pone.0093299.

18. Wilkins W, Rajić A, Parker S, Waddell L, Sanchez J, Sargeant J, Waldner C. 2010. Examining heterogeneity in the diagnostic accuracy of culture and PCR for *Salmonella* spp. in swine: a systematic review/meta-regression approach. *Zoonoses Public Health* **57**(Suppl 1)**:**121–134.

19. O'Connor AM, Anderson KM, Goodell CK, Sargeant JM. 2014. Conducting systematic reviews of intervention questions. I. Writing the review protocol, formulating the question and searching the literature. *Zoonoses Public Health* **61**(Suppl 1)**:**28–38.

20. Sargeant JM, O'Connor AM. 2014. Introduction to systematic reviews in animal agriculture and veterinary medicine. *Zoonoses Public Health* **61**(Suppl 1)**:**3–9.

21. Sargeant JM, Kelton DF, O'Connor AM. 2014. Study designs and systematic reviews of interventions: building evidence across study designs. *Zoonoses Public Health* **61**(Suppl 1)**:**10–17.

22. Sargeant JM, Kelton DF, O'Connor AM. 2014. Randomized controlled trials and challenge trials: design and criterion for validity. *Zoonoses Public Health* **61**(Suppl 1)**:**18–27.

23. Sargeant JM, O'Connor AM. 2014. Conducting systematic reviews of intervention questions. II. Relevance screening, data extraction, assessing risk of bias, presenting the results and interpreting the findings. *Zoonoses Public Health* **61**(Suppl 1)**:**39–51.

24. Wisener LV, Sargeant JM, O'Connor AM, Faires MC, Glass-Kaastra SK. 2014. The evidentiary value of challenge trials for three pre-harvest food safety topics: a systematic assessment. *Zoonoses Public Health* **61:**449–476.

25. Conn VS, Isaramalai SA, Rath S, Jantarakupt P, Wadhawan R, Dash Y. 2003. Beyond MEDLINE for literature searches. *J Nurs Scholarsh* **35:**177–182.

26. Crumley ET, Wiebe N, Cramer K, Klassen TP, Hartling L. 2005. Which resources should be used to identify RCT/CCTs for systematic reviews: a systematic review. *BMC Med Res Methodol* **5:**24.

27. McKibbon KA, Wilczynski NL, Haynes RB, Hedges Team. 2009. Retrieving randomized controlled trials from MEDLINE: a comparison of 38 published search filters. *Health Info Libr J* **26:**187–202.

28. Alpi KM, Stringer E, Devoe RS, Stoskopf M. 2009. Clinical and research searching on the wild side: exploring the veterinary literature. *J Med Libr Assoc* **97:**169–177.

29. Grindlay DJ, Brennan ML, Dean RS. 2012. Searching the veterinary literature: a comparison of the coverage of veterinary journals by nine bibliographic databases. *J Vet Med Educ* **39:**404–412.

30. Meade MO, Richardson WS. 1997. Selecting and appraising studies for a systematic review. *Ann Intern Med* **127:**531–537.

31. Higgins JPT, Green S (ed). 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.handbook.cochraneorg.

32. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* **155:**529–536.

33. Broen MP, Braaksma MM, Patijn J, Weber WE. 2012. Prevalence of pain in Parkinson's disease: a systematic review using the modified QUADAS tool. *Mov Disord* **27:**480–484.

34. Shamliyan TA, Kane RL, Ansari MT, Raman G, Berkman ND, Grant M, Janes G, Maglione M, Moher D, Nasser M, Robinson KA, Segal JB, Tsouros S. 2011. Development quality criteria to evaluate nontherapeutic studies of incidence, prevalence, or risk factors of chronic diseases: pilot study of new checklists. *J Clin Epidemiol* **64:**637–657.

35. Lewis S, Clarke M. 2001. Forest plots: trying to see the wood and the trees. *BMJ* **322:**1479–1480.

36. Egger M, Smith GD, Phillips AN. 1997. Meta-analysis: principles and procedures. *BMJ* **315:**1533–1537.

37. Egger M, Davey Smith G, Altman DG (ed). 2001. *Systematic Reviews in Health Care: Meta-Analysis in Context*, 2nd ed. BMJ Publishing Group, London, United Kingdom.

38. Friedrich JO, Adhikari NK, Beyene J. 2007. Inclusion of zero total event trials in meta-analyses maintains analytic consistency and incorporates all available data. *BMC Med Res Methodol* **7:**5.

39. DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. *Control Clin Trials* **7:**177–188.

40. Khan KS, Kunz R, Kleijnen J, Antes G. 2005. *Systematic Reviews to Support Evidence-Based Medicine: How to Review and Apply Findings of Healthcare Research*. Royal Society of Medicine Press, London, United Kingdom.

41. Higgins JPT, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21:**1539–1558.

42. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. 2008. Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Med Res Methodol* **8:**79.

43. Higgins JP. 2008. Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol* **37:**1158–1160.

44. Egger M, Smith GD. 1998. Bias in location and selection of studies. *BMJ* **316:**61–66.

45. Thornton A, Lee P. 2000. Publication bias in meta-analysis: its causes and consequences. *J Clin Epidemiol* **53:**207–216.

46. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. 2002. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* **31:**115–123.

47. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JP, Simes J, Williamson PR. 2008. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* **3:**e3081. doi:10.1371/journal.pone.0003081.

48. Snedeker KG, Totton SC, Sargeant JM. 2010. Analysis of trends in the full publication of papers from conference abstracts involving pre-harvest or abattoir-level interventions against foodborne pathogens. *Prev Vet Med* **95:**1–9.

49. Sterne JAC, Egger M, Smith GD. 2001. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* **323:**101–105.

50. Begg CB, Mazumdar M. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50:**1088–1101.

51. Egger M, Davey Smith G, Schneider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315:**629–634.

52. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P, PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* **62:**1006–1012.

53. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* **6:**e1000100. doi:10.1371/journal.pmed.1000100.

54. Jüni P, Altman DG, Egger M. 2001. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* **323:**42–46.

55. Kjaergard LL, Villumsen J, Gluud C. 2001. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* **135:**982–989.

56. Kunz R, Oxman AD. 1998. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* **317:**1185–1190.

57. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. 1998. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* **352:**609–613.

58. Sargeant JM, Saint-Onge J, Valcour J, Thompson A, Elgie R, Snedeker K, Marcynuk P. 2009. Quality of reporting in clinical trials of preharvest food safety interventions and associations with treatment effect. *Foodborne Pathog Dis* **6:**989–999.

59. O'Connor AM, Sargeant JM, Gardner IA, Dickson JS, Torrence ME, Dewey CE, Dohoo IR, Evans RB, Gray JT, Greiner M, Keefe G, Lefebvre SL, Morley PS, Ramirez A, Sischo W, Smith DR, Snedeker K, Sofos J, Ward MP, Wills R, Consensus Meeting Participants. 2010. The REFLECT statement: methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety by modifying the CONSORT statement. *Zoonoses Public Health* **57:**95–104.

60. Sargeant JM, O'Connor AM, Gardner IA, Dickson JS, Torrence ME, Dohoo IR, Lefebvre SL, Morley PS, Ramirez A, Snedeker K, Consensus Meeting Participants. 2010. The REFLECT statement: reporting guidelines for randomized controlled trials in livestock and food safety: explanation and elaboration. *Zoonoses Public Health* **57:**105–136.

61. Wellman NG, O'Connor AM. 2007. Meta-analysis of treatment of cattle with bovine respiratory disease with tulathromycin. *J Vet Pharmacol Ther* **30:**234–241.