Herd management practices associated with multi-drug resistance in fecal bacterial commensals and *Salmonella* shed by cull dairy cows, a machine learning approach

Pranav S Pandit [Corresp., 1], **Deniece R. Williams** [2], **Paul Rossitto** [2], **John Adaska** [3], **Richard Pereira** [4], **Terry W. Lehenbauer** [2, 4], **Barbara A. Byrne** [5], **Xunde Li** [6], **Robert Atwill** [6], **Sharif S. Aly** [Corresp. 2, 4]

[1] EpiCenter for Disease Dynamics, One Health Institute, School of Veterinary Medicine, University of California Davis, Davis, California, The United States of America

[2] Veterinary Medicine Teaching and Research Center, University of California, Davis, Tulare, California, The United States of America

[3] California Animal Health and Food Safety Laboratory System, University of California, Davis, Tulare, California, The United States of America

[4] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, Davis, California, The United States of America

[5] Department of Pathology, Microbiology & Immunology, School of Veterinary Medicine, University of California, Davis, California, The United States of America

[6] Western Institute for Food Safety and Security, University of California, Davis, Davis, California, The United States of America

Corresponding Authors: Pranav S Pandit, Sharif S. Aly
Email address: pspandit@ucdavis.edu, saly@ucdavis.edu

Background: Understanding the effects of herd management practices on the prevalence of multidrug-resistant pathogenic *Salmonella* and commensals *Enterococcus* spp*. and Escherichia coli* in dairy cattle is key in reducing antibacterial resistant infections in humans originating from food animals. Our objective was to explore the herd and cow level features associated with the multi-drug resistant, and resistance phenotype shared between *Salmonella*, *E. coli,* and *Enterococcus* spp. using machine learning algorithms. **Methods:** Randomly collected fecal samples from cull dairy cows from six dairy farms in central California were tested for multi-drug resistance phenotypes of *Salmonella, E. coli,* and *Enterococcus* spp. Using data on herd management practices collected from a questionnaire, we built three machine learning algorithms, decision tree classifier, random forest, and gradient boosting decision trees, to predict the cows shedding multidrug-resistant *Salmonella* and commensal bacteria. **Results:** The decision tree classifier identified rolling herd average milk production as an important feature for predicting fecal shedding of multi-drug resistance in *Salmonella* or commensal bacteria. The number of culled animals, monthly culling frequency and percentage, herd size, and proportion of Holstein cows in the herd were found to be influential herd characteristics predicting fecal shedding of multidrug-resistant phenotypes based on random forest models for *Salmonella* and commensal bacteria. Gradient boosting models showed that higher culling frequency and monthly culling percentages were associated with fecal shedding of multidrug

resistant *Salmonella* or commensal bacteria. In contrast, an overall increase in the number of culled animals on a culling day showed a negative trend with classifying a cow as shedding multidrug-resistant bacteria. Increasing rolling herd average milk production and spring season were positively associated with fecal shedding of multidrug- resistant *Salmonella*. Only 6 individual cows were detected sharing tetracycline resistance phenotypes between *Salmonella* and either of the commensal bacteria. **Discussion:** Percent culled and culling rate reflect the increase in culling over time adjusting for herd size and were associated with shedding multidrug resistant bacteria. In contrast, number culled was negatively associated with shedding multidrug resistant bacteria which may reflect economic decisions that prioritize the culling of cows based on milk or beef prices (with respect to dairy beef) where a producer may elect to cull dairy cows that wouldn't be culled otherwise. Using data-driven suite of machine learning algorithms we identified generalizable and distant associations between antimicrobial resistance in *Salmonella* and fecal commensal bacteria, that can help develop a producer-friendly and data-informed risk assessment tool to reduce shedding of multidrug-resistant bacteria in cull dairy cows.

