Check for updates

SYSTEMATIC REVIEW

REVISED Efficacy of antimicrobial treatment for bovine respiratory disease: a systematic review and meta-analysis for European Network for Optimization of Veterinary Antimicrobial Therapy guidelines

[version 2; peer review: 3 approved with reservations]

Karolina Scahill[1,2], Mickael Cargnel[3], Ramazan Yildiz[4], Sébastien Buczinski[5], Clair Firth[6], Bart Pardon[7], Maria Stokstad[8], Pierre-Louise Toutain[9], Amelia R Woolums[10], Lise Marie Ånestad[11], Lisbeth Rem Jessen[1], Luca Guardabassi[12], Luis Pedro Carmo[13]

[1]Department of Veterinary Clinical Science, University of Copenhagen, Frederiksberg C, Denmark
[2]Infectious disease, The University of Edinburgh College of Medicine and Veterinary Medicine, Edinburgh, Scotland, UK
[3]Infectious Diseases in Animals, Sciensano, Brussels, 1050, Belgium
[4]Internal medicine, Burdur Mehmet Akif Ersoy University, Burdur, Burdur Province, Turkey
[5]Département des sciences cliniques, Universite de Montreal, Montreal, Québec, Canada
[6]Clinical Department for Farm Animals and Food System Science, University of Veterinary Medicine, Vienna, Vienna, 1210, Austria
[7]UGent Calf Health, Department of Internal Medicine, Reproduction and Population Medicine, Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, Merelbeke, 9820, Belgium
[8]Department of Production Animal Clinical Sciences, Norwegian University of Life Sciences, Ås, Norway
[9]Comparative Biomedical Sciences, The Royal Veterinary College, London, England, UK
[10]Department of Pathobiology and Population Medicine, Mississippi State University College of Veterinary Medicine, Mississippi State University, Mississippi, USA
[11]Norwegian Veterinary Institute, Ås, Norway
[12]Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C, 1870, Denmark
[13]Veterinary Public Health Institute, University of Bern, Bern, Switzerland

v2 **First published:** 12 Feb 2026, **6**:41
https://doi.org/10.12688/openreseurope.21879.1
Latest published: 30 Apr 2026, **6**:41
https://doi.org/10.12688/openreseurope.21879.2

Abstract

Bovine respiratory disease (BRD) is a leading cause of illness and death in cattle and a major driver of antimicrobial use. This systematic review and meta-analysis assessed the clinical effectiveness of five commonly used antimicrobial classes (penicillins, tetracyclines, phenicols, macrolides, and fluoroquinolones) in the treatment of cattle with natural BRD. Cattle from all production systems (feedlot, dairy, slaughter, and stocker) were included, and metaphylaxis trials were excluded. Peer-reviewed trials were searched in several databases, and all reports were screened in duplicate for data

Open Peer Review

Approval Status ? ? ?

	1	2	3
version 2 (revision) 30 Apr 2026	? view		? view
version 1 12 Feb 2026	? view	? view	

1. **Elina B Reinoso**, National University of Río

extraction and risk of bias. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) methodology was employed to evaluate the certainty of evidence. Twenty-eight randomized controlled trials were included. We did not detect clinically meaningful differences between antimicrobial classes in reducing the need for re-treatment (a surrogate for short-term therapeutic failure). Certainty of evidence ranged from very low to moderate, and results should be interpreted cautiously. Subgroup estimates were broadly consistent for production systems (feedlot and non-feedlot) and for trials with *Mycoplasma*-positive animals; however, these subgroup analyses were small and imprecise and do not allow firm conclusions. Mortality, relapse, and growth outcomes were measured inconsistently and could not be reliably pooled. Overall, we did not identify any class with a clinically meaningful short-term advantage based on our prespecified thresholds. The results of this systematic review and meta-analysis will inform the BRD Antimicrobial Use Guidelines of the European Network for Optimization of Antimicrobial Therapy (ENOVAT).

Plain Language Summary

Bovine respiratory disease (BRD) is the leading cause of illness and a major cause of death in cattle. It is also a key reason antibiotics are used in herds. Using antibiotics widely can fuel antimicrobial resistance. We asked: when a sick animal with BRD is treated, do some antibiotic classes work better than others?

We systematically searched for peerreviewed randomised controlled trials comparing five commonly used antibiotic classes—penicillins, tetracyclines, florphenicols, macrolides and fluoroquinolones—for individual treatment (not mass preventive treatment) of natural BRD in any production system worldwide. Two reviewers independently screened studies, extracted data and assessed bias. We pooled results when possible and judged certainty using the GRADE approach. The main outcome was the need for retreatment (a proxy for shortterm treatment failure).Twentyeight trials met the criteria. Across all headtohead comparisons, no antibiotic class showed a clinically important advantage over another in reducing retreatment. This pattern held in subanalyses of feedlot versus other systems and in studies where Mycoplasma bovis (a bacterium naturally resistant to some drugs) was present. Other important outcomes—death, relapse, and weight gain—were reported inconsistently and could not be combined. Certainty of the evidence ranged from very low to moderate, mainly because the outcome (retreatment) was a surrogate and because several trials had risks of bias.

For firstline treatment of BRD, the five antibiotic classes studied appear to perform similarly in the short term. This allows vets and farmers to give more weight to responsible use and publichealth considerations when choosing a drug. Many studies were small, used different followup times, or lacked details on key outcomes. Results do

Cuarto,, Río Cuarto, Argentina

2. **Michael D Apley**, Kansas State University, Manhattan, USA

3. **Majed H Mohammed**, University of Adelaide, Adelaide, Australia

Any reports and responses or comments on the article can be found at the end of the article.

not show effects on growth or survival. These findings will inform ENOVAT BRD treatment guidelines. Policymakers are also encouraged to use them.

Keywords
Antimicrobial therapy, therapy failure, respiratory tract infections, mycoplasma, inflammation, bovine

This article is included in the Animal and Dairy Science gateway.

This article is included in the COST Actions gateway.

Corresponding author: Karolina Scahill (karolina.scahill@sund.ku.dk)

Author roles: Scahill K: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Cargnel M**: Data Curation, Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Yildiz R**: Data Curation, Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Buczinski S**: Conceptualization, Investigation, Writing – Review & Editing; **Firth C**: Conceptualization, Investigation, Writing – Review & Editing; **Pardon B**: Conceptualization, Investigation, Writing – Review & Editing; **Stokstad M**: Conceptualization, Investigation, Writing – Review & Editing; **Toutain PL**: Conceptualization, Investigation, Writing – Review & Editing; **Woolums AR**: Conceptualization, Investigation, Writing – Review & Editing; **Ånestad LM**: Conceptualization, Investigation, Project Administration, Writing – Review & Editing; **Rem Jessen L**: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Guardabassi L**: Conceptualization, Investigation, Project Administration, Writing – Review & Editing; **Carmo LP**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Review & Editing

Competing interests: SB has received funding for research projects on BRD from Merck Animal Health, Zoetis, and CEVA Animal Health. He has also received honorarium from pharmaceutical companies selling products for the treatment and prevention of BRD (Merck Animal Health, Zoetis, Vetoquinol, CEVA Animal health and HIPRA) to provide continuing education and seminars to veterinary practitioners. In the past decade, BP has received speaker honorarium, provided consultancy activities, or received research funding from pharmaceutical (Zoetis, MSD Animal Health, Boehringer Ingelheim, Vetoquinol, Ceva, Dechra, Elanco and Hipra) agricultural (Trouw Nutrition, Boerenbond n.v.), and chemical (Proviron) companies, as well as from non-profit organizations (DGZ Vlaanderen and AMCRA vzw). BP is a co-founder of qTUS, a company active in the training and certification of lung ultrasonography in calves. ARW received funding for research and consulting from companies that market antimicrobials for BRD. LG received funding for research collaboration and consultation with CEVA. ARW has received speaker honoraria, provided consultancy activities for, or received research funding from companies marketing pharmaceuticals and/or biologicals (Bayer Animal Health, Boehringer Ingelheim, Ceva, Elanco, Medgene, Merck/MSD Animal Health, Phibro, SAb Biotherapeutics, and Zoetis), and is a member of the advisory board of Medgene. KS and LRJ are members of GRADE.

Grant information: This project has received funding from the European Union's Framework Programme for Research & Innovation as part of the COST Action [CA18217], as supported by the COST Association (European Cooperation in Science and Technology).

Copyright: © 2026 Scahill K *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Scahill K, Cargnel M, Yildiz R *et al.* **Efficacy of antimicrobial treatment for bovine respiratory disease: a systematic review and meta-analysis for European Network for Optimization of Veterinary Antimicrobial Therapy guidelines [version 2; peer review: 3 approved with reservations]** Open Research Europe 2026, **6**:41
https://doi.org/10.12688/openreseurope.21879.2

First published: 12 Feb 2026, **6**:41 https://doi.org/10.12688/openreseurope.21879.1

REVISED **Amendments from Version 1**

This revised version incorporates clarifications and corrections in response to peer review, without changing the study's conclusions.

Scope and interpretation: We added brief statements in the introduction and methods to clarify that this SR synthesizes class level, head to head RCT evidence and does not make recommendations; those are addressed in the companion ENOVAT guideline's Evidence to Decision framework. We also softened wording throughout (Abstract, Discussion, Conclusion) to emphasize cautious interpretation given very low to moderate certainty in some PICOs.

Terminology: We standardized nomenclature for Mycoplasma bovis (syn. Mycoplasmopsis bovis), used "Mycoplasma positive/M. bovis–positive" consistently, and retitled Table 4 "Mycoplasma status" with a clarifying footnote.

Outcomes: We clarified in Methods that re treatment was used as a surrogate for short term therapeutic failure and that certainty was downgraded for indirectness; "relapse" was not pooled.

Heterogeneity/inconsistency: We added a paragraph explaining our threshold based approach (absolute effects against prespecified clinical thresholds) and illustrated it with an influential study (Schunicht 2007); conclusions were unchanged.

Data corrections: We removed Catry et al. (2008) from PICO 6 (TET vs FLO) because it was a metaphylaxis arm. The study is retained only under PICO 8 (FLO vs MAC) using affected animal treatment data. Accordingly, we updated the PICO 6 pooled estimate to RR 0.96 (95% CI 0.70–1.30), I^2=0%, and revised Table 5 and Figure 5; corrected the production system subgroup for PICO 6 (Table 8 and text); removed PICO 6 from the Mycoplasma positive subgroup (Table 7 and text).

Subgroups: We softened the Mycoplasma positive and production system subgroup text to note small, imprecise estimates that do not allow firm conclusions.

Applicability/language: We revised the US feedlot paragraph and acknowledged the English language restriction as a limitation.

These changes improve clarity and transparency; the overall conclusions (no clinically meaningful between class differences on re treatment; very low to moderate certainty) are unchanged.

Any further responses from the reviewers can be found at the end of the article

Introduction

Bovine respiratory disease (BRD) is the leading cause of morbidity and a major cause of mortality in cattle, especially in systems that purchase and commingle calves from different origins, such as the feedlot, veal calf, and dairy beef industries (Duff & Galyean, 2007; Renaud & Pardon, 2022; USDA, 2011). BRD is a multifactorial syndrome that involves interactions between host susceptibility, stressors (weaning, commingling, transport, and weather), and a suite of viral and bacterial pathogens. Because BRD affects young stock at critical growth stages and feedlot cattle during high-value finishing periods, it carries substantial and recurrent economic costs that extend well beyond treatment (Blakebrough-Hall *et al.*, 2020; Wang *et al.*, 2018). The global economic losses associated with BRD have not been well characterized but have been estimated to exceed $3 billion annually (Watts & Sweeney, 2010).

Prevalence estimates vary widely according to the production system, region, season, and risk status of the cattle (Duff & Galyean, 2007; USDA, 2011). In US feedlots, BRD is the most frequently treated clinical disease and a leading cause of morbidity and mortality, making it a dominant health issue in this sector (USDA NAHMS Feedlot, 2011). Reported rates vary widely by risk class, season, and management; high-risk, newly received cattle can experience markedly higher BRD burdens than lower-risk placements (Duff & Galyean, 2007). In dairy systems, the incidence of preweaning BRD in heifer calves can fall between 4% and 25%, with respiratory disease responsible for a notable fraction of preweaning deaths and for much of the antimicrobial use in young stock (Donlon *et al.*, 2023; Urie *et al.*, 2018; USDA, 2011; Windeyer *et al.*, 2014). In the veal calf system, BRD is the most frequent cause of morbidity and mortality (Pardon *et al.*, 2012; Pardon *et al.*, 2013).

Given the prevalence described, BRD shapes priorities in welfare, antimicrobial stewardship, and sustainability. Respiratory outbreaks drive a large share of antimicrobial use in cattle, and chronic cases are associated with poor welfare outcomes (USDA, 2017). Although antimicrobial therapy is an effective method for managing BRD, its use can drive the development of antimicrobial resistance (AMR) (Bell *et al.*, 2014). AMR is a serious global health threat, with significant economic consequences due to treatment failure and increased morbidity and mortality in humans (Cassini *et al.*, 2019). Owing to the occurrence of resistant bacteria in animals and the zoonotic risk associated with animal-to-human transmission, antimicrobial stewardship in livestock is warranted to minimize this risk. Because antimicrobials are widely used in livestock and can select for resistant bacteria that can be transmitted to humans and contribute to this

burden (Morel, 2019), prudent antimicrobial use in the veterinary sector is essential to minimize zoonotic risk and preserve the effectiveness of critically important antimicrobials (EMA, 2025; WHO, 2024). As BRD is one of the leading reasons for antimicrobial usage in cattle production (Guardabassi *et al.*, 2018), establishing evidence-based guidelines for antimicrobial therapy in BRD is crucial for mitigating the risk of AMR while maintaining animal health and welfare. Achieving this requires a thorough evaluation of the clinical effectiveness of the antimicrobials commonly used to treat BRD.

The primary aim of this systematic review and meta-analysis was to assess the clinical effectiveness of commonly used classes of antimicrobials for the treatment of BRD. The results will inform the BRD Antimicrobial Use Guidelines of the European Network for Optimization of Antimicrobial Veterinary Therapy (ENOVAT) and European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group for Veterinary Microbiology (ESGVM). The ENOVAT guidelines encourage policymakers to use the results of this study to develop regional or national antimicrobial treatment guidelines for BRD.

This systematic review was prespecified to estimate comparative effects at the antimicrobial class level from direct, head-to-head randomized trials. It does not issue recommendations or place value judgements on the results; those decisions are made in the companion ENOVAT guideline using a structured Evidence-to-Decision (EtD) framework. Within-class (substance-level) comparisons and implementation trade-offs (feasibility, stewardship, values and preferences) are beyond the scope of this review and are addressed in the guideline.

Materials and methods

Reporting standards and protocol

This systematic review adhered to the PRISMA guidelines (Page *et al.*, 2021). The PRISMA checklist and protocol deviations are shown in Supplementary File 1. The protocol for this systematic review and meta-analysis was registered with Syreaf. org in July 2022 (https://syreaf.org/wp-content/uploads/2022/07/BRD-PRISMA-protocol_version-for-upload_2022-07.pdf).

Guideline and panel organization

The guideline group was chaired by LG. KS, LPC, and LRJ formed the methodological oversight committee, which was responsible for the harmonization of methodological approaches across the ENOVAT consortium. The panel consisted of seven bovine experts and clinicians (AW, SB, CF, MS, BP, RY, and LMÅ), one microbiologist (LG), and one pharmacologist (PLT). The methodological task force were non-voting members of the group (KS, MC, LRJ, and LPC), except for one member that was also part of the panel (RY).

Population, intervention, comparator and outcome (PICO) generation

The clinicians of the panel drafted ten PICO questions (Table 1) that compared the efficacy of different antimicrobial classes. The included antimicrobials are EMA categories D, C, and B antimicrobials that European bovine practitioners may choose between in their clinical decision-making process. In Scandinavia, the choice of treatment is generally between penicillin (PNC) and tetracyclines (TET). In other parts of Europe, such as Belgium, clinicians often choose between florfenicol (FLO) and macrolides (MAC). Because *Mycoplasma bovis* (syn. *Mycoplasmopsis bovis*) is intrinsically resistant to certain antimicrobial classes, trials assessing the antimicrobial efficacy of BRD associated with this organism were analyzed as a subgroup. Hereafter, we use *Mycoplasma bovis* (*M. bovis*). Feedlot animals were also analyzed as a subgroup because their status could potentially act as an effect modifier in the pooled results. Risk in feedlots varies by cohort; newly received, commingled, lightweight calves are often higher-risk for BRD, whereas many yearling placements are lower-risk. Because some feedlot cohorts may also have had prior BRD treatment, we explored feedlot versus non-feedlot systems as a potential effect modifier in an exploratory subgroup analysis.

The review question and analyses were defined a priori at the antimicrobial class level to inform guideline decisions typically made at the class level (aligned with WHO/EMA–AMEG stewardship frameworks). This review summarizes effects and certainty; it does not make recommendations. Substance-level (within-class) comparisons were out of scope and were not sought or synthesized. The companion ENOVAT guideline integrates these effect estimates with feasibility, acceptability, values/preferences, and stewardship considerations in EtD frameworks to produce recommendations.

Table 1. PICO questions including sub-groups, intervention and comparator, outcomes and critical outcomes. The European Medicine Agency (EMA) antimicrobial categories A-D are shown in brackets.

