Quality assessment of systematic reviews and meta-analyses that examine preventive antibiotic uses and management practices designed to prevent disease in livestock

Rachael Vriezen[1], Jan M. Sargeant[1,2] (iD), Ellen Vriezen[1], Charlotte B. Winder[1] (iD) and Annette M. O'Connor[3] (iD)

[1]Department of Population Medicine, Ontario Veterinary College, Guelph, ON, Canada; [2]Centre for Public Health and Zoonoses, University of Guelph, Guelph, ON, Canada and [3]Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University, Ames, IA, USA

Abstract

To implement effective stewardship in food animal production, it is essential that producers and veterinarians are aware of preventive interventions to reduce illness in livestock. Systematic reviews and meta-analyses (SR/MA) provide transparent, replicable, and quality-assessed overviews. At present, it is unknown how many SR/MA evaluate preventive antibiotic use or management practices aimed at reducing disease risk in animal agriculture. Further, the quality of existing reviews is unknown. Our aim was to identify reviews investigating these topics and to provide an assessment of their quality. Thirty-eight relevant reviews were identified. Quality assessment was based on the AMSTAR 2 framework for the critical appraisal of systematic reviews. The quality of most of the reviews captured was classified as critically low (84.2%, $n = 32/38$), and only a small percentage of the evaluated reviews did not contain critical weaknesses (7.9%, $n = 3/38$). Particularly, a small number of reviews reported the development of an *a priori* protocol (15.8%, $n = 6/38$), and few reviews stated that key review steps were conducted in duplicate (study selection/screening: 26.3%, $n = 10/38$; data extraction: 15.8%, $n = 6/38$). The development of high-quality reviews summarizing evidence on approaches to antibiotic reduction is essential, and thus greater adherence to quality conduct guidelines for synthesis research is crucial.

Background

Presently, antibiotics are a crucial tool for both the prevention and treatment of diseases in humans (WHO, 2015) and in animals (FAO, 2016; OIE, 2016). The application of antimicrobial drugs, particularly antibiotics, has prompted substantive improvements in health outcomes since such medicines were introduced 75 years ago (Laxminarayan *et al.*, 2013; FAO, 2016; OIE, 2016). However, there are growing global concerns about antimicrobial resistance (AMR) and the threat that AMR poses for the continuing efficacy of treatments for many important infectious diseases (FAO, 2016). In response to this threat, plans to combat AMR have been developed by major international organizations including the World Health Organization (WHO) (WHO, 2012, 2015), the Food and Agriculture Organization of the United Nations (FAO) (FAO, 2016), and the World Organization for Animal Health (OIE) (OIE, 2016).

Antibiotics are used in livestock production around the world, contributing to the emergence of AMR (WHO, 2012). In modern food animal production systems, antibiotics are used both for disease prevention and treatment (WHO, 2012, 2015; Laxminarayan *et al.*, 2013), and may be administered at the group level (WHO, 2012). A major concern is that the widespread use of antibiotics in animal agriculture will accelerate resistance to important treatment options for infectious diseases in humans and animals. For those microorganisms that affect animals, the implications of AMR include increased morbidity and mortality, reduced animal welfare, and production losses in livestock industries (Laxminarayan *et al.*, 2013; WHO, 2015).

The WHO, the FAO, the American Veterinary Medical Association (AVMA), the Canadian Veterinary Medical Association (CVMA), and others call for the reduction of antibiotic use in food animal production, and for improved antibiotic stewardship in the sector (WHO, 2012; Laxminarayan *et al.*, 2013; FAO, 2016; CVMA, 2017; AVMA, 2019*a*). In the context of veterinary medicine, this stewardship requires the implementation of preventive management strategies to reduce both the incidence of infectious diseases as well as the need for antibiotic treatments while maintaining animal health and welfare (AVMA, 2019*a*). There are two key

approaches that can be implemented to prevent and control diseases in food animal production: prophylactic or metaphylactic applications of antibiotics among at-risk livestock populations (AVMA, 2019b); and adhering to livestock management practices to prevent diseases that would otherwise require antibiotic treatments. Preventive uses of antibiotics include applications, usually in feed or water, to control the spread of a specific pathogen in a defined group of livestock, as well as routine treatments given to a group of animals during known times of higher disease risk to prevent infection (Laxminarayan et al., 2013). Interventions such as vaccinations, biosecurity, and hygiene best practices may also reduce infections requiring antibiotic treatment in livestock (WHO, 2015).

In order to appropriately implement disease prevention strategies, it is crucial that the animal health, production, and economic consequences of each approach are rigorously and scientifically assessed. Systematic reviews, either alone or in combination with statistical meta-analyses, are powerful synthesis tools that provide researchers, industry groups, government agencies, and clinicians with concise, quality-assessed summaries of primary research for use in evidence-based decision-making (Higgins and Green, 2011; Khan et al., 2011; Sargeant and O'Connor, 2014a). Existing systematic reviews of preventive antibiotic use and non-antibiotic management practices designed to prevent disease are a potential source of high-level evidence on the efficacy of these approaches, and thus these reviews may be crucially important in livestock management decision-making aiming to optimize antibiotic use. At present, it is unknown how many synthesis studies on these topics exist, and the quality of any such reviews is similarly unknown. Both the quality of the conduct of a review and the transparency and completeness of review reporting impact the ability of end-users to judge the relevance and accuracy of the review results and therefore affect the usefulness of the results for informed decision-making. This review focuses specifically on an evaluation of the execution of systematic review and meta-analysis studies. Well-executed reviews that minimize selection bias and information error give end-users greater confidence in the review results, and so the results of reviews that incorporate strategies to reduce or eliminate sources of bias and error are considered to be at a lower risk of bias than the results of reviews that fail to address these points (Sargeant and O'Connor, 2014a).

The purpose of this study was 2-fold: first, to inventory all systematic reviews and meta-analyses that examine antibiotic and non-antibiotic approaches to disease prevention in livestock production published in the last 25 years; and second, to evaluate the quality of these systematic reviews and meta-analyses using the AMSTAR-2 tool (Shea et al., 2017), both to determine whether there are methodological aspects that warrant improvement and to provide an indication of the level of confidence that end-users should have in the rigor and accuracy of the results of those existing reviews.

Methods

Eligibility criteria

To be included in the present analysis, a review article had to include at least one systematic review and/or meta-analysis (as defined by the review authors) that examined at least one intervention constituting either preventative antibiotics or management practices designed to reduce antibiotic use in at least one livestock or poultry

species. Reviews also had to include at least one clinically relevant outcome: a measure of clinical morbidity, mortality, gross pathological lesions, and/or condemnation at slaughter. Additionally, reviews were only included if the full text was available in English.

Search and relevance screening

The search and initial relevance screening was conducted as part of a larger scoping review on synthesis research in animal health, animal performance, and on-farm food safety over the past 25 years (1995–2018) (Vriezen et al., 2019). A subset of the papers identified in that study was selected for inclusion in the present analysis. The search and original screening methods are described in detail elsewhere (Vriezen et al., 2019), and the complete search string is available in Supplementary Appendix 1. Briefly, a comprehensive search was conducted in three electronic databases (MEDLINE via PubMed, CAB Direct, and AGRICOLA via ProQuest) for systematic reviews and/or meta-analyses combined with search terms related to livestock, companion animals, and wildlife. Two reviewers independently screened the title and abstract of each citation identified in the search for publications that used the terms 'systematic review' or 'meta-analysis' and were of relevance to animal health, performance, or on-farm food safety. The titles and abstracts of all articles that passed the first screening were subject to data characterization based on the information provided in the title and abstract, again by two independent reviewers. Ultimately, 1787 synthesis studies were included in the scoping review.

For the present analysis, an additional screening process was applied to the 1787 review articles included in the scoping study. Two reviewers independently screened each abstract to determine whether the review targeted one or more livestock or poultry species, whether the review investigated an intervention or exposure, whether the review examined at least one health outcome, and whether the subject matter of the review concerned either the use of antibiotics designed to reduce disease incidence and thus future antibiotic use, or non-antibiotic management practices intending to reduce illness or future antibiotic use. Full-text articles were obtained for the review articles that passed this screening round. Two reviewers then independently screened the full texts to ensure that each article met all of the eligibility criteria outlined above. All conflicts in either screening round were resolved by consensus.