Herd management practices associated with multi-drug resistance in fecal bacterial commensals and *Salmonella* shed by cull dairy cows, a machine learning approach

Pranav S. Pandit[1], Deniece R. Williams[2], Paul Rossitto[2], John Adaska[3], Richard Pereira[4], Terry W. Lehenbauer[2,4], Barbara A. Byrne [5], Xunde Li[6], Robert Atwill[6], Sharif S. Aly[2,4]*

[1]EpiCenter for Disease Dynamics, School of Veterinary Medicine, University of California Davis, Davis, CA, USA

[2] Veterinary Medicine Teaching and Research Center, University of California, Davis, Tulare, CA, USA

[3]California Animal Health and Food Safety Laboratory System, University of California, Davis, Tulare, CA, USA

[4]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, Davis, CA, USA

[5]Department of Pathology, Microbiology & Immunology, School of Veterinary Medicine, University of California, Davis, California

[6]Western Institute for Food Safety and Security, University of California, Davis, Davis, CA, USA

*Corresponding Author: Sharif Aly; email address: saly@ucdavis.edu

Abstract

Background: Understanding the effects of herd management practices on the prevalence of multidrug-resistant pathogenic *Salmonella* and commensals *Enterococcus* spp. *and Escherichia coli* in dairy cattle is key in reducing antibacterial resistant infections in humans originating from food animals. Our objective was to explore the herd and cow level features associated with the multi-drug resistant, and resistance phenotype shared between *Salmonella*, *E. coli,* and *Enterococcus* spp. using machine learning algorithms.

Methods: Randomly collected fecal samples from cull dairy cows from six dairy farms in central California were tested for multi-drug resistance phenotypes of *Salmonella, E. coli,* and *Enterococcus* spp. Using data on herd management practices collected from a questionnaire, we built three machine learning algorithms, decision tree classifier, random forest, and gradient boosting decision trees, to predict the cows shedding multidrug-resistant *Salmonella* and commensal bacteria.

Results: The decision tree classifier identified rolling herd average milk production as an important feature for predicting fecal shedding of multi-drug resistance in *Salmonella* or commensal bacteria. The number of culled animals, monthly culling frequency and percentage, herd size, and proportion of Holstein cows in the herd were found to be influential herd characteristics predicting fecal shedding of multidrug-resistant phenotypes based on random forest models for *Salmonella* and commensal bacteria. Gradient boosting models showed that higher culling frequency and monthly culling percentages were associated with fecal shedding of multidrug resistant *Salmonella* or commensal bacteria. In contrast, an overall increase in the number of culled animals on a culling day showed a negative trend with classifying a cow as shedding multidrug-resistant bacteria. Increasing rolling herd average milk production and spring

season were positively associated with fecal shedding of multidrug-resistant *Salmonella*. Only 6 individual cows were detected sharing tetracycline resistance phenotypes between *Salmonella* and either of the commensal bacteria.

Discussion: Percent culled and culling rate reflect the increase in culling over time adjusting for herd size and were associated with shedding multidrug resistant bacteria. In contrast, number culled was negatively associated with shedding multidrug resistant bacteria which may reflect economic decisions that prioritize the culling of cows based on milk or beef prices (with respect to dairy beef) where a producer may elect to cull dairy cows that wouldn't be culled otherwise. Using data-driven suite of machine learning algorithms we identified generalizable and distant associations between antimicrobial resistance in *Salmonella* and fecal commensal bacteria, that can help develop a producer-friendly and data-informed risk assessment tool to reduce shedding of multidrug-resistant bacteria in cull dairy cows.

Introduction

The Centers for Disease Control and Prevention (CDC) estimates that more than 2.8 million antimicrobial resistant infections occur in the U.S. with more than 35,000 deaths annually (Control & Prevention 2019). Amongst the resistant bacteria, the CDC classifies nontyphoidal *Salmonella enterica* as a serious threat (Control & Prevention 2019). *Salmonella* is an important foodborne zoonotic agent in the U.S. (Scallan et al. 2011) and several studies reported on its prevalence in cull cattle. Troutt et al found that the prevalence of *Salmonella* in cecal contents of dairy cattle at slaughter in the Western U.S. ranged between 9.6% and 35.6% in the winter, and between 32.3% and 93% in the summer (Troutt et al. 2001). More recent studies reported on the associations between herd management and seasonal differences on the prevalence of *Salmonella* in fecal samples of cull dairy cows collected quarterly from seven California dairies with an overall *Salmonella* shedding prevalence of 3.42% (95% CI 1.28, 5.56) (Abu Aboud et al. 2016). Pereira et al. (Pereira et al. 2019) followed six of the same California dairies for a second year showing an increase in *Salmonella* shedding prevalence (31%; 95% CI 26.0-35.0). The latter study explored how the study herd and cow level features were associated with shedding of antimicrobial resistant *Salmonella* and although 60% of the isolates were pan-susceptible, the remaining isolates were found to be resistant to different medically important antimicrobial drugs (MIAD) defined as antimicrobial drugs (AMD) that are important for therapeutic use in humans, with 12% of the isolates being multidrug resistance to more than two drug classes.