PICO	Population	Intervention (EMA category)	Comparator (EMA category)	Outcomes	Critical outcomes
1	Cattle with undifferentiated pneumonia	Penicillins (D)	Tetracyclines (D)	Therapeutic failure (resolution of clinical disease <7 days). Relapse (need for re-treatment) Mortality Adverse effects Weight gain.	Therapeutic failure (resolution of clinical disease <7 days). Relapse (need for re-treatment) Mortality Adverse effects
2	Cattle with undifferentiated pneumonia	Penicillins (D)	Macrolides (C)		
3	Cattle with undifferentiated pneumonia	Penicillins (D)	Phenicols (C)		
4	Cattle with undifferentiated pneumonia	Penicillins (D)	Fluoroquinolones (B)		
5.	Cattle with undifferentiated pneumonia	Tetracyclines (D)	Macrolides (C)		
6	Cattle with undifferentiated pneumonia	Tetracyclines (D)	Phenicols(C)		
7	Cattle with undifferentiated pneumonia	Tetracyclines (D)	Fluoroquinolones (B)		
8	Cattle with undifferentiated pneumonia	Phenicols (C)	Macrolides (C)		
9	Cattle with undifferentiated pneumonia	Macrolides (C)	Fluoroquinolones (B)		
10	Cattle with undifferentiated pneumonia	Phenicols (C)	Fluoroquinolones (B)		

Open Research Europe 2026, 6:41 Last updated: 15 OCT 2026

Outcome definitions

- Preferred critical outcome: Short-term therapeutic failure, defined a priori as lack of clinical resolution within 7 days after initial antimicrobial treatment.

- Surrogate used for pooling: Need for retreatment, defined as the first administration of an additional antimicrobial for the same BRD episode within the study's observation window. In most studies this occurs within the first few days (< 7 days) but is not always explicitely stated and because the observation window varied across trials, we used retreatment as a surrogate for short-term therapeutic failure and rated down for indirectness.

- Relapse: Recurrence of BRD clinical signs requiring treatment after documented initial clinical resolution. This generally occurs after another antimicrobial has been administered. Relapse was therefore not pooled.

Clinical thresholds and stakeholder involvement

Clinical thresholds are used to assess the magnitude of potential health benefits and harms in order to develop sensible recommendations. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach categorizes these effect sizes as large, moderate, small, or trivial effects with explicit thresholds that are of clinical importance to stakeholders. These thresholds are also used to judge the precision of an effect (Zeng *et al.*, 2022). The clinicians of the panel drafted clinical threshold surveys. Stakeholders, including farmers (n = 23) and field clinicians (n = 28) from Norway, Austria, Belgium, Canada, and the US, were consulted in the prioritization of outcomes and development of clinically relevant thresholds using semi-structured interviews and surveys. Questions posed in the veterinarian and farmer surveys and results can be found in Supplementary File 2. Surveys were completed via telephone. Clinicians from each country were responsible for contacting the surveyed stakeholders and recording their opinions on acceptable treatment failure and mortality rates for animals with BRD. The median values of acceptable and unacceptable therapeutic failure and mortality by stakeholders were used to establish small and large clinical thresholds. Values between small and large were used to establish a moderate effect. The clinical thresholds in Table 2 were prespecified for therapeutic failure. When re-treatment was used as the surrogate for therapeutic failure, these thresholds were applied to the absolute effects.

Search strategy and eligibility criteria

The search was performed according to the protocol on the July 5, 2022, and was updated on the June 1, 2025. Librarians with expertise in search development and information retrieval were consulted for the design of the search strategy. Peer reviewed controlled trials of cattle with natural infections, and the clinical signs of BRD were eligible for inclusion. Cattle of all ages and production systems, irrespective of bacterial etiology, were included, as well as trials with unknown bacterial agents. Other co-treatments, including metaphylaxis prior to trial enrolment, were accepted if both the intervention and comparator groups received the treatment. Trials that solely evaluated the effect of metaphylaxis, articles written in languages other than English, and conference proceedings were excluded. Grey literature was not searched, but the reference lists of all included studies and a previous systematic review (O'Connor *et al.*, 2016) were searched for relevant studies.

Study selection and data collection

Three reviewers (MC, RY, and KS) independently screened titles, abstracts, and full texts in duplicate using an online tool (www.rayyan.ai) for systematic reviews. Covidence (www.covidence.org) was used in the updated search for 2025. Calibration exercises were conducted for every step of the study and the data collection processes. This was done to fine-tune the process and increase standardization across reviewers. Data extraction was performed independently in duplicate (KS, RY, MC) and included information on the study population and baseline characteristics (age, production system), assessment methodology, microbiological samples (methods, isolated bacteria, presence of *Mycoplasma*, *in vitro* antimicrobial resistance), intervention and comparator (antimicrobial, dosage, frequency, and duration). Potential

Table 2. Clinical thresholds. The number of acceptable treatment failures or mortality per 1000 animals. The thresholds were used in the assessment of the imprecision and inconsistency domain to determine a clinically relevant effect size.

Outcome	Trivial effect	Small effect	Moderate effect	Large effect
Therapeutic failure	<200 animals	200–250 animals	250–300 animals	>300 animals
Mortality	<35 animals	35–50 animals	50–75 animals	>75 animals

Table 3. Certainty definitions and short descriptions of the five domains used to grade the evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE). [a]Cited from Guyatt *et al.* (2025); [b]Adapted from Guyatt *et al.* (2025); [c]CI, confidence interval.

Certainty[a]	
Very low	The true effect is probably markedly different from the estimated effect
Low	The true effect might be markedly different from the estimated effect
Moderate	The authors believe that the true effect is probably close to the estimated effect
High	The authors have a lot of confidence that the true effect is similar to the estimated effect
Domains[b]	
Risk of bias	Bias occurs when the results of a study do not represent the truth because of inherent limitations in design or conduct of a study. The body of evidence is rated at the outcome level rather than the study level.
Inconsistency	When several studies show consistent effects the certainty of evidence is higher. Similarity of point estimates and the overlap of CI are evaluated in the rating process.
Indirectness	Evidence is most certain when studies directly compare the interventions of interest in the population of interest, and report the outcome(s) critical for decision-making. Certainty can be rated down if the patients studied are different from those for whom the recommendation applies. Indirectness can also occur when the interventions studied are different than the real outcomes.
Imprecision[c]	This domain relies on thresholds and CI of absolute effects as a primary criterion for imprecision rating (i.e., CI approach). Certainty is downgraded when confidence intervals cross predefined clinical thresholds (trivial, small, moderate, large) that are of importance to stakeholders. If the relative effect is large the optimal information size is also considered.
Other considerations	Publication bias can contribute to further downgrading, but the consideration of large effects, plausible confounding, and dose response gradients can occasionally lead to upgrading of the certainty of evidence.

conflicts of interest and funding were extracted into a data management tool (Excel, Microsoft). Disagreements between reviewers were resolved through discussion.

Data synthesis and certainty (quality) assessment

A direct pairwise meta-analysis was performed using a random-effects model to estimate the pooled effect risk ratio across studies. Cochrane Review Manager software edition 7.12.0 (revman.cochrane.org) was used to generate the meta-analyses. Descriptive information was extracted for the data that could not be pooled. The GRADE methodology was employed to evaluate the certainty of the evidence and was assessed by KS and LRJ (Guyatt *et al.*, 2025). The panel and methodology task forces were involved in the discussion and final decisions regarding certainty assessment. Inconsistency among the included studies was primarily evaluated by visual inspection to identify outliers in relation to the chosen thresholds, and not based on I^2 statistics, which is in line with the updated GRADE guidelines (Guyatt *et al.*, 2023). Imprecision was assessed using confidence intervals (CI), and the effect estimate was considered precise if the 95% CI did not cross clinical thresholds. (Zeng *et al.*, 2022). The GRADEpro software (www.gradepro.org) was used to produce summary of findings (SoF) tables and risk difference calculation for the anticipated absolute effect. There was no risk of bias tool developed specifically for cattle, but the modified Cochrane risk-of-bias tool (RoB 2.0) for randomized trials in swine (Moura *et al.*, 2019) was used, as it was also determined to be relevant to cattle. The risk of bias assessment was performed independently by two reviewers (KS and LRJ) in duplicate. Conflicts were resolved through discussions. The certainty of evidence definitions and descriptions of the five domains used to judge the certainty of evidence are presented in Table 3.

Results

Included studies (PICO 1–10)

Twenty-eight randomized controlled trials met the inclusion criteria. The number of excluded reports and reasons for the exclusion of retrieved texts are reported in Figure 1. Study characteristics such as country, production system, age, antimicrobial substance including dosage, duration of treatment, study duration, and *Mycoplasma* status are reported for all trials that were pooled in Table 4. Mortality, relapse, and weight gain outcomes have been reported after the use of several different antimicrobials in most studies, and causation could not be reliably established. These outcomes are, therefore, not pooled but are reported in Supplementary File 3. The data extraction rationale and further details of the diagnostic assessments and funding are also reported in Supplementary File 3.

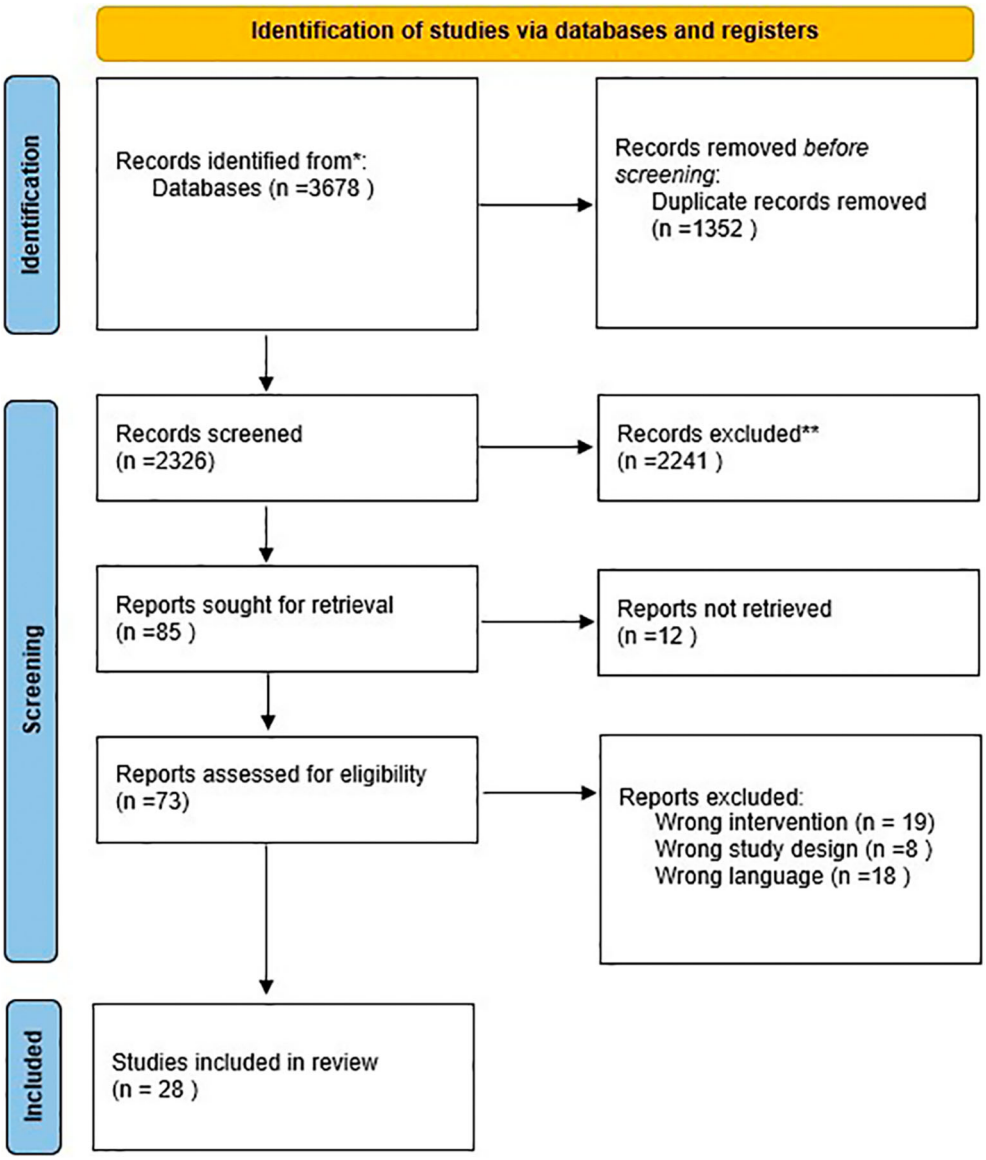

Figure 1. PRISMA flow diagram show the number of screened records, reason for exclusion and number of included studies for PICO 1–10.

PICO 1: Three trials compared penicillin and tetracycline (PNC vs. TET) in a total of 518 animals (Bateman *et al.*, 1990; Mechor *et al.*, 1988; Welling *et al.*, 2020).

PICO 2: No trial compared penicillin and macrolides (PNC vs. MAC).

PICO 3: One trial compared penicillin and florfenicol (PNC vs. FLO) in 77 animals (Welling *et al.*, 2020).

PICO 4: No trials compared penicillin and fluoroquinolones (PNC vs. QUI).

PICO 5: Three trials compared tetracycline with macrolides (TET vs. MAC). Three studies were pooled and included 171 animals (Keita *et al.*, 2007; Laven & Andrews, 1991; Musser *et al.*, 1996).

PICO 6: Two trials compared tetracycline and florfenicol (TET vs. FLO) and provided sufficient information to pool the results, including 873 animals (Schunicht *et al.*, 2002; Welling *et al.*, 2020). One trial assessed lung aeration as an outcome and was described narratively (Jourquin *et al.*, 2022) in Supplementary file 3.

Table 4. Study characteristics for pooled studies in PICO 1–10. [a]Days when pooled outcome was measured. i.m. = intramuscular, s.c. = subcutaneous, SID = once a day, EOD = every other day, PNC = penicillin, OX = oxytetracycline, FLO = florfenicol, MAC = macrolide, QUI = fluoroquinolone. [b]*M.bovis* indicates that *Mycoplasma bovis* was detected in some animals otherwise only *Mycoplasma* spp. were reported.

PICO	First author (year)	Country	Production system	Age	Intervention	Comparator	Duration of treatment	Study duration [a]	*Mycoplasma* status[b]
1	Bateman *et al.* (1990)	Canada	Feedlot	4–7 months	Penicillin (45.000 IU/kg, s.c.,SID)	Oxytetracycline (10 mg/kg, i.m., SID)	4 days (if responding to treatment, if no response on day 3 antimicrobial was changed to second line)	3	Unknown. No sampling
1	Mechor *et al.* (1988)	Canada	Feedlot	Recently weaned (250–320 kg)	Penicillin (65.000 IU, i.m. SID)	Oxytetracycline (20 mg/kg, i.m., SID)	3–4 days. If temperature was abnormal on day 4 treatment was changed to second line.	60	Unknown. Samples not tested for Mycoplasma
1, 3, 6	Welling *et al.* (2020)	Sweden	Slaughter	<6 months	Penicillin (40 mg/kg, i.m., SID, 5 doses)	Oxytetracycline (20 mg/kg, i.m., EOD, 2 doses) AND Florfenicol (20 mg/kg, i.m., EOD, 2 doses)	PNC 5 days, OX and FLO two long-acting injections 48 h apart.	60	Unknown. No sampling.
5	Keita *et al.* (2007)	France	Slaughter	8–90 days	Oxytetracycline (20 mg/kg, i.m.)	Tilmicosin (10 mg/kg, s.c.)	One injection	10	Yes. Cultured on 2 farms and positive seroconversion (*M.bovis*) on 3 farms
5	Laven & Andrews (1991)	United Kingdom	Beef	4–8 months	Oxytetracycline (20 mg/kg, i.m.)	Tilmicosin (10 mg/kg, s.c.)	Most animals were treated with one or two injections. One animal was given a third injection.	28	Unknown. Samples collected but results not reported
5	Musser *et al.* (1996)	USA	Dairy	<3 months	Oxytetracycline (20 mg/kg, i.m.)	Tilmicosin (10 mg/kg, s.c.)	One injection	Not reported	Yes in 22/40 samples (*M.bovis*)
8	Catry *et al.* (2008)	Netherlands	Veal	2 weeks to 3 months	Tilmicosin (12,5 mg/kg, BID, p.o.)	Florfenicol (40 mg/kg, s.c.)	Not clear, but tilmicosin seem to have been given for 5 days.	7	Yes (*M. bovis)*

Table 4. *Continued*

PICO	First author (year)	Country	Production system	Age	Intervention	Comparator	Duration of treatment	Study duration [a]	*Mycoplasma* status[b]
6	Schunicht *et al.* (2002)	Canada	Feedlot	7–10 months	Oxytetracycline (20 mg/kg, i.m.) and 30 mg(kg, i.m)	Florfenicol (20 mg/kg, sc).	Long-acting oxytetracycline given once and florfenicol twice	Until slaughter (days not mentioned)	Unknown. No sampling.
7	Lekeux & Art (1988)	Belgium	Feedlot	10 months	Oxytetracycline (10 mg/kg, i.m.)	Enrofloxacin (2.5 mg/kg, i.m)	3	30	Unknown. Not tested.
7	Giles *et al.* (1991)	France	Beef	Not reported	Oxytetracycline (10 mg/kg, i.m.)	Danofloxacin (1.25 mg/kg)	3	5	Unknown. Not tested.
8	Godinho *et al.* (2005)	France, Germany, Italy, Spain	Not reported	3 weeks to 12 months	Florfenicol (40 mg/kg, s.c.)	Tulathromycin (2,5 mg/kg)	One injection	14	Yes
8	Nutsch *et al.* (2005)	USA	Stocker	Not reported	Florfenicol (40 mg/kg, s.c.)	Tulathromycin (2,5 mg/kg)	One injection	60	Unknown. No sampling
8	Schunicht *et al.* (2007)	USA	Feedlot	5–9 months	Florfenicol (40 mg/kg, s.c.)	Tulathromycin (2,5 mg/kg)	One injection	28	Unknown. No sampling.
8	Skogerboe *et al.* (2005)	USA	Feedlot	Not reported (Weight = 315 lb–745 lb)	Florfenicol (40 mg/kg, s.c.)	Tulathromycin (2,5 mg/kg)	One injection	28	No
8	Hoar *et al.* (1998)	Canada	Feedlot	>5 months and < 200 kg	Florfenicol (20 mg/kg, i.m.) given twice with 48 h interval	Tilmicosin (10 mg/kg, s.c.) administered once	Florfenicol was administered twice and tilmicosin once.	28	Yes. In 94 calves (3 died). *M. bovis* was grown from 7 postmortem specimens."
8	Van Donkersgoed *et al.* (2008)	Canada	Feedlot	Not reported (Weight 650–750 lb)	Florfenicol (40 mg/kg, s.c.)	Tulathromycin (2,5 mg/kg, s.c.)	One injection	Until slaughter	Unknown. No sampling.
8	Perrett *et al.* (2008)	USA	Feedlot	Not reported	Florfenicol (40 mg/kg, s.c.)	Tulathromycin (2,5 mg/kg, s.c.)	One injection	Until slaughter	Unknown. No sampling.
8	Jim *et al.* (1999)	Canada	Feedlot	7–10 months	Florfenicol (20 mg/kg, i.m., EOD)	Tilmicosin (10 mg/kg, s.c.)	Florfenicol was administered twice and tilmicosin once.	Until slaughter	No
8	Van Donkersgoed & Hendrick (2013)	Canada	Feedlot	6–10 months	Florfenicol (40 mg/kg, s.c.)	Gamithromycin (6 mg/kg, s.c.)	One injection	Not reported (probably until slaughter)	Unknown. No sampling.