All papers that passed the full-text screening were then subject to data characterization, which was also conducted independently by two reviewers, and was based on the full texts of the included reviews. Conflicts were resolved by consensus. The framework for the analysis of the quality of existing reviews was based on the AMSTAR 2 tool, which is designed for the critical appraisal of systematic reviews of healthcare interventions (Shea et al., 2017). The data screening form was created following the AMSTAR 2 checklist, with some modifications made by the authors to adapt the tool for this study. Additional items were added to the charting form to capture the target commodity group(s) or livestock species, as well as the interventions and outcomes that were examined in the review.

Data collection

Data were collected regarding the livestock or poultry commodity group targeted in the review, as well as the types of interventions and outcomes assessed in each article. One reviewer (RV) then

grouped the interventions and outcomes into broad categories, and another reviewer (JS) verified the categorization. A list of each of the interventions and outcomes included in each category is presented in Supplementary Appendix 2.

The first item on the AMSTAR 2 tool requires the review research question and eligibility criteria for primary study inclusion to follow the PICO/PECO framework. The acronym 'PICO' represents the components of a research question that is designed to evaluate the efficacy of an intervention, namely Population or Participants (P), Intervention(s) (I)/Exposure(s) (E), Comparison group(s) (C), and Outcome(s) (O) (EFSA, 2010; Higgins and Green, 2011). Each of the included reviews was examined to determine whether the elements of the PICO/PECO acronym could be identified in the review question, objectives, or eligibility criteria. In some cases, the comparator group was not explicitly described in the methods section of the review, but comparison groups were described and discussed in the results and other sections of the article; this was judged to be sufficient consideration. We also recorded whether or not the review authors clearly identified eligibility criteria for the inclusion of studies in the review, either through the incorporation of the PICO/PECO elements or by explicitly specifying alternate criteria. The AMSTAR 2 framework further recommends that reviews identify the timeframe for follow-up, but this was not assessed in the present study because follow-up times could differ across outcomes within a single review.

The second AMSTAR 2 item requires review authors to establish an *a priori* protocol and to justify any deviations from that protocol to ensure that the review did not evolve as literature was identified, which could create selection bias (Higgins and Green, 2011). The AMSTAR 2 checklist includes a list of requirements that the protocol must meet, and also requires that the protocol be registered. Review protocols were not obtained and examined in the current study; the criterion applied in this study required only that an *a priori* protocol be specified.

The third AMSTAR 2 item requires review authors to justify the selection of eligible study designs for inclusion in the review (i.e. randomized trials, non-randomized trials, or both). As different types of study designs may be appropriate for addressing research questions related to animal health, this item was evaluated based on whether the review authors identified which study designs were eligible both during the search and data extraction phases.

The fourth AMSTAR 2 item details the requirements of an appropriate literature search strategy: at least two databases must have been searched; the search string or key words used in the search must be provided; any search restrictions (e.g. language) must be justified; grey literature must have been searched, including contacting experts in the field and searching trial registries; the reference lists of eligible studies must have been searched; and the search must have been completed within 2 years of the review. Most of these criteria were evaluated in the current study, with the exception of the requirement to search trial registries, since trial registries for studies related to animal health have only recently been established and are not commonly used. When evaluating whether or not review authors searched the bibliographies of included or eligible primary studies, the criterion was only considered fulfilled if the authors searched the reference lists of all included studies, as opposed to searching only one or several bibliographies. It is challenging to assess the requirement that the search must have been conducted within 24 months of the completion of the review because the date of

completion of any study is not typically reported. The year of publication is more commonly available, but due to the length of the submission and review process, this date does not generally correspond to study completion. Nonetheless, stating the date of the search in the methods section of the review is useful, as it informs the reader of the timespan of the studies captured in the review. Therefore, we recorded whether or not review authors specified the date of the search as a part of our evaluation of this AMSTAR 2 item. To meet this particular criterion, authors must have identified the month and year during which the literature search was conducted.

The fifth and sixth AMSTAR 2 domains require authors to conduct study selection and data extraction in duplicate; these criteria were considered fulfilled if review authors stated that these steps were conducted independently by at least two reviewers, or if they were conducted by one reviewer and verified by a second.

The seventh AMSTAR 2 item requires authors to provide a list of any studies excluded following the retrieval of full-text articles and to justify those exclusions. In the present analysis, this condition was considered partially fulfilled if the review authors listed the reasons for any exclusions following the full-text screening, as well as the numbers of studies excluded for each reason. The criterion was considered fully met if the authors provided a bibliography of those exclusions, as well as a justification for exclusion for each study in the bibliography.

Item eight on the AMSTAR 2 tool lists all of the elements that should be presented in review articles for each of the included primary studies, including the research designs, populations, interventions, outcomes, comparators, settings, and follow-up timeframes; according to the AMSTAR 2 framework, each of these elements must be described 'in adequate detail', but little additional guidance is provided for the assessment of that level of detail (Shea *et al.*, 2017). Although it is important to describe these details, we judged that the evaluation of what constitutes an adequate level of detail for each of these areas would be subjective. For this criterion, we opted to assess only whether review authors identified the languages and study designs of the primary studies included in their review.

The ninth AMSTAR 2 item details the requirements for the risk of bias assessment for each of the studies included in the review, with separate requirements for reviews incorporating randomized and non-randomized control trials. For this item, the present study evaluated whether a risk of bias assessment was conducted or if features of bias, such as randomization, were incorporated into the eligibility criteria, whether the authors identified the tool used in the risk of bias assessment, and whether the assessment was conducted in duplicate.

The tenth AMSTAR 2 item, which requires authors to describe the funding sources of all studies included in the review, was not assessed in the present study because there is an ongoing debate as to the relevance of sources of funding as a potential source of bias (Bero, 2013; Sterne, 2013). Additionally, some of the bias that might be introduced based on funding sources for individual studies might also be captured under the checklist items devoted to the risk of bias and publication bias assessments that are applied to the review as a whole.

Item 11 on the AMSTAR 2 checklist focuses on the evaluation of meta-analytic methods. The criteria used to assess this item were simplified because several studies reported multiple meta-analyses with varying levels of detail, and it was not feasible to evaluate each meta-analysis separately. Thus, the criteria

evaluated herein were: whether the study designs included in each meta-analysis were clearly identified, whether natural disease exposure trials and challenge trials (trials with deliberate induction of the outcome) were combined in the same meta-analysis, whether observational and experimental data were combined in a single meta-analysis, and whether at least one summary effect measure was reported. This item and those subsequent items that are specific to meta-analyses were only applied to those reviews with a meta-analytic component.

The twelfth AMSTAR 2 domain was not evaluated in the present analysis. This item requires either that meta-analyses include only those primary studies at low risk for bias, or that review authors undertake additional analyses to investigate and discuss the potential impact that bias might have on the summary effect measure. Framed in this way, this domain invites subjective judgments about what level of analysis of the impact of bias is sufficient. Further, this is a potentially controversial quality criterion because the decision to exclude low-quality studies could be made either during eligibility screening or after data extraction; if the latter occurs, this may introduce reviewer subjectivity and potential bias (O'Connor and Sargeant, 2014). For the thirteenth item, AMSTAR 2 requires review authors to account for the risk of bias in individual primary studies in their interpretation of the results of their review. Similar to item 12, the appropriate nature and extent of this discussion are not elaborated in the AMSTAR 2 documentation, and so the application of this domain was considered to be ambiguous and subjective. Therefore, the present analysis simply evaluated whether any studies were removed from the review on the basis of the risk of bias assessment, and reviews with no such exclusions were evaluated positively.

The fourteenth AMSTAR 2 item requires authors to explore and explain any heterogeneity in the results of the review. This item was considered fulfilled in the present analysis if the review authors evaluated heterogeneity by using an I^2 statistic for at least on meta-analysis; since many of the reviews contained multiple meta-analyses, it was not practical to evaluate discussions of heterogeneity for each individual meta-analysis.