Fecal commensals such as *E. coli* and *Enterococcus* spp. can acquire mobile gene elements that encode antimicrobial resistance to these species (Aidara-Kane et al. 2018). Given the documented antimicrobial resistance in *Salmonella* isolated from cull dairy cows, further

research on the similarity between resistance patterns of fecal commensals and *Salmonella* shed in feces of dairy cattle is needed.

Traditional risk factor approaches (mixed-effects modeling) can often have limitations due to high-dimensional, imbalanced, and non-linear data and can perform poorly in cases of large number predictive variables. To overcome these, we used a suite of classification tree models, a group of supervised machine learning models that can handle various types of data and handle interactions between predictive variables. Classification tree models like Random Forest, gradient boosting have been found useful in investigating the prevalence and associated risk factors of bovine viral diarrhea virus (Machado et al. 2015), swine pneumonia (Mollenhorst et al. 2019), and mastitis in dairy cattle (Hyde et al. 2020). In the study reported here, our objective was to explore the herd and cow level features associated with the multi-drug resistance and resistance phenotype shared between *Salmonella*, *E. coli,* or *Enterococcus* spp. using machine learning algorithms.

Materials & Methods

Farms surveys and sampling

The study was approved by the University of California, Davis's Institutional Animal Care and Use Committee (protocol number 18019). Six dairy farms located in the San Joaquin Valley of California were reenrolled as a convenience sample and followed up for a second year after being part of an earlier study (Abu Aboud et al. 2016). Briefly, cull cows were identified for fecal sampling once during each season between 2015 and 2016, specifically during summer (July 1–September 30, 2015), fall (October 1–December 31, 2015), winter (January 1–March 31, 2016) and spring (April 1–June 30, 2016). The choice of sampling week to collect fecal samples

from the cull cows during any of the four seasons was also by convenience. From the list of cows selected by the dairy farms for culling, 10 cows were randomly selected for fecal sampling on the day of their removal from the herd using a random number generator (Excel; Microsoft Corp., Redmond, WA, USA). Several lists of random numbers were prepared in multiples of 10 ranging from 11–20 to 91–100 cows and the respective list was selected depending on the number of cows presented for culling on the day of sampling. If a producer had less than 11 cows presented for culling, all the cows were sampled. An individual disposable sleeve was used to manually collect fecal samples from the rectum of each selected cow. The fecal samples were transported to the Aly Lab (Dairy Epi Lab) on wet ice for processing on the same day.

A survey was also completed with the help of the herd manager on the same sampling day. The survey questions were described in an earlier report (Pereira et al., 2019). Briefly, the questions targeted management of the herd in the previous 4 months and collected information on herd size, breed, rolling herd average, cull rate, frequency of culling per month, the proportion of cows sold for beef (compared to as dairy), proportion and reasons for condemnation of culled cows. The survey also collected information on the proportion of culled cows that received injectable medical treatments in the 3 weeks prior to culling, the role of dairy staff allowed to treat cows on the dairy, practices to avoid drug residue violations (use of specific drug types, following withdrawal times, milk and/or urine testing prior to the cow being culling, or other practices), tracking of drug withdrawal intervals, having a drug inventory system in place, and extra-label drug use (frequency and familiarity). Finally, a backup of the herd record software was obtained to collect information on the culled cows' milk production and health events. A relational database was used to house and merge data from the surveys, dairy records, and test results (Access; Microsoft Corp., Redmond, WA, USA).