Table 4. *Continued*

PICO	First author (year)	Country	Production system	Age	Intervention	Comparator	Duration of treatment	Study duration [a]	*Mycoplasma* status[b]
9	Fodor *et al.* (2000)	Hungary	Dairy	3 days-2.5 months	Tilmicosin (12.5 mg/kg, p.o.) BID	Enrofloxacin (1.25 mg/kg, p.o.) BID	5	15	Yes. *M. Bovis*
9	Heins *et al.* (2014)	USA	Dairy	Pre-weaned (over 4 days)	Tulathromycin (2.5 mg/kg, s.c.)	Enrofloxacin (12.5 mg/kg, s.c.)	One injection	10	Yes. Detected in 36% of the submitted samples in 3 out of 4 sites.
9	Robb *et al.* (2007)	USA	Feedlot	3–10 months	Tulathromycin (2.5 mg/kg, s.c.)	Enrofloxacin (12.5 mg/kg, s.c.)	One injection	60	No
9	Rowan *et al.* (2004)	France, Ireland, United Kingdom	Rearing facilities	<15 months	Tilmicosin (10 mg/kg, s.c.)	Danofloxacin (6 mg/kg, s.c.)	One injection. One more injection danofloxacin was given after 48 h if clinical signs had not resolved. Tilmicosin was only given once.	11	Unknown. Not tested.
9	Thomas *et al.* (2001)	Belgium, France, United Kingdom	Beef	Not reported	Tilmicosin (10 mg/kg, s.c.) given as a single injection	Marbofloxacin (2 mg/kg, i.v. and s.c.) SID for 4 days	Marbofloxacin is administered for 4 days and Tilmicosin for 1 day.	28	Yes. On 2/9 sites (*M.bovis)*
10	Grandemange *et al.* (2012)	France, Germany, Italy	Not reported	14 to 515 days (mean age = 7.5 months)	Florfenicol (40 mg/kg), s.c.)	Marbofloxacin (10 mg/kg), i.m.)	One injection	21	Yes in 9/231 samples (*M.bovis*)

PICO 7: Two trials comparing tetracycline and fluoroquinolones (TET vs. QUI), including 178 animals (Giles *et al.*, 1991; Lekeux & Art, 1988).

PICO 8: Ten trials compared florfenicol with macrolides (FLO vs. MAC) including 4325 animals (Catry *et al.*, 2008; Godinho *et al.*, 2005; Hoar *et al.*,1998; Jim *et al.*, 1999; Nutsch *et al.*, 2005; Perrett *et al.*, 2008; Schunicht *et al.*, 2007; Skogerboe *et al.*, 2005; Van Donkersgoed *et al.*, 2008; Van Donkersgoed & Hendrick, 2013). Two trials did not report enough detail to pool the results and were instead reported narratively in Supplementary file 3 (Arslan *et al.*, 2010; Merrill & Tonkinson, 1989).

PICO 9: Five trials compared macrolides with fluoroquinolones (MAC vs. QUI) in 1654 animals (Fodor *et al.*, 2000; Heins *et al.*, 2014; Robb *et al.*, 2007; Rowan *et al.*, 2004; Thomas *et al.*, 2001).

PICO 10: One trial compared florfenicol with fluoroquinolones (FLO vs. QUI) in 207 animals (Grandemange *et al.*, 2012).

Treatment efficacy

No studies have shown a clinically relevant treatment effect, defined as 200 or more treatment failures per 1000 animals, in favor of either the intervention or comparator for any of the PICOs. In some pairwise comparisons, specifically, PICO 1 (PNC vs. TET), PICO 3 (PNC vs. FLO), PICO 6 (TET vs. FLO), PICO 9 (MAC vs. QUI), and PICO 10 (FLO vs. QUI), less critical antimicrobial classes even resulted in fewer treatment failures, although the differences were not clinically significant, when compared with high-priority critically important antimicrobials for human medicine. The absolute treatment effects, statistical results, and certainty of evidence for all the PICOs are presented in Table 5. The results of individual studies as well as pooled results, risk of bias assessments for all domains, and heterogeneity are visualized in forest plots in Figure 2– Figure 9. The full risk of bias assessment is reported in Supplementary File 5.

Table 5. Efficacy results for all pooled PICOs shown in absolute numbers, p-values, risk ratio with 95% confidence intervals (CI), and certainty of evidence.

PICO	Absolute effect	*p*-value	Risk ratio (95% CI)	Certainty of evidence
1	Penicillin treatment resulted in 8 fewer therapeutic failures per 1000 animals (from 65 fewer to 67 more) in comparison to tetracycline treatment	0.83	0.97 (0.76–1.25)	Very low
3	Penicillin treatment resulted in 38 fewer therapeutic failures per 1000 animals (from 93 fewer to 194 more) in comparison to florfenicol treatment	0.57	0.66 (0.16–2.75)	Low
5	Tetracycline treatment resulted in 132 more therapeutic failures per 1000 animals (from 38 fewer to 415 more) in comparison to macrolide treatment	0.15	1.45 (0.87–2.41)	Very low
6	Tetracycline treatment resulted in 7 fewer therapeutic failures per 1000 animals (from 53 fewer to 53 more) in comparison to florfenicol treatment	0.79	0.96 (0.70–1.30)	Low
7	Tetracycline treatment resulted in 179 more therapeutic failures per 1000 animals (from 60 more to 343 more) in comparison to fluoroquinolone treatment	0.002	1.45 (1.15–1.84)	Very low
8	Florfenicol treatment resulted in 27 more therapeutic failures per 1000 animals (from 16 fewer to 75 more) in comparison to macrolide treatment	0.25	1.10 (0.94 to 1.28)	Moderate
9	Macrolide treatment resulted in 48 fewer therapeutic failures per 1000 animals (from 106 fewer to 37 more) in comparison to fluoroquinolone treatment	0.22	0.79 (0.54–1.16)	Moderate
10	Florfenicol treatment resulted in 102 fewer therapeutic failures per 1000 animals (from 165 fewer to 10 more) in comparison to fluoroquinolone treatment	0.07	0.58 (0.32–1.04)	Low

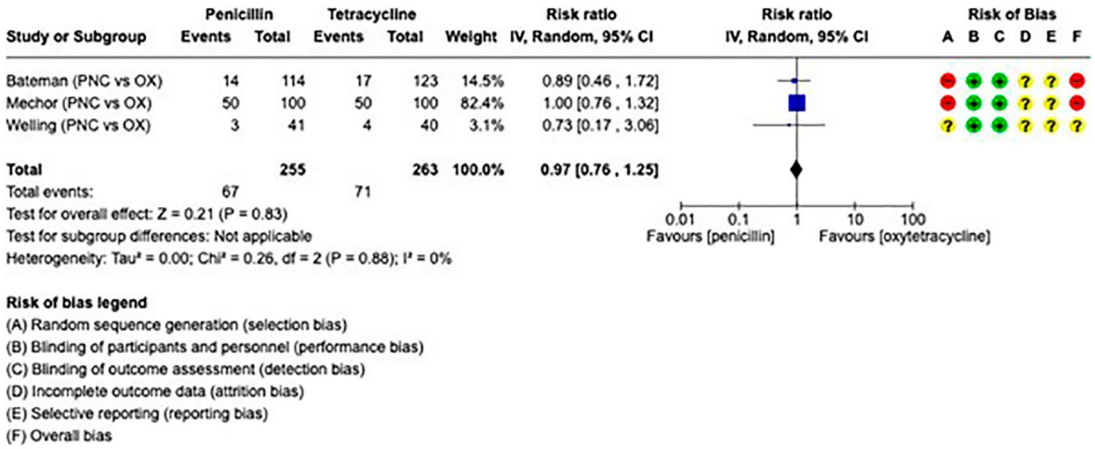

Figure 2. Forest plot for PICO 1 comparing the efficacy of penicillins vs tetracyclines. The number of events represents the need for re-treatment. PNC = penicillin, OX = Oxytetracycline. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

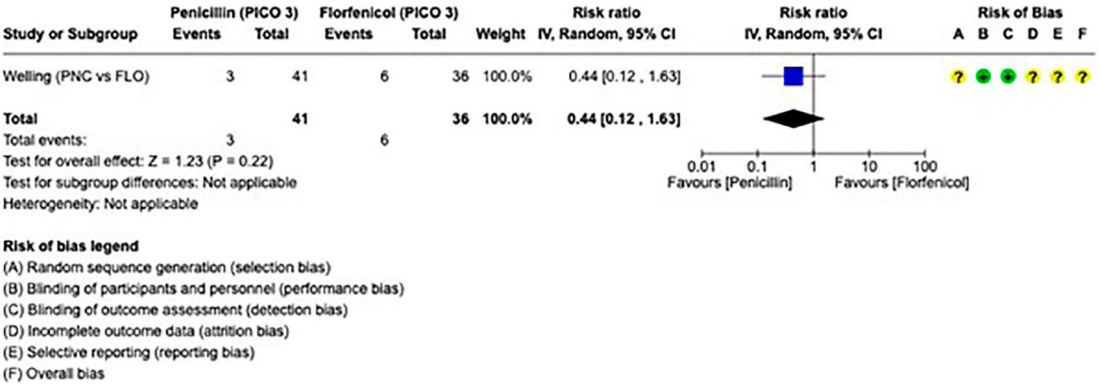

Figure 3. Forest plot for PICO 3 comparing the efficacy of penicillins vs florfenicol. The number of events represents the need for re-treatment. PNC = penicillin, FLO = florfenicol. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

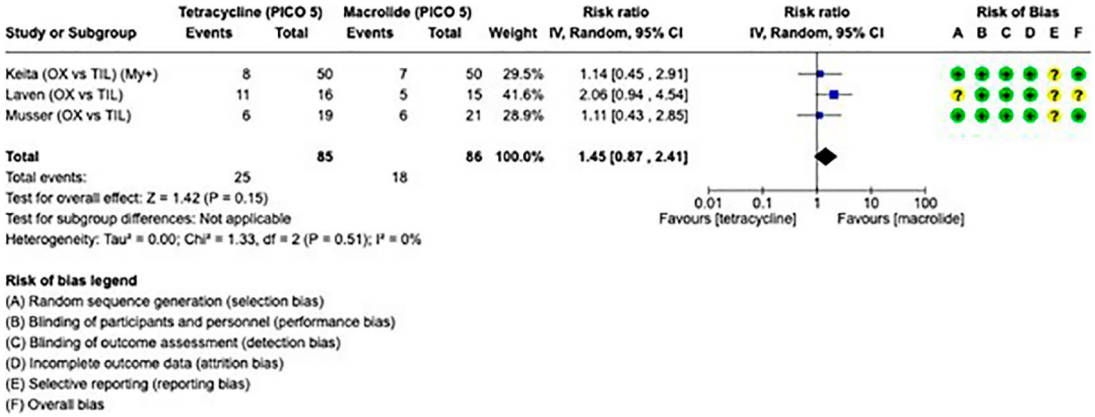

Figure 4. Forest plot for PICO 5 comparing the efficacy of tetracyclines vs macrolides. The number of events represents the need for re-treatment. OX = oxytetracycline, TIL = tilmicosin, My+ = *Mycoplasma*-positive animals included. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

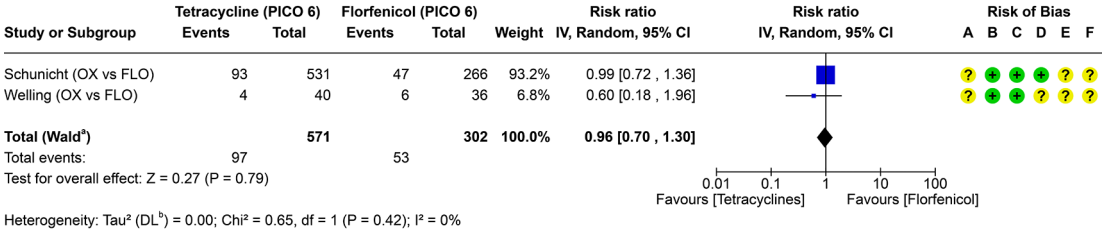

Figure 5. Forest plot for PICO 6 comparing the efficacy of tetracyclines vs florfenicol. The number of events represents the need for re-treatment. OX = oxytetracycline, FLO = florfenicol. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

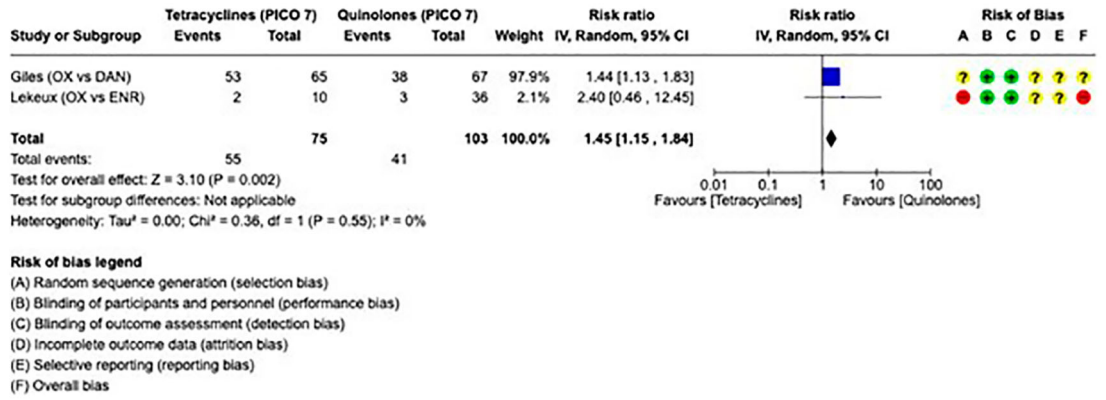

Figure 6. Forest plot for PICO 7 comparing the efficacy of tetracyclines vs quinolones. The number of events represents the need for re-treatment. OX = oxytetracycline, DAN = danofloxacin, ENR = enrofloxacin. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

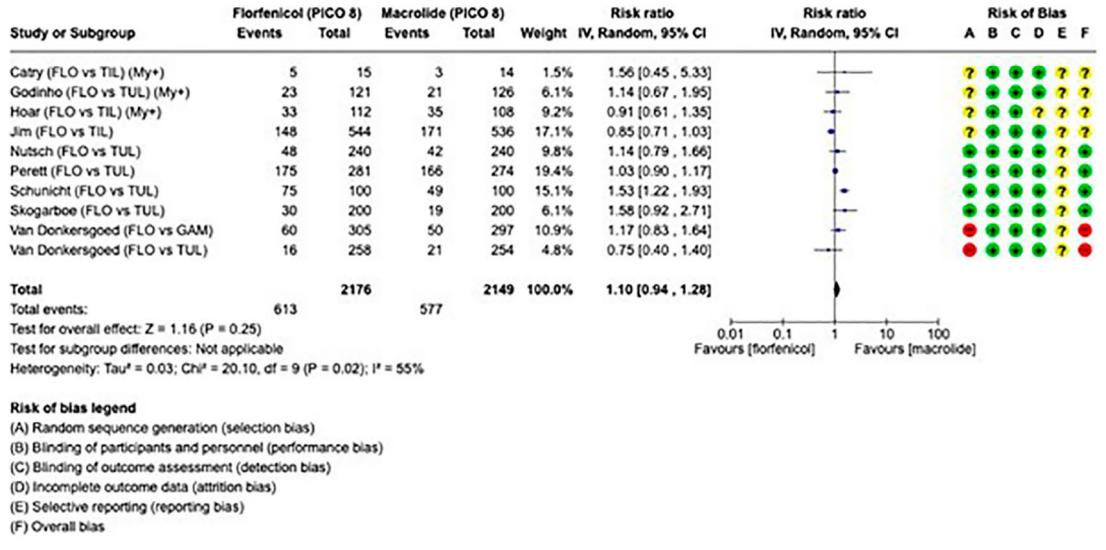

Figure 7. Forest plot for PICO 8 comparing the efficacy of florfenicol vs macrolides. The number of events represents the need for re-treatment. FLO = florfenicol, TIL = tilmicosin, TUL = tulathromycin, GAM = Gamithromycin, My+ = *Mycoplasma*-positive animals included. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