The penultimate AMSTAR 2 domain requires testing for small studies effects, which are often used to assess publication bias, as well as a discussion of the impact of any such bias on the review results. For this criterion, we assessed whether reviews incorporated any graphical or statistical tests of small studies effects; again, the number of individual meta-analyses precluded an evaluation of any discussions of the impacts of publication biases on the results of individual meta-analyses.

Finally, according to the sixteenth AMSTAR 2 domain, review authors must identify their funding sources and/or other competing interests. This criterion was considered fulfilled herein if the authors declared their source(s) of funding or specified that no external funding was received.

Quality assessment

Network meta-analysis
In the event that a network or mixed-treatment effect meta-analysis was identified for inclusion in the quality analysis, additional alterations to the evaluation process would be required. Network meta-analysis is a relatively new technique, which allows for an assessment of the relative effects of multiple interventions by synthesizing evidence across a network of trials (Cipriani *et al.*, 2013; Dias *et al.*, 2013). Since network meta-analytic methods are different from those methods employed in traditional (pairwise) meta-analyses, the evaluation criteria that apply to traditional meta-analyses might not be appropriate for assessing the quality of a network meta-analysis. Therefore, any network meta-analyses that were captured in this study were examined using only those criteria that were not specific to meta-analyses (i.e. the meta-analysis component of any such study was not assessed).

Identification of critical domains
Of the 16 items or domains in the AMSTAR 2 framework, the creators of the tool specify seven critical domains that can substantively impact the validity of a systematic review. Briefly, these are: (i) a protocol is registered *a priori* (Domain 2); (ii) an adequate literature search is performed (Domain 4); (iii) justification is provided for the exclusion of individual primary studies (Domain 7); (iv) risk of bias is assessed for individual studies (Domain 9); (v) any meta-analytic methods used are appropriate (Domain 11); (vi) risk of bias is considered in the interpretation of the results of the review (Domain 13); and (vii) publication bias is assessed and the impact of any such bias is discussed (Domain 15). If one of these domains is unsatisfactory, then the overall level of confidence in the results of the review is considered to be 'Low' according to AMSTAR 2; if more than one of the critical domains is unsatisfactory, then the confidence level is considered to be 'Critically Low' (Shea *et al.*, 2017). Each of these seven items was assessed in turn to evaluate the quality of the reviews included in this analysis.

Assessment of the seven AMSTAR critical domains
We generally followed the recommendations from the AMSTAR 2 tool for evaluating each review, with some modifications based on expected practices in the animal health literature. Each of the seven critical domains was assessed as follows:

(1) For the first domain, an article was characterized as a 'Yes' if the authors indicated that an *a priori* protocol was developed, and as a 'No' if the authors stated that no protocol was developed or if no information about an *a priori* protocol was provided.

(2) For the second domain, articles were awarded a 'Partial Yes' if at least two academic databases were searched, and if the authors provided either key words for the search or the entire search string used for at least one database; this was upgraded to a 'Yes' if there was also an attempt to search the grey literature in some form, including examining conference proceedings, contacting experts in the field, or searching industry publications. In addition, in order to achieve a 'Yes' the review authors must have searched the reference lists of all included studies, and they must have stated the date at which the database search of took place.

(3) The third critical domain received a 'Partial Yes' if the authors stated the reasons for any article exclusions that occurred during full-text screening, as well as the number of articles that were excluded during full-text screening for each of those reasons. Reviews received a 'Yes' for the third domain if the authors provided a bibliography of all studies excluded at full-text screening and if they identified the reason for exclusion for each study listed in the bibliography.

(4) For the fourth critical domain, a 'Partial Yes' was assigned if the authors stated that a risk of bias assessment was conducted or if features of bias were incorporated into the eligibility criteria (for example, if the review only examined randomized or blinded trials), and a 'Yes' was awarded for
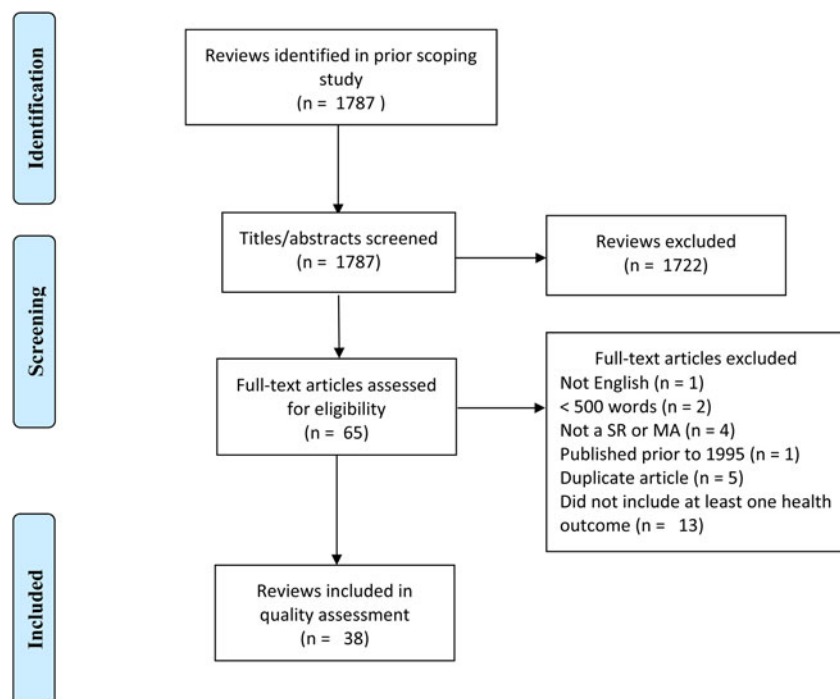
Fig. 1. Flow chart of reviews throughout the screening process. Chart illustrates the process of screening for eligibility of reviews in the quality assessment of systematic reviews (SR) and meta-analyses (MA) examining disease prevention interventions or exposures in livestock.

the fourth domain if the tool used to conduct the risk of bias assessment was also described; if neither of these conditions was met, then a 'No' was indicated.

(5) Under the fifth domain, a 'Partial Yes' was awarded if a meta-analysis was conducted, the study designs included in the meta-analysis were identified, and neither natural exposure and challenge trials nor experimental and observational studies were combined in a single meta-analysis. A 'Yes' was indicated for this domain if, in addition to the 'Partial Yes' conditions, an I^2 value and a summary effect measure were also reported as a part of at least one meta-analysis.

(6) The sixth critical domain, which requires review authors to consider the impact of the risk of bias assessment on the interpretation of their results, was not assessed in our study because the criteria associated with this domain were deemed to be subjective, ambiguous, or potentially inconsistent with best practices for conducting systematic reviews (e.g. if studies were excluded following the risk of bias assessment).

(7) Finally, articles were awarded a 'Yes' for the last critical domain if an analysis of small study effects was reported; otherwise 'No' was recorded. If no meta-analysis was conducted, then both the fifth and seventh critical domains were left blank and did not affect the final quality assessment for those reviews. If the article contained at least one meta-analysis but did not explicitly use the systematic review label, then all of the critical domains were assessed for that article.

Levels of confidence in the results of reviews

Following the evaluation of each of the seven critical domains as well as the nine non-critical domains, researchers can then combine those evaluations into an overall assessment of the confidence that a reader should have in the results of the evaluated review. Weaknesses, or unsatisfactory criteria, were divided into those related to the critical domains, described above, and non-critical weaknesses related to the other nine domains (Shea et al., 2017) (see Table 2 for a complete list of AMSTAR 2

domains assessed in this study). In this study, a critical weakness was identified if a review did not meet the criteria for one of the six critical domains; a weakness was not assigned if a review met or partially met the criteria. Each of the six domains was assessed independently for each of the reviews, and the number of critical weaknesses was evaluated for each individual review. A final confidence rating was then assigned to each review based on the number of critical weaknesses that were identified in the quality assessment process.