130

131 **Bacteriological culture**

132 The California Animal Health and Food Safety (CAHFS) laboratory conducted all the
133 study sample testing for *Salmonella* as described by Adaska et al. (Adaska et al. 2020). Briefly,
134 1 gram of feces was inoculated into tetrathionate selective enrichment broth and incubated at 37
135 ± 2 °C. The next morning a cotton swab was used to inoculate the overnight broth onto XLD and
136 XLT-4 plates and these were incubated overnight at 37 ± 2 °C. H_2S positive, *Salmonella* suspect
137 colonies from each set of plates were subcultured onto individual bi-plates (5% sheep blood
138 agar-MacConkey agar) and incubated overnight at 37 ± 2 °C. One colony from each bi-plate was
139 used for biochemical testing which included triple sugar iron, urea, motility indole ornithine,
140 citrate, O-nitrophenyl-beta-D-galactopyranoside, and lysine iron agar slants. Serogrouping and
141 serotyping were performed, using the White–Kauffmann–Le Minor scheme (Grimont & Weill
142 2007) on colonies with biochemical test results consistent with *Salmonella* (Quinn et al. 2002)
143 Fecal samples were also cultured for *E. coli* and *Enterococcus* spp. isolation as described
144 previously (Li et al. 2014). Briefly, a 40 mL solution of buffered peptone water containing 5 g of
145 feces in a 50 mL polypropylene tube is homogenized using a mechanical shaker for 15 min
146 before filtering using gauze. 1000, 100, and 10 µl of the filtered solution were then streaked on
147 CHROMAgar ECC (Chromagar, Paris) and Enterococcus Indoxyl-β-D-Glucoside agar (mEI)
148 plates (Becton, Dickinson and Company, Franklin Lakes, NJ) both incubated at 37°C for 24 h.
149 Reference strains ATCC 25922 (*E. coli*) and ATCC 29212 (*Enterococcus faecalis*) were plated
150 on agar plates as positive controls. Two pure colonies were isolated of each species after
151 presumptive colonies were confirmed using biochemical tests (*E. coli* were confirmed using
152 urea, indole, triple sugar iron, Methyl Red–Voges–Proskauer and Simmons citrate; *Enterococcus*

153 were confirmed using bile esculin, brain heart infusion agar, and growth in broth with and
154 without 6.5% NaCl).
155
156 **Antimicrobial susceptibility testing**
157 *Salmonella* and *E. coli* bacterial resistance was evaluated with a broth microdilution
158 method using a Gram-negative Sensititre plate (CMV2AGNF) and *Enterococcus* spp. evaluated
159 on gram-positive Sensititre plate (CMV3AGPF) (Sensititre, Thermo Fisher Scientific, MA,
160 USA) according to the manufacturer's instructions and as described in previous studies (Li et al.
161 2018; Pereira et al. 2019). The MIC values were the lowest concentrations of AMD that inhibited
162 visible growth of bacteria. Interpretations of antimicrobial resistance were based on breakpoints
163 recommended by the National Antimicrobial Resistance Monitoring System
164 (https://www.cdc.gov/narms/antibiotics-tested.html; and
165 https://www.ars.usda.gov/ARSUserFiles/60400520/NARMS/ABXEntero.pdf) and the Clinical
166 and Laboratory Standards Institute (CLSI 2014; CLSI 2018). Due to the inherent resistance of
167 *Salmonella* to cephalosporins, aminoglycosides, lincosamides, oxazolidinones, and
168 glycopeptides were excluded from the antimicrobial resistance analysis. In addition, the
169 following drug classes to which *E. coli* is inherently resistant were excluded from the analysis:
170 lincosamides, oxazolidinones, penicillins, streptogramins, glycopeptides. Similarly, the drug
171 classes exclude due to the inherent resistance of *Enterococcus* spp. included cephalosporins,
172 lincosamides, fluoroquinolones, aminoglycosides, aminocyclitols, sulfonamides, folate pathway
173 antagonists. Isolates from any of the three species (*Salmonella*, *E. coli*, *Enterococcus* spp.) were
174 identified as multi-drug resistant if resistance to at least one AMD in each of three or more drug
175 classes was observed (Magiorakos et al., 2012).