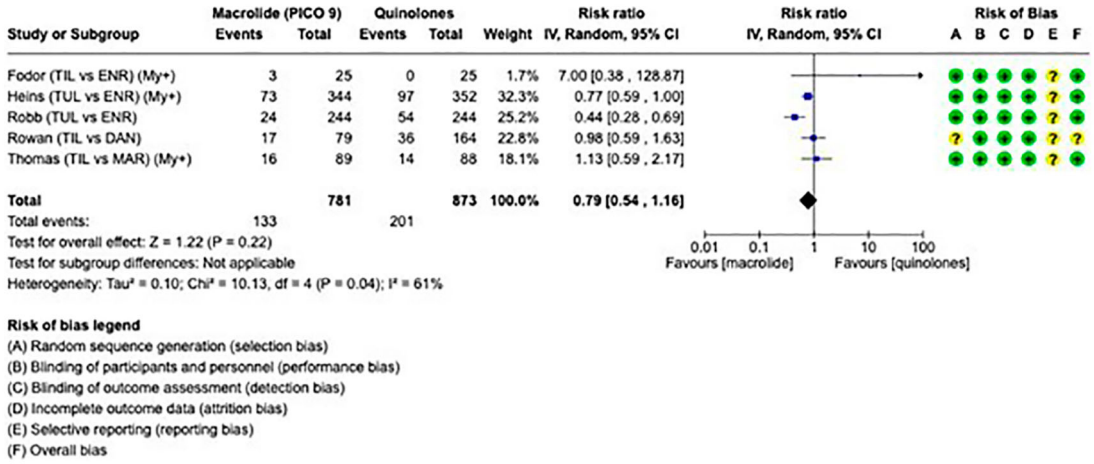

Figure 8. Forest plot for PICO 9 comparing the efficacy of macrolides vs quinolones. The number of events represents the need for re-treatment. TIL = tilmicosin, TUL = tulathromycin, DAN = danofloxacin, ENR = enrofloxacin, MAR = marbofloxacin, My+ = *Mycoplasma*-positive animals included. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

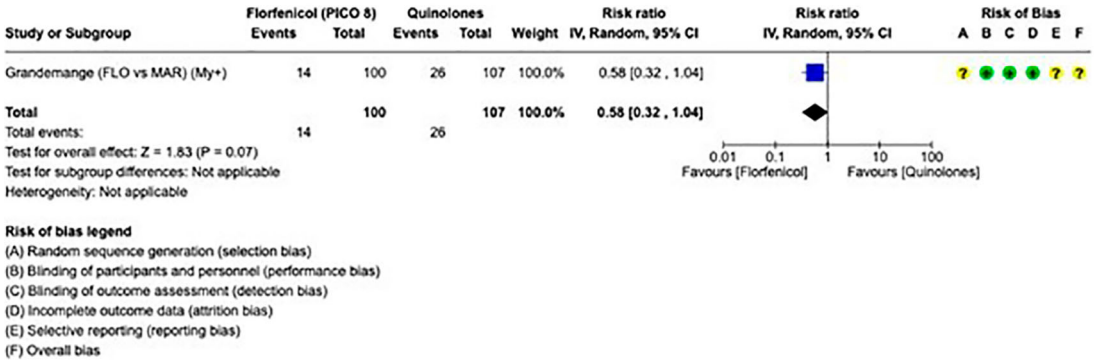

Figure 9. Forest plot for PICO 10 comparing the efficacy of florfenicol vs quinolones. The number of events represents the need for re-treatment. FLO = florfenicol, MAR = marbofloxacin, My+ = *Mycoplasma*-positive animals included. In the risk of bias assessment a red "traffic light" mean high risk of bias, yellow means that there are some concerns and green that the risk of bias is low.

Certainty of evidence

Few trials reported short-term therapeutic failure (<7 days), and the need for re-treatment was used as a surrogate outcome, irrespective of the observational period (range 3–60 days). Therefore, all pairwise comparisons were downgraded one level for indirectness. As previously mentioned, other critical outcomes, such as relapse, mortality, and weight gain could not be pooled or reliably assessed, which also contributed to the decision to downgrade all PICOs for indirectness. Only PICO 5 (TET vs. MAC) and PICO 7 (TET vs. QUI) were downgraded for imprecision (three thresholds were crossed for both). In addition to the indirectness, the main reason for downgrading was the risk of bias. PICO 8 (FLO vs. MAC) and PICO 9 (MAC vs. QUI) showed moderate evidence and were downgraded only for indirectness for the reasons mentioned above. The certainty assessment for all domains is reported in Table 6, and a summary of findings tables can be seen in Supplementary File 4.

Mycoplasma-positive subgroup analysis

Only trials reporting *Mycoplasma*-positive animals (species-level *M. bovis* when available) were included in this analysis. This information is available for PICO 5 (TET vs. MAC), PICO 8 (FLO vs. MAC), and PICO 9 (MAC vs. QUI). In this small and imprecise subgroup, point estimates were broadly consistent with those of the main analysis. Overall, these data are insufficient to conclude similar responses. Because many included trials did not assess *Mycoplasma* status, some *Mycoplasma*-positive animals were likely present in the main analysis. The effect sizes of the sub-analysis are shown in Table 7, and the forest plots are shown in Supplementary File 6. The Instrument for

Table 6. Certainty assessment for all domains and the overall certainty for the therapeutic failure for PICO 1–10.

PICO	Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	Overall certainty
1	Very serious	Serious	Not serious	Not serious	Undetected	Very low
3	Serious	Serious	Not serious	Not serious	Undetected	Low
5	Serious	Serious	Not serious	Extremely serious	Undetected	Very low
6	Serious	Serious	Not serious	Not serious	Undetected	Low
7	Very serious	Serious	Not serious	Extremely serious	Undetected	Very low
8	Not serious	Serious	Not serious	Not serious	Undetected	Moderate
9	Not serious	Serious	Not serious	Not serious	Undetected	Moderate
10	Serious	Serious	Not serious	Not serious	Undetected	Low

Table 7. Efficacy results for the *Mycoplasma* sub-analysis shown in absolute numbers, p-values, risk ratio with 95% confidence intervals (CI). Only trials with known *Mycoplasma*-positive animals were included in the analysis.

PICO	Absolute effect	*p*-value	Risk ratio (95% CI)
5	Tetracycline treatment resulted in 20 more therapeutic failures per 1000 animals (from 77 fewer to 267 more) in comparison to macrolide treatment	0.78	1.14 (0.45–2.91)
8	Florfenicol treatment resulted in 2 more therapeutic failures per 1000 animals (from 89 fewer to 150 more) in comparison to macrolide treatment	0.93	1.01 (0.64–1.61)
9	Macrolide treatment resulted in 19 fewer therapeutic failures per 1000 animals (from 103 fewer to 119 more) in comparison to fluoroquinolone treatment	0.75	0.92 (0.57–1.50)

Table 8. Efficacy results for the productions system sub-analysis shown in absolute numbers, p-values, risk ratio with 95% confidence intervals (CI). Non-feedlot productions systems include dairy, slaughter, stocker, beef and veal production systems.

PICO	Production system	Absolute effect	*p*-value	Risk ratio (95% CI)
1	Non-feedlot	Penicillin treatment resulted in 27 fewer therapeutic failures per 1000 animals (from 83 fewer to 206 more) in comparison to tetracycline treatment	0.67	0.73 (0.17–3.06)
1	Feedlot	Penicillin treatment resulted in 6 fewer therapeutic failures per 1000 animals (from 72 fewer to 81 more) in comparison to tetracycline treatment	0.89	0.98 (0.76–1.27)
6	Non-feedlot	Tetracycline treatment resulted in 67 fewer therapeutic failures per 1000 animals (from 137 fewer to 160 more) in comparison to florfenicol treatment	0.40	0.60 (0.18–1.96)
6	Feedlot	Tetracycline treatment resulted in 2 fewer therapeutic failures per 1000 animals (from 49 fewer to 64 more) in comparison to florfenicol treatment	0.96	0.99 (0.72–1.36)
7	Non-feedlot	Tetracycline treatment resulted in 250 more therapeutic failures per 1000 animals (from 74 more to 471 more) in comparison to fluoroquinolone treatment	0.003	1.44 (1.13–1.83)

Table 8. *Continued*

PICO	Production system	Absolute effect	*p*-value	Risk ratio (95% CI)
7	Feedlot	Tetracycline treatment resulted in 117 more therapeutic failures per 1000 animals (from 45 fewer to 954 more) in comparison to fluoroquinolone treatment	0.3	2.4 (0.46–12.45)
8	Non-feedlot	Florfenicol treatment resulted in 28 more therapeutic failures per 1000 animals (from 12 fewer to 78 more) in comparison to macrolide treatment	0.1	1.16 (0.93 to 1.45)
8	Feedlot	Florfenicol treatment resulted in 23 more therapeutic failures per 1000 animals (from 43 fewer to 110 more) in comparison to macrolide treatment	0.45	1.08(0.85–1.38)
9	Non-feedlot	Macrolide treatment resulted in 28 fewer therapeutic failures per 1000 animals (from 77 fewer to 40 more) in comparison to fluoroquinolone treatment	0.38	0.88 (0.67–1.17)
9	Feedlot	Macrolide treatment resulted in 55 fewer therapeutic failures per 1000 animals (from 71 fewer to 30 fewer) in comparison to fluoroquinolone treatment	0.0004	0.44 (0.28–0.69)

assessing the Credibility of Effect Modification Analysis (ICEMAN tool) was not used in the subanalysis because a clinically meaningful effect modification was not found.

Production system sub-group analysis

Feedlot production systems were separated from all other production systems in this subanalysis. Non-feedlot production systems included dairy, veal, beef, slaughter, and stocker production systems (Table 4). The point estimate of effects (Table 8) did not change in the sub-analysis, indicating that no intervention proved superior to a comparator in relation to the clinical thresholds when feedlot and non-feedlot trials were pooled separately. Furthermore, the point estimates did not cross the null value, even within the trivial effect range, suggesting that feedlot animals have a similar clinical response to treatment as non-feedlot animals for the included antimicrobials. In PICO 6 (TET vs. FLO), the point estimate for feedlot animals was closer to the null (two fewer therapeutic failures per 1000 animals in favor of TET) compared to non-feedlot animals (67 fewer therapeutic failures per 1000 animals in favor of TET), suggesting a smaller difference in treatment effect between TET and FLO in feedlot animals (Table 8). Conversely, in PICO 9 (MAC vs. QUI), the point estimate in feedlot animals was further away from the null (55 fewer therapeutic failures in 1000 animals) than in non-feedlot animals (28 fewer therapeutic failures in 1000 animals), but this was only based on one trial. Moreover, the p-value was below 0.05 ($p = 0.0004$) in this comparison, suggesting that results are less likely to be due to random error alone. The ICEMAN tool was not used in this sub-analysis because no clinically meaningful effect modification was found. Forest plots for both feedlot and non-feedlot trials are shown in Supplementary File 6.

Discussion

We did not detect clinically meaningful differences between antimicrobial classes in reducing the need for re-treatment (a surrogate for short-term therapeutic failure). Certainty of evidence ranged from very low to moderate, so these estimates should be interpreted cautiously. In GRADE terms, low certainty means the true effect might be markedly different from the estimate and very low means it is probably markedly different. Our findings therefore do not imply equivalence; rather, within the limits of the available trials and our prespecified clinical thresholds, no class demonstrated a short-term advantage.

This means that other considerations, such as the harm of AMR and its implications for public health, could weigh heavier in the evidence to decision framework and final treatment recommendations. The World Health Organization (WHO) guidelines on the use of medically important antimicrobials in food-producing animals and the European Medicines Agency's (EMA) classification of antimicrobial use in animals will be instrumental in balancing benefits and harms (EMA, 2025; WHO, 2024).

Different levels of certainty were determined for pairwise antimicrobial comparisons. Three comparisons, PICO 1 (PNC vs. TET), PICO 5 (TET vs. MAC), and PICO 7 (TET vs. QUI), were of **very low certainty** which means that the *true effect probably is markedly different from the estimated effect.* There were, however, different reasons for downgrading

the evidence, and the GRADE methodology does not distinguish between different levels in the very low spectrum. PICO 7 was downgraded for a very serious risk of bias and extremely serious imprecision, whereas PICO 1 showed no imprecision and bias was not very serious in PICO 5. PICO 7 could thus be considered to be of even lower evidence in comparison to the other PICOs, even though this is not reflected in the final certainty result.

Three comparisons were of **low certainty** evidence, which means that the *true effect could be markedly different from the estimated effect.* PICO 3 (PNC vs. FLO), PICO 6 (TET vs. FLO), and PICO 10 (FLO vs. QUI) showed precise results (within the trivial effect threshold), but were downgraded for risk of bias. Most of the trials that were downgraded for risk of bias were rated this way because no comparisons with studies with low risk of bias could be made. When there is concordance between high risk of bias studies and low risk of bias studies (meaning that all trials fall within the same threshold range), no rating down is required even if there are several high risk of bias trials included (Neumann *et al.*, 2025). The decision to rate down one level and in some cases up to two levels, for risk of bias in the included comparisons, could be considered an overly cautious judgement. However, since the requirement to report allocation concealment was excluded from the modified RoB 2 tool that was used to assess the risk of bias in the present study, and there were still "some concerns" (visualized with yellow traffic lights) in the randomization domain, a more conservative decision was made in the end.

Two comparisons, PICO 8 (FLO vs. MAC) and PICO 9 (MAC vs. QUI), were of **moderate certainty evidence** which means that the *true effect is probably close to the estimated effect.* This provides evidence that can lead to strong recommendations. The only reason for rating down the certainty of evidence was indirectness of the outcome. The panel had lengthy discussions on the definition and selection of outcomes, and ultimately, it was not possible to pool all critical outcomes. Mortality and relapse of clinical signs were generally measured over a long observational period, during which the animals often received a second and sometimes a third treatment with antimicrobials. This makes the assessment of the efficacy of the initial antimicrobial treatment unsuitable. Most trials measured the need for retreatment, and the first retreatment event (before any other antimicrobials were used) was consequently pooled, similar to previous evidence synthesis in BRD in feedlots (O'Connor *et al.*, 2016). The synthesized evidence, however, does not fully reflect the outcomes generated by the panel and stakeholders; in some trials, the need for re-treatment means that the animals require more than one antimicrobial injection during the first days of treatment, and in some trials, the need for re-treatment after several weeks, which is why all PICOs were downgraded one level for indirectness of outcomes. PICO 8 and PICO 9 could have otherwise been of high certainty, which emphasizes the importance of trialists choosing clinical outcomes that are relevant to end-users. The critical outcomes of short-term clinical cure, relapse of clinical signs after clinical resolution, and mortality can be used in the design of future trials. This study showed that there is discordance in the selection of outcomes, and ideally, a larger stakeholder group should be included in a core outcome set study using the standardized COS-STAD methodology (Kirkham *et al.*, 2017).

Although trials varied in age, production system, dose/route, follow-up window, and BRD case definitions, this diversity did not translate into clinically important inconsistency in comparative effects, which suggest limited effect modification of these parameters. We judged inconsistency (heterogeneity) using the threshold approach which means that study-level and pooled relative effects were converted to absolute risk differences per 1000 animals from a common baseline risk and examined against our a priori clinical thresholds (trivial, small, moderate, large). In PICO 8 and 9, I^2 values were around 50–60%, driven mainly by small trials with few events and wide confidence intervals rather than by opposing effects. When converted to absolute effects and thresholds, estimates were within the trivial-to-small ranges and we did not identify outliers suggesting inconsistency. As an example, one trial (Schunicht *et al.*, 2007) showed a larger effect in comparison to other trials in PICO 8 and reported higher event rates in both arms (75% vs 49%), resulting in a larger RR (1.53; 95% CI 1.22–1.93) than the other trials. Excluding this study reduced I^2 from 55% to 17% and the pooled effect to RR 1.01 (95% CI 0.90–1.12), corresponding to 3 more per 1000 (from 26 fewer to 31 more) and conclusions were unchanged. Subgroup analyses (feedlot vs non-feedlot; *Mycoplasma*-positive) showed no credible effect modification, acknowledging limited precision and few studies per comparison. We therefore did not rate down for inconsistency, but we did rate down for indirectness due to variable follow-up windows and reliance on retreatment as a surrogate for short-term failure. Disease pressure varied across trials, with some reporting high first-treatment success (e.g., ≥85–90%) and others much higher re-treatment rates; nevertheless, randomization ensures like-with-like comparisons within trials and, across trials, class-level effects still clustered within the trivial-to-small ranges against our prespecified clinical thresholds. Had baseline risk or disease severity materially modified relative effects across settings, we would expect to see clinically important inconsistency across studies (i.e., study estimates crossing into different threshold ranges); however, when judged against our prespecified clinical thresholds, study-level and pooled estimates clustered within the trivial-to-small ranges. The small number of trials in some PICOs limits power to detect effect modification, so we interpret these findings cautiously.