The creators of the AMSTAR 2 tool suggest the following rating for confidence in review results based on the evaluation of the seven critical domains: (i) High Confidence: no or one non-critical weakness; (ii) Moderate Confidence: more than one non-critical weakness; (iii) Low Confidence: one critical weakness with or without non-critical weaknesses; and (iv) Critically Low Confidence: more than one critical weakness with or without non-critical weaknesses (Shea et al., 2017). In the present analysis, only the six relevant critical domains were assessed in depth for each review. Since the distinction between 'High Confidence' and 'Moderate Confidence' involves only the number of non-critical weaknesses, these categories were combined under the heading 'High/Moderate Confidence'. For reviews that fell into this category, readers may be relatively more confident that the review results provide a complete, accurate synthesis of the existing literature compared to those reviews in the 'Low' or 'Critically Low' confidence categories.

Results and discussion

Study characteristics

Figure 1 shows the flow of review articles through the screening process. From the 1787 reviews identified in an earlier scoping review (Vriezen et al., 2019), 1722 reviews were eliminated during the abstract screening. Full texts were obtained for the remaining 65 reviews. Following the additional full-text screening, 38 reviews were retained for inclusion in the quality assessment. The citation

Table 1. Summary characteristics of 38 systematic reviews and/or meta-analyses examining preventative approaches to reducing antibiotic use

	Number
Year	
1995–1999	1
2000–2004	2
2005–2009	9
2010–2014	14
2015–2018	12
Review type	
Systematic review	4
Meta-analysis	22
Network meta-analysis[a]	1
Systematic review and meta-analysis	11
Species[b]	
Dairy cattle	15
Beef cattle	12
Cattle, unspecified	5
Swine	5
Sheep	2
Goats	0
Broiler poultry	3
Layer hens, Eggs	1
Turkeys	1
Poultry, unspecified	3
Approach to disease reduction	
Preventative antibiotic use	5
Non-antibiotic management practices	30
Both	3
Intervention types[c]	
Antibiotic applications[d]	13
Vaccines	15
Non-antibiotic feed additives[e]	9
Disinfectants/Sealants	8
Ionophores	4
Homeopathy	1
Other management practices[f]	3
Outcome types[g]	
Mortality	26
Reproductive outcomes	15
Respiratory outcomes	7
Digestive outcomes	2
Morbidity (non-specific)	10
Udder health	17
Infections (non-mastitis)	3

(Continued)

Table 1. (*Continued.*)

	Number
Peri-parturient disorders	9
Post-mortem lesions	6
Other[h]	3

[a]Network meta-analyses were assessed in the same manner as systematic reviews.
[b]May sum to more than the total number of reviews, as reviews may have examined more than one species.
[c]May sum to more than the total number of reviews, as reviews may have examined more than one intervention.
[d]Includes review articles in which antibiotics were used in control groups.
[e]Includes vitamins, probiotics, and yeast cell-wall products.
[f]Includes changes in housing practices and variations in dry period length.
[g]May sum to more than the total number of reviews, as reviews may have examined more than one outcome.
[h]Includes lameness.

information for the 38 included reviews is available in Supplementary Appendix 3.

Table 1 provides summary characteristics for the 38 reviews that were eligible for inclusion in the quality assessment. The number of systematic reviews and meta-analyses reviewing preventive approaches to improve animal health has increased over time; most (68.4%, $n = 26/38$) of the captured articles were published after 2010, and only one review (2.6%) was published between 1995 and 1999. The majority of the reviews (57.9%, $n = 22/38$) were meta-analyses that were conducted without specifying a corresponding systematic review. Dairy and beef cattle were the most common target commodity groups (39.8%, $n = 15/38$ and 31.6%, $n = 12/38$, respectively). The majority of the reviews focused on non-antibiotic management practices designed to reduce disease incidence (78.9%, $n = 30/38$); relatively few reviews examined preventive antibiotic use (13.2%, $n = 5/38$), and only three reviews (7.9%) investigated both approaches.

The reviews contained a variety of intervention approaches. Thirteen of the reviews incorporated antibiotic treatments; of these, five reviews specified that antibiotic treatments were the specific intervention under investigation, whereas the other eight reviews evaluated management practices against control groups including those treated with antibiotics. Vaccinations were the most commonly studied intervention ($n = 15$). The outcomes examined in the reviews varied, with mortality representing the most commonly studied outcome ($n = 26$), followed by udder health ($n = 17$) and reproductive outcomes including specific reproductive morbidities, infertility, and abortion ($n = 15$). Other than mastitis (captured under 'udder health'), few reviews examined interventions to reduce other infectious diseases ($n = 3$), digestive morbidity ($n = 2$), or other clinical health outcomes such as lameness ($n = 3$). One network meta-analysis was identified, which reviewed applications of antibiotics to prevent morbidity and mortality due to respiratory disease in beef cattle (Abell *et al.*, 2017).

Quality assessment

Table 2 lists each of the AMSTAR 2 items considered in the quality analysis, alongside the specific criteria used to evaluate each item. The table shows the number and percentage of review articles that met each criterion, as well as the percentage of meta-analyses that met each criterion, as appropriate.

Most of the 38 review articles (92.1%, $n = 35$) incorporated all of the PICO/PECO elements in the review question, objectives, or

Table 2. Summary of quality assessment criteria for 38 systematic reviews and/or meta-analyses examining preventative approaches to reducing antibiotic use

AMSTAR 2 domain	Criteria used in quality assessment	Number	% (Total, n = 38)	% (Meta-analyses, n = 33)
1. Did the research questions and inclusion criteria for the review include the components of PICO?				
	Research question included all of the elements of PICO[a]	35	92.1	
	Explicit eligibility criteria were provided	32	84.2	
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?				
	An *a priori* protocol was developed	6	15.8	
3. Did the review authors explain their selection of the study designs for inclusion in the review?				
	Eligible study designs were identified in the search[b]	15	39.5	
	Eligible study designs were identified during data extraction	30	78.9	
4. Did the review authors use a comprehensive literature search strategy?				
	Method for literature search described	29	76.3	
	Search string or key words were provided	23	60.5	
	At least two electronic databases were searched	28	73.7	
	Searched the reference lists of included studies	15	39.5	
	Some attempt was made to search the grey literature	14	36.8	
	Eligible languages were identified in the search[c]	23	60.5	
	Provided date of search	16	42.1	
5. Did the review authors perform study selection in duplicate?				
	Study selection was performed in duplicate	10	26.3	
6. Did the review authors perform data extraction in duplicate?				
	Data extraction was conducted in duplicate[d]	6	15.8	
7. Did the review authors provide a list of excluded studies and justify the exclusions?				
	Reasons for exclusions and numbers of studies excluded for each reason were provided[e]	15	39.5	
	A bibliography and reasons for exclusion for each study were provided[f]	4	10.5	
8. Did the review authors describe the included studies in adequate detail?				
	Languages of included studies were identified	15	39.5	
	Study designs included in the review were identified	32	84.2	

(Continued)

Table 2. (*Continued.*)

AMSTAR 2 domain	Criteria used in quality assessment	Number	% (Total, $n = 38$)	% (Meta-analyses, $n = 33$)
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?				
	A formal RoB assessment was conducted[g]	20	52.6	
	The tool used to conduct the RoB analysis was identified	8		40.0[h]
	RoB assessment was conducted in duplicate	4		20.0[i]
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?				
	A meta-analysis was performed[j]	33	86.8	100.0
	Study designs included in the meta-analysis were identified	29	76.3	87.9
	Natural disease exposure and challenge trials were not combined in the same meta-analysis	22		75.9[k]
	Observational and experimental designs were not combined in the same meta-analysis	29		100.0[l]
	At least one summary effect measure was reported	13	34.2	39.4
13. Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?				
	No exclusions were made on the basis of RoB assessment[m]	37	97.4	
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?				
	At least one I-squared value was reported	23	60.5	69.7
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?				
	Publication bias (small study effects) was assessed	13	34.2	39.4
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?				
	Funding source(s) were declared[n]	16	42.1	