Development of classification algorithms

For cows that were found positive for *Salmonella*, classification tree models were developed to test if herd management practices and features related to dairy cows can predict multi-drug resistance phenotype in *Salmonella* and fecal commensal *E. coli*, and *Enterococcus* spp. isolated from the same cows. A cow was considered to shed MDR bacteria if any of its *Salmonella, Enterococcus* spp. and/or *E. coli* isolates showed resistance for three or more antimicrobial classes (regardless of whether the species with resistance was *Salmonella, Enterococcus* spp. and/or *E. coli*), in which case the cow was labeled as 'shedding bacteria with multi-drug resistance (MDR) phenotype' (numerical label = 2). In contrast, if a cow shed bacteria that were resistant to only one or two antimicrobial classes (regardless of species) the bacteria isolated were labeled as 'shedding bacteria with antimicrobial resistance (AMR) phenotype' (numerical label = 1). If the bacteria isolated from a cow did not show any resistance across all three bacterial species to any AMD, the bacteria were labeled as non-resistant (numerical label = 0). The mutually exclusive definitions were necessary to develop a single model that predicts one of three resistance states MDR, AMR, or no resistance. Similarly, cows were also classified based on resistance phenotypes separately observed in each bacterial species isolated (three separate classification labels based on resistance phenotypes of Salmonella, *Enterococcus* spp. and *E. coli* isolates). Finally, a fifth classification was generated based on resistant phenotypes observed in commensal bacteria (*Enterococcus* spp. and *E. coli*) shed in feces collectively. Using features from herd surveys, classification tree models were trained to predict the MDR phenotype of bacteria shed in the feces of the study cows.

197 Three machine learning algorithms, specifically decision tree classifier (DTC), random
198 forest (RF), and gradient boosting (GB) were developed to explore the risk factors for resistance
199 phenotypes in the study isolates using herd survey data, specifically to predict the multilabel
200 outcome based on the resistance phenotype of bacteria shed by cows (non-resistant, AMR, and
201 MDR). For each algorithm, data from the entire study cohort described by Pereira et al. (238
202 cows) were used, except models specific to predicting AMR phenotypes in *Salmonella* were
203 restricted to the cohort of 58 *Salmonella* positive cows. Table 1 describes all classification
204 algorithms trained and developed for various bacteria-specific outcomes.
205
206 For all three classification algorithms, 25 features related to herd management and 12
207 features related to individual cows collected from the survey were used as predictive features
208 (Table 2). The DTC generates an optimum tree based on attributes by recursive selections to split
209 data into classes and it was only used to visualize the optimum tree and as a contrast to the
210 remaining classification tree models (RF and GB). The RF and GB algorithms both generate a
211 series of recursive trees of binary splits for randomly sampled predictor variables. While all tree
212 classification algorithms handle interaction effects between predictors, within GB, boosting
213 builds and combines collective models improving the predictive performance of many weak
214 models substantially, and fits complex nonlinear relationships (Elith et al. 2008). For the
215 validation of each algorithm, the data was split into training and validation datasets. To identify
216 the best hyperparameters of the classification algorithms (hyper-tuning), a grid search was
217 implemented on the training dataset. Grid search is a tuning technique that attempts to compute
218 the optimum values of model parameters by an exhaustive search performed on a set of
219 parameter values. Training datasets, composed of 80% of the data, were randomly selected for

gird search, maintaining the outcome proportional to the original dataset, with three-fold cross-validation (internal validation). Model parameters tested for hyper-tuning of RF and GB are given in Table 1. The best performing model parameters were chosen based on the accuracy of the model. For each algorithm, the external validation of the best performing model was done on the validation dataset (20% remaining random sample of the original dataset) to quantify the performance of the model on the completely independent validation dataset. The DTC and RF models were implemented using the Scikit-learn machine learning package (Pedregosa et al. 2011) and GB was implanted using XGBoost in python (Chen & Guestrin 2016).

Validated models were eventually fit on the complete dataset to produce the final predictions. The relative influence (importance) of features for the random forest model was estimated using average gini (higher the value more the relative influence of the predictor on classifying samples), permutation, and feature drop methods. The gradient boosting model with the XGboost platform was evaluated using Shapely Additive Explanations (SHAP) that assigned each predictive feature an additive feature unifying six existing methods (Lundberg & Lee 2017). Partial dependence of gradient boosted model prediction on model features (expected output response trend as a function of feature) was explored to understand the associations of the herd and cow-related features with predictions (Friedman 2001).

Results

Salmonella isolates were detected in the feces of 58 cows (24% ± 2.81 out of 238 cows) on the six dairy herds. Two herds had no *Salmonella* positive samples throughout the study period while for others the prevalence ranged from 12.5% (±5.22, n=40) to 70% (±7.24, n=40).