Sackett *et al.* (1996) emphasized that evidence-based medicine, and in extension evidence-based guidelines, integrate the best available evidence with the experience of the veterinary clinician (i.e., experience of the veterinary practitioner in this case) and values of the "patient" (i.e., values of cattle farmers in the present study), which is reflected in the clinical thresholds. Interestingly, the *a priori* clinical thresholds established by the panel and stakeholders were very similar to the overall baseline therapeutic failure across the included trials. This could indicate that the results are generalizable to field conditions, since they seem to correlate with the experiences of clinicians and farmers, which strengthens the results. It should be noted that clinical thresholds are subjective and likely to vary between countries and individuals. The choice of stakeholders to survey was also based on a convenience sample of country-representative clinicians from Europe and North America and is therefore open to bias. However, the interpretation of effect size, confidence intervals, and level of certainty is not meant to be an objective statement of right and wrong, but primarily aims to be transparent. Moreover, evidence interpretation is not supposed to be static but contextualized to the intended setting, which means that different guideline developers can arrive at different conclusions and recommendations based on the same evidence (Kahale *et al.*, 2021; Martins *et al.*, 2024).

A previous network meta-analysis (NMA) of antimicrobial treatment of clinical BRD included feedlot trials from North America (O'Connor *et al.*, 2016) whereas the present trial included trials from all production systems and countries; therefore, these analyses are not directly comparable. Another difference was that antimicrobial substances were compared in the NMA instead of comparing the antimicrobial classes included here. There are many advantages of indirect NMA comparisons: they can facilitate treatment efficacy comparisons from different trials when pairwise comparisons are lacking, increase precision, and rank a network of multiple treatments. However, ranking forest plots are based on relative effects rather than absolute effects, and clinical relevance and certainty of evidence have not been assessed previously. Oxytetracycline (OX) ranked the lowest among all included substances in the NMA, and the effect was comparable to that of the negative control. In the present study, there were more OX treatment failures in pairwise comparisons with QUI (enrofloxacin and danofloxacine) and MAC (tilmicosin), but OX did not show an inferior clinical effect compared to either class. However, the certainty of evidence was classified as very low or low, mainly due to imprecision, which means that the results could have been different if larger trials were available. FLO also ranked much higher in efficacy than OX in the NMA, but in our direct comparison, there was no clinically meaningful difference between OX and FLO. However, the certainty of evidence was low, meaning that the true effect could be markedly different from the estimated effect. The clinical experience of practicing bovine clinicians in the panel supports that OX mostly results in a satisfactory clinical response in the treatment of clinical BRD, not only in Scandinavia but also in central Europe, Turkey, Canadian dairy farms, and cow-calf operations.

A limitation of this study was that only peer-reviewed studies were included, and no grey literature was included. An attempt was made to retrieve unpublished technical reports that were included in O'Connor *et al.* (2016), but the Internet links were broken and the pharmaceutical companies (today owned by other companies) did not respond to requests, which could explain some differences in the results, besides production systems and geographical location. We restricted inclusion to English-language, peer-reviewed RCTs. Although we screened reference lists to mitigate missed studies, language bias cannot be excluded and may limit applicability in non-English contexts; this consideration is addressed in the companion guideline's Evidence-to-Decision framework. Another limitation of the present study is that clinical BRD was used as the relevant case selection criterion, and different clinical definitions and scoring systems were used in the studies, which could have introduced some bias. In addition, in recent years, the use of ultrasonography to measure lung consolidation has indicated that this could be a more reliable method to identify animals with a respiratory tract infection in need of antimicrobial therapy (Buczinski *et al.*, 2025). Studies have shown that the prevalence of subclinical pneumonia can be higher than that of clinical BRD (Jourquin *et al.*, 2023). Moreover, calves with clinical signs of BRD have higher odds of being classified as having severe pneumonia with lung consolidation $\geq$3 cm in depth and lower odds of being cured (Jourquin *et al.*, 2025). Calves detected in the subclinical phase, and hence with mild or moderate pneumonia, have better odds for cure with shorter therapy lengths (Jourquin *et al.*, 2025).

For decades, practitioners and researchers have assumed that BRD treatment efficacy increased with every novel antimicrobial formulation, based on systematic comparisons with the previous molecules marketed, and often based on non-peer-reviewed or published RCTs. The present meta-analysis suggests that differences in therapeutic efficacy between antimicrobials for BRD may be limited with different levels of certainty. This finding supports the concept that older molecules with less importance in human medicine should be re-evaluated in new trials. Unfortunately, comparative studies including PEN, the currently recommended first-choice antimicrobial for BRD in Scandinavia, where *M. bovis* prevalence is low to absent, are scarce, and the included studies outside Scandinavia date back nearly 40 years. The apparent absence of difference in efficacy may signify that AMR levels for first-intention cases are generally low in the bacterial pathogens associated with BRD, with the exception of moderate levels of resistance to specific antimicrobials in *Mannheimia haemolytica* and, to a lesser extent, *Pasteurella multocida* (EFSA Panel on Animal Health and

Welfare (AHAW) *et al.*, 2021). Another explanation for the absence of differences in efficacy between antimicrobials identified in this meta-analysis may be that the relatively small number of studies, with relatively small numbers of animals, for most of the PICOs, provided low power to detect differences. The overall importance of viral infection and the ability of the host to self-cure upon administration of bacteriostatic antimicrobials should not be underestimated.

Conclusion

In this systematic review and meta-analysis of primary antimicrobial therapy for clinical BRD (excluding metaphylaxis), we did not detect clinically meaningful differences between the five antimicrobial classes in need for re-treatment, based on prespecified clinical thresholds. Certainty of evidence ranged from very low to moderate, so results should be interpreted with caution. Important outcomes such as mortality and growth were reported too inconsistently to pool. Future trials would benefit from a core outcome set (including short-term clinical cure, relapse after resolution, and mortality) with harmonized case definitions and follow-up windows. These findings will inform ENOVAT's BRD antimicrobial treatment guideline, where broader considerations (feasibility, stewardship, and applicability across systems) are addressed in the Evidence-to-Decision framework.

Ethics and consent

Ethical approval and consent were not required.

Data availability

The supplementary files including the PRISMA checklist and protocol deviations, data extraction forms/templates, risk-of-bias forms and judgements, GRADE evidence profiles and summary-of-findings tables and stakeholder materials (surveys, clinical thresholds) are openly available at Open Science Framework: https://doi.org/10.17605/OSF.IO/KTHG2 (Scahill, 2025).

This project contains the following data:

- Supplementary file 1.
- Supplementary file 2.
- Supplementary file 3.
- Supplementary file 4.
- Supplementary file 5.
- Supplementary file 6.

Reporting guidelines

OSF: PRISMA checklist for "Efficacy of antimicrobial treatment for bovine respiratory disease: a systematic review and meta-analysis for European Network for Optimization of Veterinary Antimicrobial Therapy guidelines". https://doi.org/10.17605/OSF.IO/KTHG2 (Scahill, 2025).

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements

We would like to thank Brian Friker for developing the search strategy and for conducting the first search, and all stakeholders (farmers and field clinicians) for contributing to the establishment of clinical thresholds.

References

Arslan HH, Yavuz O, Nisbet C, *et al.*: **Comparative evaluation of the effects of florfenicol and tulathromycin on clinical recovery and acute phase proteins in undifferentiated natural bovine respiratory disease.** *Rev Med Vet.* 2010; **161**(12): 535–539.
Reference Source

Bateman KG, Martin SW, Shewen PE, *et al.*: **An evaluation of antimicrobial therapy for undifferentiated bovine respiratory disease.** *Can Vet J.* 1990; **31**(10): 689–696.
PubMed Abstract | **Free Full Text**

Bell BG, Schellevis F, Stobberingh E, *et al.*: **A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance.** *BMC Infect Dis.* 2014; **14**: 13.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Blakebrough-Hall C, McMeniman JP, González LA: **An evaluation of the economic effects of bovine respiratory disease on animal performance, carcass traits, and economic outcomes in feedlot cattle defined using four BRD diagnosis methods.** *J Anim Sci.* 2020; **98**(2): skaa005.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Buczinski S, Ollivett TL, Pardon B: ***Invited review*: Lung ultrasonography—improving our understanding and management of respiratory disease in young calves.** *J Dairy Sci.* 2025; **108**(12): 12903–12923.
PubMed Abstract | **Publisher Full Text**

Cassini A, Högberg LD, Plachouras D, *et al.*: **Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis.** *Lancet Infect Dis.* 2019; **19**(1): 56–66.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Catry B, Duchateau L, Van de Ven J, *et al.*: **Efficacy of metaphylactic florfenicol therapy during natural outbreaks of bovine respiratory disease.** *J Vet Pharmacol Ther.* 2008; **31**(5): 479–487.
PubMed Abstract | **Publisher Full Text**

Donlon JD, Mee JF, McAloon CG: **Prevalence of respiratory disease in Irish preweaned dairy calves using hierarchical Bayesian latent class analysis.** *Front Vet Sci.* 2023; **10**: 1149929.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Duff GC, Galyean ML: **Board-invited review: recent advances in management of highly stressed, newly received feedlot cattle.** *J Anim Sci.* 2007; **85**(3): 823–840.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

EFSA Panel on Animal Health and Welfare (AHAW), ; Nielsen SS, Bicout DJ, *et al.*: **Assessment of animal diseases caused by bacteria resistant to antimicrobials: cattle.** *EFSA J.* 2021; **19**(12): e06955.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

European Medicines Agency: **Categorisation of antibiotics in the European Union.** 2025; [cited 2025 Nov 21].
Reference Source

Fodor L, Reeve-Johnson L, Hodge A, *et al.*: **Efficacy evaluations of the use of oral tilmicosin in pneumonic calves.** *Vet J.* 2000; **159**(2): 194–200.
PubMed Abstract | **Publisher Full Text**

Giles CJ, Grimshaw WT, Shanks DJ, *et al.*: **Efficacy of danofloxacin in the therapy of acute bacterial pneumonia in housed beef cattle.** *Vet Rec.* 1991; **128**(13): 296–300.
PubMed Abstract | **Publisher Full Text**

Godinho KS, Wolf RMLG, Sherington J, *et al.*: **Efficacy of tulathromycin in the treatment and prevention of natural outbreaks of bovine respiratory disease in European cattle.** *Vet Ther.* 2005; **6**(2): 122–135.
PubMed Abstract

Grandemange E, Fournel S, Giboin H, *et al.*: **Efficacy of a single injection of marbofloxacin in the treatment of bovine respiratory disease.** *Rev Med Vet.* 2012; **163**(6): 287–294.
Reference Source

Guardabassi L, Apley M, Olsen JE, *et al.*: **Optimization of antimicrobial treatment to minimize resistance selection.** *Microbiol Spectr.* 2018; **6**(3): 29932044.
PubMed Abstract | **Publisher Full Text**

Guyatt G, Agoritsas T, Brignardello-Petersen R, *et al.*: **Core GRADE 1: overview of the Core GRADE approach.** *BMJ.* 2025; **389**: e081903.
PubMed Abstract | **Publisher Full Text**

Guyatt G, Zhao Y, Mayer M, *et al.*: **GRADE guidance 36: updates to GRADE's approach to addressing inconsistency.** *J Clin Epidemiol.* 2023; **158**: 70–83.
PubMed Abstract | **Publisher Full Text**

Heins BD, Nydam DV, Woolums AR, *et al.*: **Comparative efficacy of enrofloxacin and tulathromycin for treatment of preweaning respiratory disease in dairy heifers.** *J Dairy Sci.* 2014; **97**(1): 372–382.
PubMed Abstract | **Publisher Full Text**

Hoar BR, Jelinski MD, Ribble CS, *et al.*: **A comparison of the clinical field efficacy and safety of florfenicol and tilmicosin for the treatment of undifferentiated bovine respiratory disease of cattle in western Canada.** *Can Vet J.* 1998; **39**(3): 161–166.
PubMed Abstract | **Free Full Text**

Jim GK, Booker CW, Guichon PT, *et al.*: **A comparison of florfenicol and tilmicosin for the treatment of undifferentiated fever in feedlot calves in western Canada.** *Can Vet J.* 1999; **40**(3): 179–184.
PubMed Abstract

Jourquin S, Bokma J, De Cremer L, *et al.*: **Randomized field trial comparing the efficacy of florfenicol and oxytetracycline in a natural outbreak of calf pneumonia using lung reaeration as a cure criterion.** *J Vet Intern Med.* 2022; **36**(2): 820–828.
PubMed Abstract | **Publisher Full Text**

Jourquin S, Lowie T, Debruyne F, *et al.*: **Dynamics of subclinical pneumonia in male dairy calves in relation to antimicrobial therapy and production outcomes.** *J Dairy Sci.* 2023; **106**(1): 676–689.
PubMed Abstract | **Publisher Full Text**

Jourquin S, Lowie T, Debruyne F, *et al.*: **Development of a classification system for lung ultrasonographic findings in calves based on the association with cure and production outcomes.** *J Dairy Sci.* 2025; **108**(7): 7452–7471.
PubMed Abstract | **Publisher Full Text**

Keita A, Pommier P, Pagot E, *et al.*: **A combination oxytetracycline/flunixin treatment of respiratory infections in cattle.** *Rev Med Vet.* 2007; **158**: 86–91.
Reference Source

Kahale LA, Ouertatani H, Ben Brahem A, *et al.*: **Contextual differences considered in the Tunisian ADOLOPMENT of the European guidelines on breast cancer screening.** *Health Res Policy Syst.* 2021; **19**(1): 80.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kirkham JJ, Davis K, Altman DG, *et al.*: **Core Outcome Set-STAndards for Development: the COS-STAD recommendations.** *PLoS Med.* 2017; **14**(11): e1002447.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Laven R, Andrews AH: **Long-acting antibiotic formulations in the treatment of calf pneumonia: a comparative study of tilmicosin and oxytetracycline.** *Vet Rec.* 1991; **129**(6): 109–111.
PubMed Abstract | **Publisher Full Text**

Lekeux P, Art T: **Effect of enrofloxacin therapy on shipping fever pneumonia in feedlot cattle.** *Vet Rec.* 1988; **123**(8): 205–207.
Publisher Full Text

Martins RS, Nadeem S, Aziz A, *et al.*: **GRADE-ADOLOPMENT of hyperthyroidism treatment guidelines for a Pakistani context.** *BMC Endocr Disord.* 2024; **24**(1): 41.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Mechor GD, Jim GK, Janzen ED: **Comparison of penicillin, oxytetracycline, and trimethoprim-sulfadoxine in the treatment of acute undifferentiated bovine respiratory disease.** *Can Vet J.* 1988; **29**(5): 438–443.
PubMed Abstract | **Free Full Text**

Merrill JK, Tonkinson LV: **The effectiveness of Micotil for the treatment of bovine respiratory disease.** *Bovine Pract.* 1989; **24**: 26–28.
Publisher Full Text

Morel C: **Transmission of antimicrobial resistance from livestock agriculture to humans and from humans to animals.** *OECD Food, Agriculture and Fisheries Papers.* 2019; **133**
Publisher Full Text

Moura C, Totton S, Sargeant J, *et al.*: **Evidence of improved reporting of swine vaccination trials in the post-REFLECT statement publication period.** *J Swine Health Prod.* 2019; **27**: 265–277.
Publisher Full Text

Musser J, Mechor GD, Gröhn YT, *et al.*: **Comparison of tilmicosin with long-acting oxytetracycline for treatment of respiratory tract disease in calves.** *J Am Vet Med Assoc.* 1996; **208**(1): 102–106.
PubMed Abstract | **Publisher Full Text**

Neumann I, Morgan R, Langendam M, *et al.*: **Limitation in the design or execution of randomized trials.** In: Neumann I, Schünemann H. *The GRADE Book version 1.0.* The GRADE Working Group. (updated August 2025).
Reference Source

Nutsch RG, Skogerboe TL, Rooney KA, *et al.*: **Comparative efficacy of tulathromycin, tilmicosin, and florfenicol in the treatment of bovine respiratory disease in stocker cattle.** *Vet Ther.* 2005; **6**(2): 167–179.
PubMed Abstract

O'Connor AM, Yuan C, Cullen JN, *et al.*: **A mixed treatment meta-analysis of antibiotic treatment options for bovine respiratory disease – an update.** *Prev Vet Med.* 2016; **132**: 130–139.
PubMed Abstract | **Publisher Full Text**

Page MJ, McKenzie JE, Bossuyt PM, *et al.*: **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.** *BMJ.* 2021; **372**: n71.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Pardon B, De Bleecker K, Hostens M, *et al.*: **Longitudinal study on morbidity and mortality in white veal calves in Belgium.** *BMC Vet Res.* 2012; **8**: 26.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Pardon B, Hostens M, Duchateau L, *et al.*: **Impact of respiratory disease, diarrhea, otitis and arthritis on mortality and carcass traits in white veal calves.** *BMC Vet Res.* 2013; **9**: 79.
Publisher Full Text

Perrett T, Abutarbush SM, Wildman BK, *et al.*: **A comparison of florfenicol and tulathromycine for the treatment of undifferentiated fever in**

feedlot calves. *Vet Ther.* 2008; **9**(2): 128–140.
PubMed Abstract

Renaud D, Pardon B: **Preparing male dairy calves for the veal and dairy beef industry.** *Vet Clin North Am Food Anim Pract.* 2022; **38**(1): 77–92.
PubMed Abstract | Publisher Full Text