[a]PICO = Population (P), Intervention (I), Comparator (C), Outcome (O).
[b]Reviews were also included if the authors specified that no restrictions were placed on study designs eligible in the search.
[c]Reviews were also included if the authors specified that no restrictions were placed on languages eligible in the search.
[d]Reviews were included if data extraction was conducted by two independent reviewers, or if data extraction was performed by one reviewer and verified by a second reviewer.
[e]Reasons were only required for exclusions made during full-text screening. Reviews were also included if they explicitly stated that no exclusions were made during full-text screening.
[f]Bibliographic references and reasons were only required for exclusions made during full-text screening. Reviews were also included if they explicitly stated that no exclusions were made during full-text screening.
[g]Reviews were included if a formal assessment was conducted or if features of bias were included in the eligibility criteria (e.g. the review was limited to randomized trials).
[h]% of risk of bias assessments ($n = 8/20$).
[i]% of risk of bias assessments ($n = 4/20$).
[j]Excludes network meta-analyses.
[k]% of meta-analyses in which the included study designs were reported ($n = 22/29$).
[l]% of meta-analyses in which the included study designs were reported ($n = 29/29$).
[m]Reviews were included if the authors specified that no exclusions were made on the basis of the risk of bias assessment, or if no information was provided about exclusions resulting from the risk of bias assessment.
[n]Reviews were included if the funding source was declared or if the authors specified that the review received no external funding.

inclusion criteria, and most (84.2%, $n = 32$) explicitly identified eligibility criteria for the primary studies included in the review. A clearly specified review question is important because it ensures that it is possible to design a search and screening process that will capture all relevant studies, and hence reduces the potential for selection bias (O'Connor *et al.*, 2014; Sargeant and O'Connor, 2014*a*). Only 15.8% ($n = 6$) of reviews indicated that an *a priori* protocol was developed. The development of a review protocol is essential to minimize bias and promote transparency in the review process. If authors are aware of the findings of potentially relevant primary studies, these results may influence the final wording of the review question, the choice of eligibility criteria, and the identification of interventions and outcomes of interest; a protocol minimizes the risk of bias towards primary studies with specific results by establishing key decisions about the focus and methods of the review in advance (Higgins and Green, 2011).

Nearly 80% ($n = 30$) of the reviews identified the study designs that were eligible for data extraction, and 39.5% ($n = 15$) stated whether or not restrictions were placed on the study designs that were eligible during the search. Different types of studies (e.g. experimental versus observational studies) may be appropriate in different situations or for different research questions. According to the creators of AMSTAR 2, it is critical that review authors justify the inclusion of different study designs in their review (Shea *et al.*, 2017). If eligible study designs are not specified in the search but are specified later (i.e. once the authors have viewed the full-text publications during data extraction), this introduces another potential source of bias; for example, authors may only select and analyze those studies and study designs that reflect their pre-conceived ideas about the results and their relevance to the review question. Specifying eligible study designs prior to full-text screening or data extraction may help to mitigate this potential bias.

Three-quarters of the reviews (76.3%, $n = 29$) included a description of the methods used to search for relevant literature, which implies that one-quarter of review authors did not describe their search methods. Without a clear description of the search procedures, the review methods are neither transparent nor reproducible, both of which are crucial foundations of the systematic review process (Higgins and Green, 2011). Approximately 40% ($n = 15$) of the included reviews specified that the reference lists of all eligible primary studies were searched in an effort to identify other potentially relevant studies. Additionally, less than half of the captured review studies (36.8%, $n = 14$) attempted to search the grey literature, which includes conference proceedings, industry reports, and contact with experts in the field. Failing to search the grey literature might mean that the review authors missed some primary studies or other information that is relevant to the review question. In particular, identifying and including completed but unpublished studies may help to reduce bias since there are often systematic differences between those studies that are published in full and those that are completed but unpublished (e.g. studies with positive results or larger sample sizes are more likely to be published) (Higgins and Green, 2011; Scherer *et al.*, 2018). Finally, 42.1% ($n = 16$) of the reviews identified the date of the literature search. For reviews that did not specify the date of the search, readers cannot be certain of the date after which published studies would not be included in the review.

Only one-quarter of reviews (26.3%, $n = 10$) explicitly stated that the selection of primary studies for inclusion in the review was conducted in duplicate, and fewer (15.8%, $n = 6$) reported that data extraction was conducted in duplicate. For those reviews that did not indicate that these steps were conducted in duplicate, it was often impossible to determine whether study selection and data extraction were truly conducted by a single individual, or whether the reviewers failed to report that these steps were conducted in duplicate. By their nature, decisions regarding which studies to include and the interpretations of some forms of primary data during the review process may involve some subjective judgments. Conducting these steps in duplicate (i.e. by more than one reviewer) helps to reduce subjectivity and ensure that these decisions are reproducible. In addition, conducting study selection and data extraction in duplicate can help to ensure that no relevant studies are inadvertently rejected, may minimize errors in the extraction process, and may also reduce the impact of any biases based on an author's pre-formed opinions or prior experiences that might impact their evaluation of the validity and relevance of particular studies or data points (Higgins and Green, 2011).

Once the review authors have viewed the data contained within a primary study, the decision on whether to include or exclude the study may be subject to bias. It may be necessary to exclude some studies after the full text has been obtained (e.g. if the answers to some screening questions were unclear based on the title and abstract alone); however, in such cases, it is recommended that authors identify the reasons for exclusions to minimize the potential for bias due to exclusion decisions (O'Connor and Sargeant, 2014; Sargeant and O'Connor, 2014*b*; Viswanathan *et al.*, 2018). Reasons for exclusions made after the full text was retrieved, as well as the numbers of studies that were excluded for each reason, were provided in 40% ($n = 15$) of the included reviews. A smaller number of reviews (10.5%, $n = 4$) provided a bibliography of those studies excluded during or after full-text screening and identified the specific justification for exclusion for each study.

Risk of bias assessments is a crucial component of systematic reviews, as these assessments provide an evaluation of the validity of the results of the primary studies included in the review (Higgins and Green, 2011). If the results of the primary studies included in the review are biased, then the summary conclusions of the review may be similarly biased. If reviews do not include an assessment of the risk of bias of the primary studies, then end-users of the review results will not be able to judge the potential for bias. Approximately half (52.6%, $n = 20$) of the reviews in this study conducted some form of risk of bias assessment. Of these, 40% ($n = 8/20$) identified the tool used to conduct the assessment, and 20% ($n = 4/20$) stated that the assessment was conducted in duplicate. Identifying the risk of bias tool is essential for readers to judge the appropriateness of the assessment methods, and conducting the assessment in duplicate can help to reduce reviewer biases in the evaluation process.

Thirty-three reviews included at least one pairwise meta-analysis. The study designs that were included in the meta-analysis were identified in most of the reviews (87.9%, $n = 29/33$). None of the reviews combined observational and experimental study designs in the same meta-analysis; review results may be misleading if researchers combine effects from different study designs into one summary measure, particularly if study design is not being examined as a source of heterogeneity in the meta-analysis (Higgins and Green, 2011; O'Connor and Sargeant, 2014). Of the meta-analyses in this study, both natural exposure and challenge trials were combined in at least one meta-analysis in 21% ($n = 7/33$) of the reviews. Trials with natural disease exposure and challenge trials may not be comparable as disease challenge may not reflect natural disease

exposure conditions or implications, and these trials are often conducted in more highly controlled settings (Sargeant *et al.*, 2014*a*, 2014*b*). As a result, combining data from both types of trials may result in summary effect measures that are not indicative of the true effects of interventions under natural exposure conditions.

In addition, 40% (*n* = 13/33) of the publications that included a meta-analysis reported at least one summary effect measure. A summary effect measure is a single point estimate that combines the effects and associated uncertainty measures from each of the individual primary studies included in the meta-analysis (O'Connor and Sargeant, 2014). The goal of meta-analysis often provides a precise estimate of the effect of a given intervention on a given outcome, and a summary effect measure is a tool by which this is achieved (O'Connor *et al.*, 2014). If a summary effect measure is not reported, then the reader of the meta-analysis has no indicator of the pooled impact of the intervention(s) being examined on the outcome(s) of interest. However, if heterogeneity is large, the summary effect may be misleading; thus, it is not always appropriate to report a summary measure. Evaluating whether or not summary effect measure reporting was appropriate given the amount of heterogeneity in each meta-analysis was beyond the scope of the current review.