The most common reason reported for culling was low milk production (65.12%, ±3.08) followed by poor reproduction (31.09%, ±3.00). Lameness (10.50%, ±1.98) and mastitis (10.08%, ±1.95) were the other reasons reported in the survey. Of the sampled cull cows, 15.54% (±2.34) were reported as having received AMD as part of a treatment protocol for the condition resulting in their culling decision. In contrast, 58.82% (±1.25) received anti-inflammatory drugs for the condition resulting in their culling decision.

Distribution of antimicrobial resistance

Predominant resistance phenotypes detected

The prevalence of AMR and MDR phenotypes in *Salmonella* was reported previously by Pereira et al. (Pereira et al. 2019). Briefly, tetracyclines were the most prevalent drug class for which *Salmonella* were resistant and 12% of the study isolates tested positive for MDR *Salmonella* (Pereira et al. 2019). Within *Enterococcus* spp. isolates, the most common AMR phenotype detected was for nitrofuran antimicrobials, (10.83%, ± 2.47). The most common type of MDR phenotype in *Enterococcus* spp. isolates was resistance to oxazolidinones, nitrofuran antibacterials, and macrolides (5.73%, ±1.85). Frequencies of all resistance phenotypes observed in *Enterococcus* spp. are presented in Table 3.

Tetracyclines resistance was the most prevalent phenotype in *E. coli* isolates (30.86%, ±5.13). *E. coli* isolates from seven cows (8.64%, ±3.12), showed MDR phenotypes, and only one MDR phenotype was detected more than once (aminoglycosides, tetracyclines, amphenicols, 2.47%, ±1.72). Frequencies of all resistance phenotypes observed in *E. coli* are presented in Table 4.

Seasonality of multi-drug resistance

264 The overall prevalence of shedding MDR bacteria (*Salmonella*, *E. coli,* and *Enterococcus*
265 spp.) in cows was 30.5%, (±2.97, n = 238). The prevalence of cows shedding bacteria that are
266 resistant to <=2 antimicrobial drug classes (AMR) was 43.93% (±3.21, n = 238). The highest
267 prevalence of MDR was seen in the summer (50.00% ±11.18) followed by fall (34.0%, ±4.73).
268 The highest prevalence of shedding MDR bacteria was seen in herd 4 (52.00%, ± 7.89) and the
269 lowest was seen in herd 3 (10.00%± 4.74). Herds 1 and 2 each showed the lowest prevalence of
270 shedding AMR bacteria (37.50% ± 7.65), while herd 4 showed the highest prevalence of
271 shedding AMR (50.00%, ±7.90). The seasonal prevalence of MDR and AMR resistance across
272 three bacterial species is presented in Supplementary Figure 1.
273 The annual prevalence of cows shedding MDR *Salmonella* was 8.62% (±3.65, n = 58).
274 The highest prevalence for MDR was detected in the winter season (10.5%, ±7.04, n = 19) and
275 the highest AMR prevalence was detected in fall (42.3%±9.68, n = 26). *Salmonella* isolated from
276 cows in the summer season did not show any AMR or MDR phenotypes (Figure 1a). The annual
277 prevalence for MDR phenotypes within *E. coli* isolates was 2.92% (±1.09, n = 238). The highest
278 prevalence for MDR phenotypes in *E. coli* was detected in winter (4.1%±2.92, n = 49) and the
279 highest AMR prevalence in *E. coli* was detected in spring (34.3%±5.67, n = 70, Figure 1b). The
280 annual prevalence for MDR phenotypes within *Enterococcus* spp. isolates was 26.35% (±2.84, n
281 = 238). The highest prevalence for MDR *Enterococcus* spp. was detected in the summer season
282 (50.00%±11.18, n = 20) and the highest AMR prevalence in *Enterococcus* spp. was detected in
283 the winter season (49.0%±7.14, n =49, Figure 1c).
284

***Antimicrobial resistance phenotype shared between bacterial isolates*:**

Study found no perfectly shared resistance phenotype between the study isolates. However, 6 individual cows were detected sharing only tetracycline resistance phenotypes between *Salmonella* and commensal bacteria (*E. coli or Enterococcus* spp.). Out of these, three cows were from a single herd (herd 4) and shed MDR bacteria (Table 5). Within the commensal bacteria (*E. coli or Enterococcus* spp.), resistance to tetracycline was the most prevalent (11 cows), while shared resistance to each of the drugs kanamycin, streptomycin, and chloramphenicol was observed once.