Robb EJ, Tucker CM, Corley L, *et al.*: **Efficacy of tulathromycin or enrofloxacin for initial treatment of naturally occurring bovine respiratory disease in feeder calves.** *Vet Ther.* 2007; **8**(2): 127–135.
PubMed Abstract

Rowan TG, Sarasola P, Sunderland SJ, *et al.*: **Efficacy of danofloxacin in the treatment of respiratory disease in European cattle.** *Vet Rec.* 2004; **154**(19): 585–589.
Publisher Full Text

Sackett DL, Rosenberg WM, Gray JA, *et al.*: **Evidence based medicine: what it is and what it isn't.** *BMJ.* 1996; **312**(7023): 71–72.
PubMed Abstract | Publisher Full Text

Scahill KA: **Efficacy of antimicrobial treatment for bovine respiratory disease: a systematic review and meta-analysis for European Network for Optimization of Veterinary Antimicrobial Therapy guidelines.** *OSF.* 2025 Dec 30.
Publisher Full Text

Schunicht OC, Booker CW, Guichon PT, *et al.*: **An evaluation of the relative efficacy of a new formulation of oxytetracycline for the treatment of undifferentiated fever in feedlot calves in western Canada.** *Can Vet J.* 2002; **43**(12): 940–945.
PubMed Abstract | Free Full Text

Schunicht OC, Booker CW, Guichon PT, *et al.*: **An evaluation of the relative efficacy of tulathromycin for the treatment of undifferentiated fever in feedlot calves in Nebraska.** *Can Vet J.* 2007; **48**(6): 600–606.
PubMed Abstract

Skogerboe TL, Rooney KA, Nutsch RG, *et al.*: **Comparative efficacy of tulathromycin versus florfenicol and tilmicosin against undifferentiated bovine respiratory disease in feedlot cattle.** *Vet Ther.* 2005; **6**(2): 180–196.
PubMed Abstract

Thomas E, Caldow GL, Borell D, *et al.*: **A field comparison of the efficacy and tolerance of marbofloxacin in the treatment of bovine respiratory disease.** *J Vet Pharmacol Ther.* 2001; **24**(5): 353–358.
PubMed Abstract | Publisher Full Text

United States Department of Agriculture (USDA): **Feedlot 2011. Part I: management practices on U.S. feedlots with a capacity of 1000 or more head.** 2011; [cited 2025 Oct 19].
Reference Source

Urie NJ, Lombard JE, Shivley CB, *et al.*: **Preweaned heifer management on US dairy operations: part V. Factors associated with morbidity and mortality in preweaned dairy heifer calves.** *J Dairy Sci.* 2018; **101**(10): 9229–9244.
PubMed Abstract | Publisher Full Text

USDA–APHIS–VS–CEAH–NAHMS: **Antimicrobial Use and Stewardship on U.S. Feedlots, 2017.** *Fort Collins (CO).* 2017; [cited 2025 Dec 1].
Reference Source

Van Donkersgoed J, Berg J, Hendrick S: **Comparison of florfenicol and tulathromycin for the treatment of undifferentiated fever in Alberta feedlot calves.** *Vet Ther.* 2008; **9**(4): 275–281.
PubMed Abstract

Van Donkersgoed J, Hendrick S: **Clinical efficacy of gamithromycin versus florfenicol for the treatment of undifferentiated fever in winter-placed feedlot calves.** *Bovine Pract.* 2013; **47**(2): 152–158.
Publisher Full Text

Wang M, Schneider LG, Hubbard KJ, *et al.*: **Cost of bovine respiratory disease in preweaned calves on US beef cow-calf operations (2011–2015).** *J Am Vet Med Assoc.* 2018; **253**(5): 624–631.
Publisher Full Text

Watts JL, Sweeney MT: **Antimicrobial resistance in bovine respiratory disease pathogens: measures, trends, and impact on efficacy.** *Vet Clin North Am Food Anim Pract.* 2010; **26**(1): 79–88.
PubMed Abstract | Publisher Full Text

Welling V, Lundeheim N, Bengtsson B: **A pilot study in Sweden on efficacy of benzylpenicillin, oxytetracycline, and florfenicol in treatment of acute undifferentiated respiratory disease in calves.** *Antibiotics (Basel).* 2020; **9**(11): 736.
PubMed Abstract | Publisher Full Text | Free Full Text

Windeyer MC, Leslie KE, Godden SM, *et al.*: **Factors associated with morbidity, mortality, and growth of dairy heifer calves up to 3 months of age.** *Prev Vet Med.* 2014; **113**(2): 231–240.
PubMed Abstract | Publisher Full Text

World Health Organization: *WHO's list of medically important antimicrobials: a risk management tool for mitigating antimicrobial resistance due to non-human use.* Geneva: World Health Organization, Licence: CC BY-NC-SA 3.0 IGO; 2024.
Reference Source

Zeng L, Brignardello-Petersen R, Hultcrantz M, *et al.*: **GRADE guidance 34: update on rating imprecision using a minimally contextualized approach.** *J Clin Epidemiol.* 2022; **150**: 216–224.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ? ?

Version 2

Reviewer Report 02 June 2026

https://doi.org/10.21956/openreseurope.25700.r74461

© **2026 Mohammed M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Majed H Mohammed
University of Adelaide, Adelaide, Australia

This systematic review addresses an important clinical and antimicrobial-stewardship question: whether commonly used antimicrobial classes differ meaningfully in the treatment of bovine respiratory disease (BRD). The manuscript is valuable because it uses a structured systematic review design, predefined PICO questions, randomized controlled trials, and GRADE methodology, and it aims to inform future ENOVAT antimicrobial treatment guidelines. However, several points require cautious interpretation before the findings are translated into clinical or guideline decisions.

First, the central conclusion should remain cautious. The review reports that no clinically meaningful difference was detected between antimicrobial classes in reducing retreatment. However, this should not be interpreted as evidence that antimicrobial classes are therapeutically equivalent. Absence of a detected difference is not the same as proof of equal efficacy, especially when several comparisons are based on a limited number of studies and have low or very low certainty of evidence. Therefore, the safest interpretation is that the available randomized evidence did not demonstrate a clinically meaningful superiority of one antimicrobial class over another under the authors' predefined thresholds.

Second, class-level pooling is an important limitation, particularly for macrolides. Although the review was designed to answer class-level PICO questions, the included evidence does not represent all drugs within each class equally. This is especially relevant because some oxytetracycline comparisons were mainly against tilmicosin, whereas several florfenicol comparisons involved tulathromycin, with some tilmicosin also included. Therefore, broad conclusions about "macrolides" may obscure possible drug-level differences within the class. The manuscript should make clear that the findings apply to class-level evidence within the predefined review framework and should not be interpreted as proof that individual macrolides, such as tilmicosin, tulathromycin, gamithromycin, or others, are therapeutically interchangeable in BRD.

Third, retreatment is a practical field outcome, but it remains an indirect surrogate for short-term therapeutic failure. Retreatment is not identical to clinical cure, microbiological cure, relapse

prevention, mortality reduction, lung recovery, welfare improvement, or growth performance. Retreatment decisions can also be influenced by farm protocol, veterinary judgement, treatment threshold, product label, cost, labour, withdrawal period, and producer preference. Therefore, retreatment should be interpreted as a useful but limited surrogate outcome, not as a complete measure of therapeutic success.

Fourth, the variation in follow-up duration across studies needs careful interpretation. Some studies assessed outcomes within only a few days, whereas others followed animals for several weeks or until slaughter. A retreatment event shortly after initial therapy is not biologically identical to delayed relapse, chronic disease progression, or a later management-driven second treatment. Pooling retreatment across different observation windows may therefore weaken the biological and clinical interpretation of the outcome.

Fifth, BRD is a multifactorial syndrome, and the included trials varied in cattle age, production system, diagnostic criteria, disease severity, pathogen status, management conditions, and possibly previous antimicrobial exposure. BRD in pre-ruminant dairy calves, veal calves, slaughter calves, beef cattle, and feedlot cattle may not represent the same clinical situation. This heterogeneity should be emphasized when interpreting the external validity of the pooled findings.

Sixth, the feedlot subgroup should be interpreted cautiously. Feedlot cattle should not be treated as one uniform high-risk population. Newly received, lightweight, commingled calves may have a high BRD risk, whereas some yearling feedlot placements may have lower morbidity. Pooling all feedlot animals together may hide important differences in baseline risk, pathogen pressure, stress burden, prior exposure, and treatment response.

Seventh, the Mycoplasma bovis subgroup is clinically important but limited. Many studies did not test for Mycoplasma or reported Mycoplasma status incompletely. Because M. bovis has distinct biological behaviour and treatment-response patterns, stronger pathogen-level stratification would improve interpretation. Future studies should separate M. bovis-positive, M. bovis-negative, and unknown-status BRD cases wherever possible.

Eighth, several clinically important outcomes were reported inconsistently and could not be reliably pooled, including mortality, relapse, weight gain, and longer-term outcomes. This limits the ability to judge total therapeutic success using retreatment alone. A treatment comparison based only on retreatment may miss differences in survival, relapse, chronic lung damage, growth performance, welfare, or long-term recovery.

Ninth, risk of bias should be directly linked to interpretation. Because retreatment can involve subjective clinical judgement, lack of blinding, inconsistent outcome assessment, or unclear treatment-failure criteria may influence whether an animal is classified as needing additional treatment. This is particularly important in field trials where farm staff or veterinarians make retreatment decisions.

Tenth, the English-language restriction and the absence of grey-literature searching may introduce publication or regional bias. Relevant BRD treatment studies from non-English-speaking cattle-producing regions may have been missed. This matters because antimicrobial availability, disease ecology, management systems, resistance patterns, and treatment practices differ

between regions.

Finally, the review should clearly separate evidence synthesis from treatment recommendation. The review can inform guidelines, but it should not be read as a standalone recommendation that any antimicrobial class can be chosen freely. Final antimicrobial selection should integrate clinical evidence with antimicrobial stewardship, WHO/EMA categorization, local resistance data, suspected pathogen involvement, disease severity, animal welfare, withdrawal period, product availability, and veterinary clinical judgement.

Overall, the strongest message is not that antimicrobial classes are equally effective for BRD, but that current head-to-head randomized evidence is limited, heterogeneous, and often indirect. The review did not detect a clinically meaningful class-level short-term advantage using retreatment as the main surrogate outcome. Future BRD trials should use standardized case definitions, pathogen confirmation, severity scoring, fixed follow-up periods, drug-level comparisons, clear relapse definitions, mortality data, growth performance, welfare outcomes, lung assessment where feasible, and antimicrobial susceptibility information.

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results presented in the review?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My areas of expertise include veterinary microbiology, antimicrobial resistance (AMR), Mycoplasma research, molecular diagnostics, bacterial isolation and characterization, respiratory infectious diseases, antimicrobial therapy evaluation, and veterinary translational research. My assessment of this manuscript primarily focused on the microbiological, antimicrobial stewardship, methodological, and clinical interpretation components related to bovine respiratory disease and antimicrobial efficacy evaluation.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 31 Aug 2026

Karolina Scahill

We thank the reviewer for their careful reading and for emphasizing the need for cautious

interpretation. We agree with the reviewer's central points. These issues were specifically addressed in the revised manuscript (version 2), including cautious wording to avoid implying equivalence, explicit definition of retreatment as a surrogate outcome, discussion of heterogeneity and follow-up variation, cautious interpretation of the feedlot and Mycoplasma subgroup analyses, incorporation of risk-of-bias judgments into GRADE certainty, and acknowledgement of language and publication-bias limitations. We have addressed each point below and, where relevant, refer to the companion ENOVAT BRD guideline, where treatment recommendations are developed separately using Evidence-to-Decision frameworks.

Response to comment 1
We agree with this interpretation and thank the reviewer for emphasizing this important distinction. The manuscript has been revised throughout to avoid implying equivalence or non-inferiority. We now use cautious wording, such as "we did not detect clinically meaningful differences" and "no class demonstrated a clinically meaningful short-term advantage," rather than wording that could be read as therapeutic equivalence. We also emphasize in the Abstract, Discussion, and Conclusion that certainty ranged from very low to moderate and that the estimates should be interpreted cautiously. The Discussion explicitly states: "Our findings therefore do not imply equivalence; rather, within the limits of the available trials and our prespecified clinical thresholds, no class demonstrated a short-term advantage." This is the intended central message of the review and aligns with the reviewer's suggested interpretation.
Response to comment 2
The present review was designed to answer class-level PICO questions and was not intended to evaluate therapeutic differences between individual drugs within a class. The systematic review does not claim that individual macrolides are therapeutically interchangeable. Substance-level considerations, including within-class macrolide comparisons, are addressed in the recently published ENOVAT BRD guideline in response to reviewer concerns raised during peer review of this systematic review (guideline reference at the end of this document). In the guideline, agent-specific sensitivity analyses within PICO 8 and a targeted rapid review of peer-reviewed tulathromycin versus tilmicosin trials were performed. These analyses did not identify clinically meaningful differences between individual macrolides according to the same prespecified thresholds; for example, tulathromycin resulted in 64 fewer retreatments per 1000 animals compared with tilmicosin, but the confidence interval ranged from 125 fewer to 29 more, remaining within the trivial-effect range.
Response to comment 3 We agree. This is why retreatment is described throughout the manuscript as a surrogate for short-term therapeutic failure rather than as a complete measure of therapeutic success. The manuscript defines retreatment separately from relapse and notes that mortality, relapse, growth, and other longer-term outcomes were reported inconsistently and could not be reliably pooled. We also agree that retreatment decisions may be influenced by farm protocols, veterinary judgement, product label, cost, labour, withdrawal periods, and producer preference. These issues were incorporated into the GRADE assessment by downgrading all PICO comparisons for indirectness.
Response to comment 4 We agree that early retreatment, delayed relapse, and chronic disease progression are not biologically identical. This is why retreatment is interpreted as a surrogate and all PICO comparisons were downgraded for indirectness. At the same time,

the analyses were direct pairwise randomized comparisons: within each study, both treatment arms had the same follow-up window and retreatment criteria. Across studies, follow-up windows differed, but we assessed inconsistency using the prespecified threshold-based GRADE approach and did not detect clinically important heterogeneity.

Response to comment 5 We agree that BRD is heterogeneous across production systems and that this affects external validity. The manuscript now discusses variation in age, production system, dose/route, follow-up window, disease pressure, pathogen status, and BRD case definitions. We also explored production system as a potential effect modifier in the feedlot versus non-feedlot subgroup analysis. These subgroup analyses were small and imprecise, but did not show credible effect modification. We therefore interpret the findings cautiously and emphasize that the review provides class-level comparative effect estimates from eligible randomized trials rather than universal treatment recommendations. Local implementation and applicability across different production systems are addressed in the companion ENOVAT BRD guideline using the Evidence-to-Decision framework and GRADE-ADOLOPMENT approach.

Response to comment 6 We agree. The manuscript has been revised to avoid describing feedlot cattle as a uniform high-risk population. We now state that risk in feedlots varies by cohort: newly received, commingled, lightweight calves are often higher risk, whereas many yearling placements are lower risk. The feedlot versus non-feedlot subgroup analysis was exploratory and is interpreted cautiously, as it may not capture within-feedlot differences in baseline risk, pathogen pressure, stress burden, previous exposure, or management. These applicability issues are considered further in the companion ENOVAT BRD guideline, where recommendations are intended for local adaptation.

Response to comment 7

We agree. The manuscript now presents Mycoplasma status more cautiously and distinguishes trials with known Mycoplasma-positive animals from trials where status was unknown or not assessed. We also note that many included studies did not test for Mycoplasma, so some Mycoplasma-positive animals were likely included in the main analysis. The Mycoplasma-positive subgroup analysis was small and imprecise and is not used to draw firm conclusions. We agree that future trials should report M. bovis-positive, M. bovis-negative, and unknown-status animals separately whenever possible. In the companion ENOVAT BRD guideline, suspected or detected M. bovis is explicitly included as a contextual factor when selecting therapy.

Response to comment 8

We agree. This is why the manuscript does not present retreatment as a complete measure of therapeutic success. Mortality, relapse, adverse effects, and weight gain were designated as important or critical outcomes, but were reported inconsistently and could not be reliably pooled or attributed to the initial antimicrobial treatment, particularly when rescue treatments were used during follow-up. These outcomes are reported descriptively in Supplementary File 3, and the inability to pool them contributed to downgrading for indirectness in the GRADE assessment. We also highlight this as a priority for future BRD trials. The companion ENOVAT BRD guideline incorporates this uncertainty when moving from evidence synthesis to recommendations.

Response to comment 9

We agree. Retreatment can be influenced by clinical judgement and trial procedures, especially when blinding or retreatment criteria are unclear. These concerns were explicitly considered in the risk-of-bias assessment, particularly in the outcome-measurement

domain. Risk of bias directly influenced the GRADE certainty ratings: most comparisons were downgraded for risk of bias, and in some cases downgraded by two levels; only PICO 8 and PICO 9 were not downgraded for this domain. The risk-of-bias judgments are shown in the forest plots and reported in full in Supplementary File 5, and they are reflected in the cautious interpretation of the results throughout the manuscript.

Response to comment 10

We agree. The manuscript acknowledges the English-language restriction and absence of grey-literature searching as limitations. We screened reference lists of included studies and a previous systematic review to mitigate missed records, but language bias and publication/availability bias cannot be excluded. Applicability may also vary across regions with different production systems, pathogen and AMR epidemiology, antimicrobial availability, product authorization, and treatment practices. These issues are addressed further in the companion ENOVAT BRD guideline, where local adaptation using GRADE-ADOLOPMENT is encouraged.