Alternatively or in addition to the generation of a summary effect measure, a meta-analysis may evaluate sources of heterogeneity (i.e. the reasons why different studies report different results) (O'Connor *et al.*, 2014). Some authors contend that a meta-analysis should always include an evaluation of heterogeneity, with or without the calculation of a summary effect measure (O'Connor and Sargeant, 2014). More (70%, *n* = 23/33) of the meta-analyses in this study included an exploration of heterogeneity (expressed as an I^2 value). However, ten of the papers that included a meta-analytic component (30.3%) did not present either a summary effect measure or an I^2 value for any of the meta-analyses included in the review. In these cases, the purpose of undertaking a meta-analysis was therefore unclear.

Finally, 40% (*n* = 13/33) of the meta-analyses included an investigation into publication bias. Publication bias occurs when the published research on a topic is not representative of all completed studies on that topic for a variety of reasons, including but not limited to: outcome bias (reporting of the results, and the selection of results for reporting, of a primary study and the willingness of journal editors to publish those results, depends on the direction and statistical significance of those results); bias related to study size (large studies are more likely to detect statistically significant relationships than small studies and small studies are more likely to be published if they report statistically significant findings); duplication bias (some findings may be published multiple times, or may be suppressed if they are not novel); and competing interest bias (results may be selectively published based on the financial, political, or professional interests of researchers, funding agencies, journal editors, and others) (Rothstein, 2008; Higgins and Green, 2011). Where publication bias exists, the results of review studies may not be truly reflective of the existing body of research on the review topic. If publication bias is not investigated, readers of the review cannot evaluate the impact that any such bias may have on the results of the review. However, it should be noted that it is not always feasible to evaluate publication bias, particularly if the number of studies included in the meta-analysis is low; generally, at least ten primary studies must be included in a meta-analysis in order for statistical tests for publication bias to be meaningful (Sterne *et al.*, 2000; Ioannidis and Trikalinos, 2007). Determining whether or not

the meta-analyses in each of the reviews incorporated a sufficient number of studies to support an assessment of publication bias was beyond the scope of the present study.

Funding source(s) were declared by the authors in less than half of the reviews (42.1%, *n* = 16). The source of funding for any academic study could potentially introduce a source of bias, depending on the interests of the funding body (Bero, 2013; Lundh *et al.*, 2017). It should be noted that the empirical evidence concerning funding and bias is based on the human health literature, and the relationship between funding sources and the potential for bias has not been assessed in the animal health literature. However, the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines for the reporting of systematic reviews recommend that review authors report funding sources and the role(s) that funders played in the review process if any; this helps to ensure that the review process is transparent (Liberati *et al.*, 2009). Providing this information allows the reader to judge the potential for any such bias, and so failing to disclose funding information generates uncertainty about the validity of the results, thus lowering the reader's confidence in the results of the review.

Table 3 shows the number and percentage of reviews that met, partially met, or did not meet the criteria for each of the six AMSTAR 2 critical domains that were evaluated in this study. The table also includes a summary of the final quality assessment for the included reviews. According to the creators of the AMSTAR 2 tool, a weakness in any of these critical domains has serious implications for the validity of the results of a review (Shea *et al.*, 2017). When these six domains were evaluated for each review article, only three articles (7.9%) did not have any critical weaknesses (Ariza *et al.*, 2017; Nautrup *et al.*, 2017; Naqvi *et al.*, 2018). A further three reviews (7.9%) had one critical weakness, or failed to meet the criteria for one of the critical domains; these were classified as 'Low Confidence'. Finally, nearly 85% of the review articles (*n* = 32) failed under multiple critical domains, indicating that readers should have 'Critically Low Confidence' in the results of those reviews. Readers of these reviews must use caution when interpreting and applying results because these reviews may not provide an accurate, comprehensive synthesis of the existing body of literature (Shea *et al.*, 2017).

A concern is the number of meta-analyses that were undertaken without a corresponding systematic review. Two-thirds of the meta-analyses captured in the present review (66.7%, *n* = 22/33) were conducted without a corresponding systematic review component. Conducting a systematic review to inform the meta-analysis ensures that the data identification and collection process is transparent and comprehensive (Sargeant and O'Connor, 2016; O'Connor *et al.*, 2014, 2014*a*). Additionally, the risk of bias assessment component of a systematic review provides insight into the quality of the results of the primary studies, which may, in turn, impact the results of the meta-analysis (Sargeant and O'Connor, 2016). In this study, a larger proportion of meta-analyses that were conducted as a part of a broader systematic review had higher quality ratings than those reviews that presented a meta-analysis alone. Of those meta-analyses conducted without a supporting systematic review, only one rated 'High/Moderate Confidence', one rated 'Low Confidence', and 90.9% (*n* = 20/22) fell into the 'Critically Low Confidence' category. Of those studies that incorporated both a systematic review and a meta-analysis component, two received a 'High/Moderate Confidence' rating, and nine received a 'Critically Low Confidence' rating (81.8%, *n* = 9/11). When meta-analyses were conducted without a corresponding systematic

Table 3. AMSTAR 2 seven critical quality assessment domains applied to 38 systematic reviews and/or meta-analyses examining preventative approaches to reducing antibiotic use

	Yes		Partial Yes		No	
	Number	%	Number	%	Number	%
Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	6	15.8			32	84.2
Did the review authors use a comprehensive literature search strategy?	3	7.9	19	50.0	16	42.1
Did the review authors provide a list of excluded studies and justify the exclusions?	3	7.9	11	28.9	24	63.2
Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	8	21.1	12	31.6	18	47.4
If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?[a]	11	33.3	12	36.4	10	30.3
If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?[a]	13	39.4			20	60.6
	Number	%				
Final quality assessment						
High/Moderate confidence	3	7.9				
Low confidence	3	7.9				
Critically low confidence	32	84.2				

[a]This domain was only applied to those reviews in which a meta-analysis was performed. If a meta-analysis was not performed, this domain was not assessed and the score for this domain was not considered in the final quality assessment.

review, a larger proportion failed to fulfill the criteria for some of the critical domains. For example, 59.1% ($n = 13/22$) of the meta-analyses that did not involve a systematic review did not incorporate an adequate search strategy; for those meta-analyses that were supported by a systematic review, this percentage was only 18.2% ($n = 2/11$). Similarly, 63.6% ($n = 14/22$) of the unsupported meta-analyses failed to adequately assess the risk of bias, whereas only 27.3% of the combined systematic review/meta-analysis studies failed to meet this criterion. The evidence from our review supports the conclusion that meta-analyses should be conducted alongside a systematic review. However, there is room for improvement in the methodological quality of both systematic reviews and meta-analyses to ensure that they are valid sources of information for informed decision-making.

Additional considerations

In order to minimize the impact of publication bias and to ensure comprehensive coverage of the existing literature, the literature search components of systematic reviews should be extensive (Glanville *et al.*, 2013). Based on the AMSTAR 2 criteria, only 11 reviews (28.9%) incorporated a satisfactory literature search strategy. Within those reviews that reported at least some detail of their search, the number and types of databases searched varied considerably. Searching multiple sources increases the likelihood of retrieving all relevant records, but there is no consensus as to the correct number of databases that should be examined. The actual number of databases searched should depend on the nature of the review question, as well as time and budget constraints (Glanville *et al.*, 2013). For those articles that identified at least one database, the number of databases searched ranged from 1 to 14. The most commonly searched databases were MEDLINE via PubMed or other platforms, which was searched in 28 reviews (73.7%), followed by CABI or CAB Direct (60.5%, $n = 23$), AGRICOLA (31.6%, $n = 12$), Web of Science (23.7%, $n = 9$), Scopus (15.8%, $n = 6$), and Google Scholar (13.2%, $n = 5$). Thirty-five additional sources of published literature were identified in various reviews, each of which was searched in a small number of reviews (one, two, or three reviews). Some research questions require searches of more specialized information sources, such as LILACS, which covers medical journals from Latin America and the Caribbean (Glanville *et al.*, 2013), and was searched in one of the included reviews. Similarly, a wide range of grey literature sources was searched (16 different sources), the most common of which were pharmaceutical sources (21.1%, $n = 8$), conference proceedings (18.4%, $n = 7$), and contacting experts in the field (7.9%, $n = 3$).

The authors of the Cochrane Handbook identify three databases that they argue are the most important sources of primary studies for inclusion in Cochrane systematic reviews of human healthcare interventions: CENTRAL, MEDLINE, and EMBASE (now Embase) (Higgins and Green, 2011). Since the CENTRAL (Cochrane Controlled Register of Trials) database contains primarily randomized controlled trials in human medical research, this is not likely to be an appropriate source for the reviews captured in this study. MEDLINE was the most commonly searched database across the reviews in the present study ($n = 28$, 73.7%). Two reviews (5.3%) searched Embase. Although the databases recommended for reviews on human health topics may also contain some articles that are relevant to animal health, such databases may not be the most appropriate or useful sources to search for animal health reviews. In veterinary medicine, an investigation into the databases containing the best coverage of an extensive list of veterinary journals revealed that CAB Abstracts

had the most extensive coverage (90.2%) (Grindlay et al., 2012). In addition, CAB Abstracts indexed the highest number of unique journals. The authors of that study concluded that CAB Abstracts should be included in all searches designed to locate all published evidence on any veterinary topic and that the addition of other databases (Scopus, Science Citation Index/Web of Science, BIOSIS Previews, and Zoological Record) would slightly increase journal coverage. They further concluded that MEDLINE or Embase alone cannot be relied upon to provide comprehensive coverage of the animal health literature (Grindlay et al., 2012). Of the reviews captured in this study, 60.5% (n = 23) included CAB or CAB Direct in the search strategy, which implies that 40% of reviews did not. Those reviews that did not search CAB via CAB Direct or another platform may have missed relevant studies that were published in journals that were only indexed in that database.

Limitations

The quality assessment framework used in this study was based on the AMSTAR 2 critical appraisal tool. There were some challenges associated with the use of this tool to assess the quality of systematic reviews and meta-analyses in animal health literature. Some of the items were difficult to adapt for the present analysis due to the differences between human and animal health research; for instance, item nine on the AMSTAR 2 checklist details the features of an appropriate risk of bias assessment for synthesis studies in the human medical literature, but some of the components of human healthcare bias assessments may not be relevant in studies of animal health. For example, in confined livestock populations such as beef cattle, swine, and poultry, in which the animals are housed in groups, all animals are typically enrolled in a trial and the differential economic value of each animal may not be known at the time of allocation to treatment groups. As such, evaluating whether concealment occurred during the allocation process may not be relevant to a broader risk of bias assessment (Moura et al., 2019).

The number of meta-analyses presented in some of the reviews also made it difficult to apply the AMSTAR 2 criteria, as several of these criteria require detailed information about and evaluations of components of each individual meta-analysis. Two of the items on the AMSTAR 2 tool (items 12 and 13) suggest that authors include only randomized controlled trials with a low risk of bias in a meta-analysis or that authors include some form of discussion about the potential impact of including higher risk of bias studies. The tool does not include an elaboration on the appropriate features that such a discussion should include, and so an objective evaluation of these criteria is difficult. If authors choose to include only studies at low risk of bias, this decision must be made a priori and not following the risk of bias assessment; once trials have been assessed for risk of bias, deciding to include or exclude a trial on the basis of that assessment may introduce bias into the meta-analysis. The AMSTAR 2 tool does not provide an indication of the stage of the review process at which the decision to include studies based on features of bias must be made. As a result, these items were not evaluated according to the AMSTAR 2 criteria in the present study.

Additionally, the AMSTAR 2 tool allocates the same weight to all items on the assessment checklist, which may not be appropriate. The identification of the seven critical domains is useful, but these domains are not clearly indicated within the AMSTAR 2 checklist itself, and no weighting scheme is suggested beyond

the confidence level ratings. Further, some elements required by the AMSTAR 2 checklist do not seem to be particularly useful or relevant to the quality assessment of systematic reviews. For example, the fourth item on the AMSTAR 2 checklist, which details the criteria for an appropriate literature search, requires that review authors search the reference lists and bibliographies of all included primary studies in an effort to locate additional relevant studies. This is a time-consuming process and may yield few results. For example, in their review of antibiotics used to control bovine respiratory disease, Nautrup et al. (2017) identified only three citations in their hand search of the reference lists of relevant studies, compared to the 707 citations that were identified through database searches (Nautrup et al., 2017). In another animal health-related review, Theurer et al. (2015) found 1751 citations when they combined their search terms and applied them to three online databases, but they only identified one additional study by manually searching reference lists (Theurer et al., 2015); similarly, Larson and Step (2012) identified 703 potentially relevant studies through database searches but identified only four studies by hand-searching bibliographies (Larson and Step, 2012). That same checklist item requires the search to be conducted within 24 months of the completion of the review. However, determining the date of the completion of a review can be difficult, as the date at which academic papers are submitted for publication is seldom made available, and submission, peer review, and other steps that are a part of the publishing process can vary in duration.

Finally, some items on the AMSTAR 2 checklist are more relevant to the completeness of reporting, as opposed to methodological quality. Specifically, items 1, 8, 10, and 16 concern the reporting of the review research question, details of the studies included in the review, funding sources for the included studies, and sources of funding for the review, respectively. Although clear, comprehensive is essential, research quality and research reporting are separate issues. Comprehensive reporting guidelines for synthesis research are currently available elsewhere, such as PRISMA (Liberati et al., 2009).

Other quality assessment tools for synthesis research exist. For instance, the European Food Safety Authority (EFSA) developed a critical appraisal tool to evaluate the quality of systematic reviews of intervention studies. Similar to the AMSTAR 2 tool, the EFSA framework lists appraisal questions related to important steps of the systematic review process, such as 'Was the extensive literature search performed in an appropriate way?' and 'Were preventive steps taken to minimise bias and errors in the study selection process?' (EFSA, 2015). Each appraisal question is presented alongside a list of criteria to consider when evaluating each item, and space is provided to summarize the information presented in the review and to justify the final appraisal for each item in the tool. The actual appraisal is conducted on a four-point scale from 'Definitely appropriate' to 'Definitely not appropriate', with an additional 'Not Applicable' option (EFSA, 2015). However, this evaluation system does require subjective judgments about the adequacy of each element, and therefore the results of an evaluation based on this tool may not be reproducible. In the AMSTAR 2 framework, the criteria for appraisal are arguably more objective. For example, in applying the AMSTAR 2 checklist, the evaluator determines if more than two databases were searched and if the search string was provided, whereas the EFSA tool requires the evaluator to make a subjective determination as to whether 'too many' search concepts were used (EFSA, 2015).

AMSTAR 2 was developed for evaluations of systematic reviews related to human health. Much of the framework can be adapted for evaluating reviews of animal health topics, and the tool provides valuable insight into the strengths and weaknesses of review studies in general. An empirical investigation into those elements of the research process that specifically impact the results of systematic reviews in animal health may help to further refine quality assessment frameworks for applications in animal health research.

Conclusion

Thirty-eight reviews examining a broad range of commodity groups, interventions, and outcomes were identified for inclusion in this analysis. Based on our application of the AMSTAR 2 framework (Shea *et al.*, 2017), the quality of most of the systematic reviews and meta-analyses on preventive approaches to disease reduction is critically low, which implies that decision-makers must use caution when relying on the results of these reviews. Although there were challenges associated with the use of the AMSTAR 2 tool in this analysis, the AMSTAR 2 framework represents a comprehensive and objective tool available for the evaluation of systematic reviews and meta-analyses. An empirical investigation into which elements of quality assessment tools are most relevant to synthesis research evaluations in the animal health field may provide important insights for the continuing refinement of quality assessment frameworks and their applications for animal health research.