Reference to the guideline Scahill K, Buczinski S, Firth CL et al. European Network for Optimization of Veterinary Antimicrobial Therapy (ENOVAT) guidelines for antimicrobial and non-steroidal anti-inflammatory drug use in bovine respiratory disease [version 1; peer review: awaiting peer review]. Open Res Europe 2026, 6:296 (https://doi.org/10.12688/openreseurope.24099.1)

Competing Interests: No competing interests were disclosed.

Reviewer Report 20 May 2026

https://doi.org/10.21956/openreseurope.25700.r73367

© 2026 Reinoso E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Elina B Reinoso
National University of Río Cuarto,, Río Cuarto, Argentina

The authors have made a substantial and good-faith effort to address the concerns raised in the previous review. The revisions have improved the clarity, transparency, and cautious interpretation of the findings.

The revised manuscript now clarifies that the analysis was prespecified at the antimicrobial class level and that substance-level comparisons were outside the scope of the review.
However, I recommend that this limitation be made more explicit in the Discussion, to ensure that readers do not overinterpret class-level findings. In particular, class-level pooling may obscure clinically relevant differences among individual drugs within the same antimicrobial class.

Overall, the revised manuscript has adequately addressed the main concerns, but I recommend this minor clarification before final approval.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Veterinary microbiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 31 Aug 2026

Karolina Scahill

We apologize for the delayed response and thank the reviewer for their constructive assessment of the revised manuscript. The delay was because we considered the companion ENOVAT BRD guideline to be the most appropriate place to address the substance-level and implementation issues raised here, and we wanted to be able to refer the reviewer to the published guideline. We agree that readers should not overinterpret class-level findings as evidence of therapeutic interchangeability between individual drugs within the same antimicrobial class. However, we view this primarily as an issue of scope and interpretation rather than a methodological limitation. The present systematic review was designed and prespecified to answer class-level PICO questions; adding substance-level comparisons would have meant answering a different research question with different PICOs, eligibility criteria, and analyses. Presenting such results within this review would therefore have risked methodological inconsistency rather than strengthening the prespecified synthesis. We took the reviewer's concern seriously and addressed it explicitly in the ENOVAT BRD guideline, where substance-level considerations belong within the Evidence-to-Decision process. Specifically, the guideline includes agent-specific sensitivity analyses within PICO 8 and a targeted rapid review/meta-analysis of peer-reviewed macrolide trials. These analyses did not identify clinically meaningful differences between individual macrolides according to the same prespecified thresholds. Full reference to guideline below and direct link here: https://open-research-europe.ec.europa.eu/articles/6-296 Scahill K, Buczinski S, Firth CL et al. European Network for Optimization of Veterinary Antimicrobial Therapy (ENOVAT) guidelines for antimicrobial and non-steroidal anti-inflammatory drug use in bovine respiratory disease [version 1; peer review: awaiting peer review]. Open Res Europe 2026, 6:296 (https://doi.org/10.12688/openreseurope.24099.1)

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Report 06 April 2026

https://doi.org/10.21956/openreseurope.23668.r70707

© 2026 Apley M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Michael D Apley
Kansas State University, Manhattan, Kansas, USA

General comments:
First, a very well written paper as far as being clear and able to follow the processes. I greatly appreciate the transparency in methods and also in the data used.

However, it appears, from the overall conclusion of the paper, that through the meta-analysis process the plural of underpowered superiority studies is now a non-inferiority conclusion based on low quality data? This is a major leap in how far the results of multiple small studies apply to different locations and different types of cattle. The disease pressure in quite a few of the papers appears to be relatively mild as evidenced by the treatment success rates and case fatality, leading to a good number of spontaneous cures which greatly dilutes treatment differences. Much of this paper in which there are direct comparisons of penicillin or oxytetracycline to other drugs is based on moderate to relatively low risk cattle populations.

Relying on my view, I would opine that there are very, very few high-risk BRD protocols in the USA that would lead off with, or even include in the first two treatments, a form of oxytetracycline. The reason is that the larger systems have done their own internal work, and there is a large difference of oxytetracycline from drugs like tulathromycin and florfenicol for serious BRD challenges. I realize that this information may not available to the authors, but strengthens my resolve that this paper does not accurately reflect the true nature of relative efficacy for these antimicrobials.

In the body of the last paragraph, the authors did not “show no difference” (non-inferiority), they in fact detected no difference (non-significance in a superiority study). The next sentence says it better, “no clear differences were detected”.

I have commented below on the problem with considering macrolides a homogenous group. One could argue therapeutic similarity within the macrolides (for bovine respiratory disease) for tildipirosin and tulathromycin, maybe gamithromycin, but definitely not tilmicosin. Note that oxytetracycline comparisons are against tilmicosin, while florfenicol is primarily against tulathromycin (some tilmicosin). Tulathromycin and tilmicosin are very different drugs when considering efficacy for bovine respiratory disease. If it isn’t possible to get sufficient numbers of studies within each comparison without lumping antimicrobial classes, then the analysis should not be done. A requirement for numbers doesn’t justify the use of scientifically unsound groupings.

If considered too biased by the authors due to a company funded review, the paper below should at least create the need to consider that tulathromycin and tilmicosin are different in therapeutic response.

Nautrup, et al. J Anim Sci 2013 Dec;91(12):5868-77. doi: 10.2527/jas.2013-6709

Specific comments:
The prevalence estimates for U.S. feedlots misrepresent the overall prevalence. (Introduction, second paragraph). For overall morbidity, where in the cited USDA report is there an estimate of respiratory disease morbidity? This is a survey of practices at the feedlot level. I went back to this report and searched morbidity and treatment as terms, and went back through the sections, and could find no morbidity estimates. Even if there were, they would be estimates only and not appropriate for reference as to morbidity rates without specifically indicating they are estimates. How are the Abell (a metaphylaxis meta-analysis) and Baptista (a paper on Brazil, which is not north America) references appropriate for overall estimates of morbidity in U.S. beef feedlots? The morbidity cited is possible for high-risk cattle but is not representative of the population as a whole. The authors should either find appropriate references and put the morbidity estimates within the correct type of cattle, or delete these estimates.

"Feedlot cattle are a high-risk population" (page 4). Some are, the majority are not. The vast majority of placements are still yearlings, with expected respiratory morbidity rates well below 10%, with 1-5% not uncommon.

I went through the oxytet comparison studies to see how they shook out.

In looking at studies for **oxytet vs. macrolides**, a major error is extrapolating from tilmicosin to other macrolides in the treatment of bovine respiratory disease.
Keita, et al. OTC, OTC/flunixin, Tilmicosin This study compares OTC to the lowest efficacy macrolide (tilmicosin), which consistently loses comparisons to tulathromycin. Considering macrolides as one group, efficacy-wise, is a major error. This is also in pre-ruminant calves. How does that apply to a 300 kg feedlot animal? Rectal temp mean at treatment was around 40.0 to 40.1 C, indicating there were few animals displaying high rectal temperatures in the mix, supporting the concept of a low potency respiratory challenge. Also, someone please convince me of the applicability of an overall clinical score in assessing efficacy of respiratory disease treatment. Assuming 50 animals per treatment, with an average between groups of 8 retreated, that is a success rate of 84%, which would be outstanding treatment response, likely due to a very mild disease challenge. Mild disease challenge makes it more difficult to find a difference in efficacy if there is one.

Laven, et al. OTC vs. tilmicosin. Again, extrapolating from tilmicosin to all macrolides, a major error. 16 calves and 15 calves in the two treatment groups, a very small study. Just looked at the abstract, but appears even with the low numbers that tilmicosin had a lower retreat rate?

Musser, et al. 1996 OTC vs. tilmicosin Very low numbers, 21 and 20 head in treatment groups. The macrolide won on 3 day post-treatment signs but no significant difference in the other outcomes.

Oxytet and florfenicol
Catry, et al. 2008. The authors considering doxycycline equivalent to oxytetracycline? We consider tetracycline as a class representative in antimicrobial susceptibility testing, but to consider oxytet and doxy therapeutically equivalent as well is unwarranted. Also, this was metaphylaxis. You are including metaphylaxis or not? Your abstract and text says not.

Schunicht, et al. 2002 This study used florfenicol at the 20 mg/kg Q48H initial regimen (this was 1997), not the currently dominant 40mg/kg once regimen. Florfenicol won on chronics, which is pretty important.

Welling, et al. Again, treatment success rates (first treatment) of 86-90% tells me it wasn't a very severe challenge. I see no indication this study was masked or blinded. Small study. How do these young calves equate to older animal BRD? The authors note that Mycoplasma really hasn't been a problem for them. An unblinded study where the authors goal is to support the penicillin recommendation for initial BRD therapy? Why was the bias acceptable for inclusion of this paper and not others?

Oxytet and fluoroquinolones
Giles, et al. 1991 Oxytet and danofloxacin. The article link wasn't working, so just looked at the abstract. A small study, danofloxacin apparently won. In the table, danofloxacin was evidently a daily dose of 1.25 mg/kg. The US label dose is 8 mg/kg once, making your data relevance to commonly used doses questionable.

Leukex, et al. The link wasn't working for me so just looked at the abstract. Masked evaluators? Enrofloxacin dose is daily at 2.5 mg/kg or 5 mg/kg IM. The US label dose is single injection of 7.5 – 12.5 mg/kg, usually in the middle. Probably fairly equivalent to the 5 mg/kg daily.

What do I suggest?

- In your conclusion, put additional emphasis on the wide spectrum of disease severity in these studies, the very short term evaluation in some of the studies, the lack of consideration of other important therapeutic outcome metrics, and the limitations on projection to high-risk cattle for comparisons involving penicillin and oxytetracycline. The emphasis on indicating the studies which document the presence of M. bovis is very appropriate, and supports the conclusion that this is an incredibly disparate set of papers being synthesized into one conclusion.
- The Catry, et al., study was metaphylactic treatment of cattle in close exposure to other cattle which were treated. This was not treatment of acute disease and should be removed.
- The consideration of tilmicosin as therapeutically equivalent to other macrolides, especially tulathromycin and tildipirosin is an error. This inflates the proposed lack of difference between penicillins/oxytetracycline and "macrolides" as a group. You need a direct comparison of penicillin and oxytetracycline to tulathromycin before commenting on comparison to the macrolides.

There is a lot of work in this paper, presented with transparency and access to methods and additional data. Unfortunately, the authors make an unsupported statement in the abstract which states "For first-line treatment of BRD, the five antibiotic classes studied appear to perform similarly in the short term". The analysis, and the data, are insufficient to make this statement. The specific outcomes of performance, case fatality, and relapses (an arbitrary division of retreatment based on time to retreatment) were noted to be absent from the analysis. A major limitation is that the paper is based on only a partial picture of therapeutic outcomes without recognizing that animal welfare and producer economic survival need to be in balance; the authors fail to articulate that those are also important aspects of BRD therapy (which are left out of the analysis) in addition to the stated objectives of responsible use and public health considerations. The appropriate conclusion would be a lack of sufficient data to compare penicillin and oxytetracycline to other antimicrobials in the therapy of bovine respiratory disease,

not that they “perform similarly”.

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results presented in the review?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Clinical pharmacology, bovine respiratory disease, beef production medicine

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 Apr 2026

Karolina Scahill

Thank you for your kind words about the clarity and transparency of the work, and for sharing a high-risk US feedlot perspective. We agree that local experience, feasibility, authorization, and stewardship priorities shape protocol choices—and that many large US systems might not lead off with oxytetracycline in high-risk cohorts. That is not claimed anywhere in this manuscript. Your points have been very helpful to us in drafting the companion ENOVAT guideline, where feasibility and stewardship are weighed explicitly alongside the trial effects. The guideline will be open for consultation (hopefully within 1-2 months) to include further perspectives, and we would genuinely value your input also there. **Response to general comments Methodology and scope** This paper is a systematic review (SR) designed to answer prespecified PICO questions at the antimicrobial class level, using all eligible head-to-head randomized trials we could find. Our starting point is the clinical questions we need to answer, not where evidence happens to be abundant. When evidence is sparse, the GRADE approach does not advise abandoning the question; it asks us to synthesize the best available evidence, be transparent about limitations, and rate down certainty accordingly. That is what we have done. Choosing not to analyze because numbers are small would leave an important decision questions unanswered and shift attention only to topics with plentiful data, and evidence gaps that are important to inform practice remains obscured. This would also introduce serious publication bias since published trials are often (industry) driven by the formulation of new molecules as mentioned in the discussion. We are not making recommendations in this publication, nor are we placing additional value judgements on the results—we simply report all studies that

fit the inclusion criteria. We synthesized what exists; we cannot analyze direct comparisons that have not been conducted, or unpublished internal data we could not systematically retrieve (non-use of grey literature was mentioned as a limitation in the discussion). Excluding eligible studies (e.g., those using tilmicosin) because of perceived performance would introduce selective reporting. Historically, SRs have been "hidden" in the supplements of evidence-based guidelines but are nowadays often published as standalone accompanying papers that don't include recommendations and Evidence-to-Decision considerations. We added a statement in the Introduction and Methods that this SR estimates comparative effects at the antimicrobial class level and does not make recommendations or place value judgements on the results. Substance-level (within-class) comparisons and implementation trade-offs are beyond the SR's scope and are handled in the companion ENOVAT guideline's Evidence-to-Decision framework. **Interpretation of results** To our knowledge, GRADE has not previously been applied in bovine medicine, and its terminology is likely unfamiliar to many in the field, which can cause confusion. Regarding interpretation, we do not make firm conclusions based on low-quality data. On the contrary, the quality (certainty in GRADE terminology) of the evidence is explicit and central to our conclusions. Low and very low certainty literally means that the true effects may (or probably do) differ markedly from the point estimates. Certainty ratings are central to the GRADE approach and are results in their own right. However, very low certainty does not mean results should be withheld. We report that we did not detect clinically meaningful between-class differences on our primary outcome (need for re-treatment as a surrogate), judged against prespecified clinical thresholds, which states and signals how much we trust the results. To reflect this, we have softened wording throughout. We have removed wording such as "appear to perform similarly" and "showed no difference" and have strengthened the emphasis on caution in the interpretation of results. **Class-level synthesis** We respectfully disagree that class-level synthesis is "scientifically unsound." The review was prespecified to answer class-level PICOs to support guideline decisions that are often made at the class level, consistent with stewardship frameworks such as WHO priorities and EMA/AMEG categories. In other words, the decision unit we needed to inform was the class, not the individual substance. Our aim was to reflect the evidence landscape across production systems, not only high-risk North American feedlots (and we do not claim or recommend that oxytetracycline should be used in those settings). In many European contexts, tetracyclines remain common first-line options, and penicillin in Scandinavia. In the EU, veterinary prescribing is governed by stewardship legislation and guidance that operationalise the EMA/AMEG categorisation; clinicians are expected to prioritise lower-category options where appropriate and restrict higher-category agents. Class-level comparisons are therefore directly relevant for those users. This review fills an important evidence gap in that context. **Macrolides** For readers seeking substance-level rankings (macrolide vs macrolide) in high-risk feedlot settings, prior syntheses exist (e.g., the network meta-analysis by O'Connor cited in our discussion). Our SR answers a different, prespecified question and reports the best available class-level RCT evidence with transparent certainty judgements. The search and inclusion criteria did not encompass within-class head-to-head trials (macrolide vs macrolide), so they were not sought or synthesized. This SR therefore does not make substance-level efficacy claims. Tilmicosin is not extrapolated for all macrolides in the comparison against oxytetracyclines; it is simply the macrolide tested in those eligible trials. In PICO 8 and PICO 9 different macrolides were represented and we did not detect clinically meaningful inconsistency in regards to the performance of TUL and TIL

against FLO and QUO. There were trivial-to-small absolute differences against our prespecified threshold also in sensitivity analysis when only TUL and TIL were pooled. The results are beyond the scope of the present manuscript but will be adressed in the guideline.We are careful not to imply real-life equivalence; we state that no clinically meaningful between-class differences were detected within the limits of the available RCTs and certainty ratings. **Response to specific comments** **Reviewer comment: "In the body of the last paragraph, the authors did not 'show no difference' (non-inferiority), they in fact detected no difference (non-significance in a superiority study). The next sentence says it better, 'no clear differences were detected'."** We agree with the distinction. We have revised the wording to better reflect this. We now state that we "did not detect clinically meaningful differences" (based on prespecified clinical thresholds) and use "no clear differences were detected" rather than "showed no difference." This wording has been applied consistently in the abstract, subgroup summaries, discussion and conclusion. **Reviewer comment: "Welling et al.: high success rates suggest a mild challenge; appears unblinded; small study; how do these young calves equate to older animal BRD? Why was the bias acceptable for inclusion of this paper and not others?"** We did not exclude trials on the basis of risk of bias. All eligible head-to-head RCTs of individual treatment were included; risk of bias was assessed domain-by-domain and reported transparently (In figures and in Supplementary File 5). Excluding studies because they are small, unblinded, or show high success rates would introduce selection bias. Interestingly, no heterogeniety was detected despite the difference in populations. The SR intentionally synthesized across production systems to reflect the broader evidence landscape. Within each trial, randomization balances severity and case mix between arms, so the direct comparison remains valid for that setting. We also explored feedlot vs non-feedlot in an exploratory subgroup; estimates were imprecise but did not show credible effect modification. **Reviewer comment: "The prevalence estimates for U.S. feedlots misrepresent the overall prevalence. (Introduction, second paragraph). For overall morbidity, where in the cited USDA report is there an estimate of respiratory disease morbidity? This is a survey of practices at the feedlot level. I went back to this report and searched morbidity and treatment as terms, and went back through the sections, and could find no morbidity estimates. Even if there were, they would be estimates only and not appropriate for reference as to morbidity rates without specifically indicating they are estimates. How are the Abell (a metaphylaxis meta-analysis) and Baptista (a paper on Brazil, which is not north America) references appropriate for overall estimates of morbidity in U.S. beef feedlots? The morbidity cited is possible for high-risk cattle but is not representative of the population as a whole. The authors should either find appropriate references and put the morbidity estimates within the correct type of cattle, or delete these estimates."** Thank you for this careful read. We agree that our original wording over-generalized and that the cited sources did not directly support population-wide morbidity estimates for US feedlots. We have revised the Introduction to avoid presenting specific incidence ranges and now state that BRD is the most frequently treated disease and a leading cause of morbidity and mortality in US feedlots. We qualified that rates vary by risk class and management, and—where we mention higher burdens in high-risk, newly received cattle—this is now framed as context rather than a population-wide estimate. We also removed Abell (metaphylaxis meta-analysis) and Baptista (Brazil) from this US-specific morbidity sentence. **Reviewer comment: "'Feedlot cattle are a high-risk population' (page 4). Some are, the majority**