References

Abell KM, Theurer ME, Larson RL, White BJ and Apley M (2017) A mixed treatment comparison meta-analysis of metaphylaxis treatments for bovine respiratory disease in beef cattle. *Journal of Animal Science* **95**, 626–635.

Ariza JM, Relun A, Bareille N, Oberle K and Guatteo R (2017) Effectiveness of collective treatments in the prevention and treatment of bovine digital dermatitis lesions: a systematic review. *Journal of Dairy Science* **100**, 7401–7418.

AVMA (2019a) *Antimicrobial Stewardship Definition and Core Principles.* Schaumburg, IL: American Veterinary Medical Association. Available at https://www.avma.org/KB/Policies/Pages/Antimicrobial-Stewardship-Definition-and-Core-Principles.aspx (Accessed 11 March 2019).

AVMA (2019b) *AVMA Definitions of Antimicrobial Use for Treatment, Control and Prevention.* Shaumburg, IL: American Veterinary Medical Association. Available at https://www.avma.org/KB/Policies/Pages/AVMA-Definitions-of-Antimicrobial-Use-for-Treatment-Control-and-Prevention.aspx (Accessed 26 November 2019).

Bero LA (2013) Why the Cochrane risk of bias tool should include funding source as a standard item. *Cochrane Database of Syst Reviews* **12**, ED000075.

Cipriani A, Higgins JPT, Geddes JR and Salanti G (2013) Conceptual and technical challenges in network meta-analysis. *Annals Internal Medicine* **159**, 130–137.

CVMA (2017) *Veterinary Oversight of Antimicrobial Use: A Pan-Canadian Framework for Professional Standards for Veterinarians.* Ottawa: Canadian Veterinary Medical Association.

Dias S, Sutton AJ, Ades AE and Welton NJ (2013) Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making* **33**, 607–617.

European Food Safety Authority (2010) Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal* **8**, 1–90.

European Food Safety Authority (2015) *Tools for Critically Appraising Different Study Designs, Systematic Reviews and Literature Searches.* Parma, Italy: European Food Safety Authority (EFSA).

FAO (2016) *The FAO Action Plan on Antimicrobial Resistance 2016–2020.* Rome: Food and Agriculture Organization of the United Nations.

Glanville J, Wood H, Arber M, Varley D, Frampton G and Brazier H (2013) *Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety.* Parma, Italy: European Food Safety Authority (EFSA).

Grindlay DJ, Brennan ML and Dean RS (2012) Searching the veterinary literature: a comparison of the coverage of veterinary journals by nine bibliographic databases. *Journal of Veterinary Medical Education* **39**, 404–412.

Higgins JPT and Green S (eds) (2011) *Cochrane Handbook for Systematic Reviews of Interventions 5.1.0 [Updated March 2011].* The Cochrane Collaboration. Available at https://training.cochrane.org/handbook (Accessed 30 November 2019).

Ioannidis JP and Trikalinos TA (2007) The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal* **176**, 1091–1096.

Khan K, Kunz R, Kleijnen J and Antes G (2011) *Systematic Reviews to Support Evidence-Based Medicine: How to Review and Apply Findings of Healthcare Research,* 2nd Edn, London, UK: Hodder Arnold.

Larson RL and Step D (2012) Evidence-based effectiveness of vaccination against *Mannheimia haemolytica, Pasteurella multocida,* and *Histophilus somni* in feedlot cattle for mitigating the incidence and effect of bovine respiratory disease complex. *Veterinary Clinics: Food Animal Practice* **28**, 97–106.

Laxminarayan R, Duse A, Wattal C, Zaidi AK, Wertheim HF, Sumpradit N, Vlieghe E, Hara GL, Gould IM, Goossens H, Greko C, So AD, Bigdeli M, Tomson G, Woodhouse W, Ombaka E, Peralta AQ, Qamar FN, Mir F, Kariuki S, Bhutta ZA, Coates A, Bergstrom R, Wright GD, Brown ED and Cars O (2013) Antibiotic resistance-the need for global solutions. *The Lancet Infectious Disease* **13**, 1057–1098.

Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis J, Clarke M, Devereaux P, Kleijnen J and Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine* **6**, e1000100.

Lundh A, Lexchin J, Mintzes B, Schroll JB and Bero L (2017) Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews* **2**, MR000033.

Moura CAA, Totton S, Sargeant J, O'Sullivan T, Linhares D and O'Connor AM (2019) Evidence of improved reporting of swine intervention trials in the post-REFLECT statement publication period. *Journal of Swine Health and Production* **27**, 265–277.

Naqvi SA, Nobrega DB, Ronksley PE and Barkema HW (2018) Invited review: effectiveness of precalving treatment on postcalving udder health in nulliparous dairy heifers: a systematic review and meta-analysis. *Journal of Dairy Science* **101**, 4707–4728.

Nautrup BP, Ilse Van Vlaenderen DVM, Decker M and Cleale RM (2017) Antimicrobial drug use for control and treatment of bovine respiratory disease in US feedlot cattle: a meta-analysis of comparative studies versus tulathromycin. *The Bovine Practitioner* **51**, 1–13.

O'Connor AM and Sargeant JM (2014) Meta-analyses including data from observational studies. *Preventive Veterinary Medicine* **113**, 313–322.

O'Connor AM, Sargeant J and Wang C (2014) Conducting systematic reviews of intervention questions III: synthesizing data from intervention studies using meta-analysis. *Zoonoses and Public Health* **61**(S1): 52–63.

OIE (2016) *The OIE Strategy on Antimicrobial Resistance and the Prudent use of Antimicrobials.* Paris: World Organization for Animal Health.

Rothstein HR (2008) Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology* **4**, 61–81.

Sargeant J and O'Connor AM (2014a) Introduction to systematic reviews in animal agriculture and veterinary medicine. *Zoonoses and Public Health* **61** (S1): 3–9.

Sargeant JM and O'Connor AM (2014b) Conducting systematic reviews of intervention questions II: relevance screening, data extraction, assessing risk of bias, presenting the results and interpreting the findings. *Zoonoses and Public Health* **61**(S1): 39–51.

Sargeant J and O'Connor AM (2016) Potential for meta-analysis in the realm of preharvest food safety. *Microbiology Spectrum* **4**, 1–12.

Sargeant J, Kelton D and O'Connor AM (2014a) Study designs and systematic reviews of interventions: building evidence across study designs. *Zoonoses and Public Health* **61**(S1): 10–17.

Sargeant JM, Kelton DF and O'Connor AM (2014b) Randomized controlled trials and challenge trials: design and criterion for validity. *Zoonoses and Public Health* **61**(S1): 18–27.

Scherer RW, Meerpohl JJ, Pfeifer N, Schmucker C, Schwarzer G and von Elm E (2018) Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews* **11**, MR000005.

Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E and Henry DA (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* **358**, j4008.

Sterne JA (2013) Why the Cochrane risk of bias tool should not include funding source as a standard item. *Cochrane Database of Systematic Reviews* **12**, ED000076.

Sterne JA, Gavaghan D and Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* **53**, 1119–1129.

Theurer ME, Larson RL and White BJ (2015) Systematic review and meta-analysis of the effectiveness of commercially available vaccines against bovine herpesvirus, bovine viral diarrhea virus, bovine respiratory syncytial virus, and parainfluenza type 3 virus for mitigation of bovine respiratory disease complex in cattle. *Journal of the American Veterinary Medical Association* **246**, 126–142.

Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, Murad MH, Treadwell JR and Kane RL (2018) Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. *Journal of Clinical Epidemiology* **97**, 26–34.

Vriezen R, Sargeant JM, Vriezen E, Reist M, Winder CB and O'Connor AM (2019) Systematic reviews and meta-analyses in animal health, performance, and on-farm food safety: a scoping review. *Animal Health Research Reviews*. doi: 10.1017/S1466252319000197.

WHO (2012) *The Evolving Threat of Antimicrobial Resistance: Options for Actions*. Geneva: World Health Organization.

WHO (2015) *Global Action Plan on Antimicrobial Resistance*. Geneva: World Health Organization.