are not. The vast majority of placements are still yearlings, with expected respiratory morbidity rates well below 10%, with 1–5% not uncommon." We agree our wording was too categorical. Our intent was to justify exploring production system as a potential modifier, not to imply that all feedlot placements are high-risk.We have made the following change in the manuscript: "Risk in feedlots varies by cohort; newly received, commingled, lightweight calves are often higher-risk for BRD, whereas many yearling placements are lower-risk. Because some feedlot cohorts may also have had prior BRD treatment, we explored feedlot versus non-feedlot systems as a potential effect modifier in an exploratory subgroup analysis".We also note in the results that this subgroup analysis was small and imprecise and did not show a credible difference in relative effects between production systems. **Reviewer comment: "The Catry, et al., study was metaphylactic treatment of cattle in close exposure to other cattle which were treated. This was not treatment of acute disease and should be removed."** We are very grateful you noticed this error. Metaphylaxis was an exclusion criteria, but Catry also included acute disease arms. On re-checking the details, doxycycline appears as a metaphylactic positive control (trial B), and there is no individual-treatment arm comparing doxycycline with florfenicol so we have removed this from PICO 6 but kept it in PICO 8 where there is a comparison of animals with acute clinical disease.With Catry excluded, the pooled estimate for PICO 6 is RR 0.96 (95% CI 0.70–1.30), I^2=0%, corresponding to about 7 fewer re-treatments per 1,000 (from 56 fewer to 56 more) so there is still no clinically meaningful difference in the analysis and certainty of evidence remains low. The removal of Catry in PICO 6 did not change the conclusion. **Reviewer comment: "In your conclusion, put additional emphasis on the wide spectrum of disease severity in these studies, the very short term evaluation in some of the studies, the lack of consideration of other important therapeutic outcome metrics, and the limitations on projection to high risk cattle for comparisons involving penicillin and oxytetracycline. The emphasis on indicating the studies which document the presence of M. bovis is very appropriate, and supports the conclusion that this is an incredibly disparate set of papers being synthesized into one conclusion."** Our review is restricted to direct, head-to-head randomized trials (pairwise meta-analyses). Within each trial, randomization balances disease severity and baseline risk between arms, so each comparison is like-with-like in that outbreak context. If this were a network meta-analysis, arms would be compared across studies where severity and baseline risk might differ, raising transitivity concerns; this is one reason we did not pursue an NMA. Between trials we judged inconsistency against prespecified clinical thresholds. Study-level and pooled estimates clustered within the trivial-to-small ranges, and we did not observe clinically important inconsistency (heterogeneity). In other words, while settings differed in disease pressure, our analyses did not show a pattern suggesting moderate or large between-class differences. Interpretation remains cautious, which is reflected in our downgrading for indirectness (and imprecision where applicable). We have added a paragraph in the discussion to make explicit the variation in disease pressure, age, follow-up window, and our heterogeneity approach. Finally, this SR does not include recommendations and does not claim that penicillin or oxytetracycline should be used in high-risk cattle. There is no single overarching "conclusion" beyond the PICO-specific findings: we report what the eligible trials show and pair this with GRADE certainty ratings. Recommendations will be provided in the companion guideline after Evidence-to-Decision considerations. We welcome your opinion and perspective also there.

Competing Interests: No competing interests were disclosed.

Reviewer Report 31 March 2026

https://doi.org/10.21956/openreseurope.23668.r70705

© 2026 Reinoso E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Elina B Reinoso
National University of Río Cuarto,, Río Cuarto, Argentina

This manuscript addresses a highly relevant topic because bovine respiratory disease (BRD) is a major cause of morbidity and mortality in cattle, a major driver of antimicrobial use, and an important issue in the context of AMR and One Health. The systematic review and meta-analysis approach appears appropriate. However, there are several aspects that should be clarified or strengthened to ensure a more robust and cautious interpretation of the findings.

1. Please ensure consistency in terminology throughout the manuscript, particularly regarding Mycoplasma, *Mycoplasmopsis bovis*, *M. bovis*, and Mycoplasma-positive animals.
2. Please revise minor stylistic and language issues throughout the manuscript.
3. Although the review is structured by antimicrobial class, there is substantial clinical heterogeneity across the included trials with respect to cattle age, production systems, dosing regimens, route of administration, follow-up period, and diagnostic criteria for BRD. Please discuss more explicitly how this heterogeneity may affect comparability across studies and the pooled estimates.
4. A key limitation is that the analysis is performed at the antimicrobial class level rather than at the level of individual drugs. This should be discussed more explicitly, since clinically relevant differences may exist within antimicrobial classes.
5. Please clarify the relationship between the outcomes “therapeutic failure,” “relapse,” and “need for re-treatment.” In Table 1, therapeutic failure and relapse are listed as separate outcomes, whereas Table 2 only presents thresholds for therapeutic failure, and the abstract refers to need for re-treatment. This creates possible terminological inconsistency. Please standardize the terminology and clarify whether the same thresholds were applied across all comparisons.
6. The exclusion of non-English studies should be acknowledged more explicitly as a limitation in the Discussion, as it may introduce language bias.
7. The hypothesis that feedlot cattle would have poorer treatment outcomes than non-feedlot cattle, particularly when treated with antimicrobial classes of lower importance in human medicine, implies a subgroup-by-treatment interaction. Please clarify whether this hypothesis was pre-specified in the original protocol or developed during the analysis.
8. The decision not to rely on I^2 statistics for inconsistency assessment is in line with updated GRADE guidance. However, it would be helpful if the manuscript explained more clearly how variability across studies was judged to be clinically important in relation to the predefined

thresholds.

9. The statement that “M+ animals respond to treatment similarly to non-positive animals” appears too strong given the limited subgroup evidence and, in some cases, reliance on a single small study with an imprecise estimate. This conclusion should be stated more cautiously.
10. Please verify whether the EMA (2025) reference is publicly available.
11. Please verify the citation “USDA of Agriculture (2011).
12. There appears to be an inconsistency in the Discussion: PICO 6 corresponds to TET vs FLO, not TET vs MAC. Please revise this point for consistency.
13. The use of a modified RoB 2.0 tool originally developed for swine trials should be better justified, including why the authors consider it applicable to cattle studies.
14. The conclusions should be phrased more cautiously, especially where the certainty of evidence is low or very low. In several PICOs, the absence of a clinically meaningful difference should not be interpreted as conclusive evidence of equivalence.

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Partly

Is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results presented in the review?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Veterinary microbiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 Apr 2026

Karolina Scahill

We thank you for your thoughtful and constructive review. Your comments have helped us strengthen this systematic review by clarifying terminology, expanding limitations (e.g., heterogeneity and language restriction), and making our certainty and interpretation more cautious and transparent. Issues such as within-class drug differences, stewardship trade-offs (WHO/AMEG), feasibility, disease severity/setting are addressed in the companion guidelines Evidence-to-Decision framework. The guideline will be open for consultation (hopefully within 1-2 months) to include further perspectives. We have no representation from South America in the panel (this is a weakness) and would value your input also in the

guideline. **Reviewer comment 1: Please ensure consistency in terminology throughout the manuscript, particularly regarding Mycoplasma, Mycoplasmopsis bovis, M. bovis, and Mycoplasma-positive animals.** Thank you for highlighting this. We have standardized nomenclature throughout: at first mention we use Mycoplasma bovis (syn. Mycoplasmopsis bovis), thereafter Mycoplasma bovis (M. bovis). We use the hyphenated term Mycoplasma-positive as the generic descriptor, and specify M. bovis–positive when species-level identification was available. Table 4 is now headed Mycoplasma status with a footnote clarifying species vs genus reporting. We hope this resolves the terminology and formatting concerns. **Reviewer comment 2: Please revise minor stylistic and language issues throughout the manuscript.** We agree style and language is important for readability, but it's difficult to act on this request since it's very general. Style is subjective, and language issues also can be to some extent. The manuscript went through a light language editing program provided by the journal before publication. We have tried to focus on the scientific terminology standardization mentioned above. **Reviewer comment 3: "Although the review is structured by antimicrobial class, there is substantial clinical heterogeneity across the included trials with respect to cattle age, production systems, dosing regimens, route of administration, follow-up period, and diagnostic criteria for BRD. Please discuss more explicitly how this heterogeneity may affect comparability across studies and the pooled estimates." Reviewer comment 8: "The decision not to rely on I^2 statistics for inconsistency assessment is in line with updated GRADE guidance. However, it would be helpful if the manuscript explained more clearly how variability across studies was judged to be clinically important in relation to the predefined thresholds."** Response to 3 and 8: Thank you for raising heterogeneity and the role of I^2. We agree it would be helpful to add some clarity to our certainty judgements with a specific focus on inconsistency (heterogeneity). We began from an inclusive stance on PICO-element variability rather than assuming a large effect of age, system, dose/route, or follow-up a priori. If there would have been clinically important inconsistency, we would have explored it in more detail (as per the GRADE methodology), but the results did not suggest that these differences had a clinically meaningful effect. We assessed inconsistency using threshold-based absolute effects rather than I^2 alone: within each PICO, point estimates clustered within the same threshold range, confidence intervals overlapped, and we did not observe outliers implying large, clinically important deviations. Because the threshold-based way of judging inconsistency is still new in veterinary medicine, we have added a paragraph in the discussion and have tried to illustrate the process by including an example of a potentially influential study (Schunicht 2007) and how it affected both I^2 and the absolute effects. **Reviewer comment 4: A key limitation is that the analysis is performed at the antimicrobial class level rather than at the level of individual drugs. This should be discussed more explicitly, since clinically relevant differences may exist within antimicrobial classes.** The review was prespecified to answer class-level PICOs to support guideline decisions that are often made at the class level, consistent with stewardship frameworks such as WHO priorities and EMA/AMEG categories. In other words, the decision unit we needed to inform was the class, not the individual substance. Trials comparing drugs within the same class (e.g., macrolide vs macrolide) were therefore out of scope and not sought or synthesized, so this SR cannot make within-class efficacy claims. Within the included class-to-class comparisons, several macrolide substances (e.g., tulathromycin, tilmicosin, gamithromycin) were represented, and we did not observe a coherent pattern or clinically important inconsistency that would suggest one macrolide consistently performed

better than others. We have clarified the scope of the manuscript in the introduction and methods and will address substance selection and its stewardship/feasibility implications in the companion guideline. **Reviewer comment 5: Please clarify the relationship between the outcomes 'therapeutic failure,' 'relapse,' and 'need for re-treatment.' In Table 1, therapeutic failure and relapse are listed as separate outcomes, whereas Table 2 only presents thresholds for therapeutic failure, and the abstract refers to need for re-treatment. This creates possible terminological inconsistency. Please standardize the terminology and clarify whether the same thresholds were applied across all comparisons."** We agree this causes confusion in the manuscript. Our protocol prioritized short-term therapeutic failure (<7 days) as the preferred outcome, but it was rarely and inconsistently reported. Therefore, we used "need for retreatment" (first additional antimicrobial given for the same BRD episode within the study's observation window) as a surrogate for short-term therapeutic failure and rated down for indirectness. Relapse refers to recurrence after documented clinical resolution; it was reported too variably to pool. To make this clearer, we have defined the outcomes and surrogate outcome in more detail in the methods and stated that the clinical thresholds in Table 2 were prespecified for therapeutic failure and were applied to retreatment when used as its surrogate. Relapse was not pooled and therefore not assessed against thresholds. Standardized wording in the abstract and main text to "retreatment (a surrogate for short-term therapeutic failure)." **Reviewer comment 6: "The exclusion of non-English studies should be acknowledged more explicitly as a limitation in the Discussion, as it may introduce language bias."** We fully agree with you in regards to applicability in the interpretation of results in an international context. We prespecified an English-language restriction for feasibility and consistency. While this SR's effect estimates are based on head-to-head RCTs and we do not expect the language restriction to systematically change the direction of relative effects, we agree that language bias cannot be excluded and that it may affect applicability beyond English-language contexts. We have added this as a limitation in the discussion to acknowledge this and will consider applicability across languages explicitly (as well as international representation) in the guideline's Evidence-to-Decision process. **Reviewer comment 7: "The hypothesis that feedlot cattle would have poorer treatment outcomes than non-feedlot cattle, particularly when treated with antimicrobial classes of lower importance in human medicine, implies a subgroup-by-treatment interaction. Please clarify whether this hypothesis was pre-specified in the original protocol or developed during the analysis."** All major protocol deviations are reported in Supplementary File 1 (after the PRISMA checklist) including rationale and the production system subgroup analysis is one of them. This work was part of a COST-funded capacity-building effort in guideline methodology and as far as we know this is the first time GRADE is used on bovine medicine. The protocol deviations reflects the steep learning curve that the team experienced during the process. We considered to make protocol amendments along the process but reasoned it was more transparent to keep it as it is and report deviations. The ICEMAN is a tool used to judge the credibility of claimed subgroup effects (i.e., when results suggest that treatment works differently in one subgroup than another). In our analyses we did not detect a clinically meaningful subgroup-by-treatment interaction, so there was no claim to appraise and ICEMAN was not applied. If any subgroup had suggested a credible difference, we would have evaluated it with ICEMAN and reported that assessment explicitly. **Reviewer comment 9: The statement that 'M+ animals respond to treatment similarly to non-positive animals' appears too strong given the**

limited subgroup evidence and, in some cases, reliance on a single small study with an imprecise estimate. This conclusion should be stated more cautiously." Agreed. We have softened the wording to reflect the small and imprecise evidence base. It now reads: Only trials reporting *Mycoplasma*-positive animals (species-level *M. bovis* when available) were included in this analysis. This information is available for PICO 5 (TET vs. MAC), PICO 8 (FLO vs. MAC), and PICO 9 (MAC vs. QUI). In this small and imprecise subgroup, point estimates were broadly consistent with those of the main analysis. Overall, these data are insufficient to conclude similar responses. Because many included trials did not assess *Mycoplasma* status, some *Mycoplasma*-positive animals were likely present in the main analysis. The effect sizes of the sub-analysis are shown in Table 7, and the forest plots are shown in Supplementary File 6. The Instrument for assessing the Credibility of Effect Modification Analysis (ICEMAN tool) was not used in the subanalysis because a clinically meaningful effect modification was not found. **Reviewer comment 10: Please verify whether the EMA (2025) reference is publicly available.** Response: Yes it's publicly available and the link is provided in the reference section. The link works so we have not changed anything. **Reviewer comment 11: Please verify the citation 'USDA of Agriculture (2011).** The 2011 USDA report is publicly available, and the manuscript already cites it with a working link in the references. Maybe the links don't work on the downloaded manuscript? They work in the online version. **Reviewer comment 12: "There appears to be an inconsistency in the Discussion: PICO 6 corresponds to TET vs FLO, not TET vs MAC. Please revise this point for consistency."** Thank you for catching this! PICO 6 is tetracyclines vs florfenicol. We have corrected this in the discussion. **Reviewer comment 13: "The use of a modified RoB 2.0 tool originally developed for swine trials should be better justified, including why the authors consider it applicable to cattle studies."** The same point could be made about not using the modified swine-tool. The original RoB 2.0 is designed for trials in human medicine and not herd-level field trials in cattle so swine-adapted signaling questions is the closest available fit for livestock RCTs. The core bias domains (randomization, deviations from intended interventions, missing data, outcome measurement, selective reporting) are included and the swine adaptation simply makes the signaling questions more realistic for production-animal trials (e.g., group enrolment, occasional cluster randomization, practical limits on allocation concealment). The GRADE methodology does not prescribe a specific risk of bias tool so this is not a methodological error that necessary needs further justification. **Reviewer comment 14: "The conclusions should be phrased more cautiously, especially where the certainty of evidence is low or very low. In several PICOs, the absence of a clinically meaningful difference should not be interpreted as conclusive evidence of equivalence."** Agreed. In GRADE, low certainty means the true effect might be markedly different from the estimate, and very low means it is probably markedly different; these ratings do not support firm conclusions. We recognise this terminology is not yet widely used in veterinary medicine, so we have softened the language throughout the abstract, discussion and conclusion emphasizing cautious interpretation of the findings.

Competing Interests: No competing interests were disclosed.