Model performance and tuning results: Grid search of model parameters with 3-fold cross-validation yielded satisfactory results in classifying MDR cows based on herd management practices and cow-related features. The sensitivity of models ranged from 0.47 to 0.74 with precision ranging from 0.46 to 0.66. Details related to hyperparameters, best performing decision tree classifiers, random forest, and XGboost models and the performance of the selected models are given in Table 1.

Association between herd management practices and cow-related features with shedding multi-drug resistant bacteria

Decision tree classification models:

For all DTC models, the impurity, a measure of the heterogeneity of the outcome in a subset of samples resulting from a split in a decision tree, was reduced in samples the most (gini or entropy) by the rolling herd average milk production, denoting its highest position in decision trees generated by all five models. Other management features that formed nodes showing high

measures of split quality (gini or entropy) including the proportion of Jersey cows in the herd, administration of tetracyclines, monthly culling percentage, culling frequency, and the number of culled individuals. Of the five DCT models, all but the model for overall resistance across *Salmonella* and commensal bacterial species were able to generate nodes that classify cows based on their fecal shedding of bacteria to a single class of AMD, or not resistant (Supplementary figures 2-5) with 100% purity (homogenous subsets). While only the DCT model for commensal bacteria, resulted in pure nodes for MDR cows, none of the other models were able to classify cows into pure MDR nodes (Supplementary figures 2). Figure 2 shows the optimum decision tree for classifying cows into MDR, AMR, or not resistant based on phenotypes of all bacteria (*Salmonella, Enterococcus* spp*., E. coli*).

Random forest models:

Results of random forest models indicated common herd management practices that influence the shedding of MDR and AMR phenotypes of *Salmonella, Enterococcus* spp*.,* and *E. coli* separately, as well as individually (Figure 3). The number of culled animals, monthly culling frequency and percentage, herd size, and proportion of Holstein cows in the herd were found to be influential herd characteristics (top ten features by relative influence) predicting MDR phenotypes in all algorithms. Random forest algorithms for predicting AMR phenotype in *Enterococcus* spp., commensal bacterial species combined, and in all bacterial species showed the same top ten most influential herd management features. Individual-level features such as culling due to milk production, mastitis, reproductive and other reasons appeared important for predicting resistance phenotypes in *Salmonella*. The use of the chalk method for withdrawal determination was in the top ten most influential features and exclusive only for the *E. coli* model.

Gradient boosting classification models:

Herd characteristics that showed higher variable importance in SHAP values for predicting cows shedding MDR resistance bacteria based on all bacterial species were herd size, the proportion of Jersey cows, sampling season, the frequency of extra-label drug use, rolling herd average, and culling related features. These features also showed high marginal contributions in predicting AMR phenotypes. For predicting AMR phenotypes, features such as the number of monthly veterinary treatments, antibiotic dose tracking, and the proportion of Holstein cows also showed higher SHAP values (Figure 4).

Partial dependence of gradient boosting model prediction on important features showed the possible relationships of these herd and cow-related features in predicting the AMR of their fecal bacteria as MDR. Higher culling frequency and monthly culling percentages were associated with cows with MDR phenotypes from all bacteria, whereas an overall increase in the number of culled animals from the herd showed a negative trend with classifying a cow as shedding MDR bacteria. Winter season was negatively associated with MDR phenotype bacteria shed by cows compared to cows sampled in Spring. Similarly, herds with more than 10,432 kgs of rolling average milk production showed a positive trend with MDR positive cows. Cows from herds with a higher proportion of Jersey cows were negatively associated with being classified as shedding MDR bacteria by the gradient boosting algorithm (Figure 5). Within other herd characteristics that were identified as important, herd size showed a varying trend with

353 classifying cows shedding MDR bacteria, with some herd sizes showing a positive association of
354 classifying cows as shedding MDR bacteria (Figure 5).

355 For predicting MDR phenotypes in *Salmonella* shed by cows, rolling herd average milk
356 production, sampling season, chalk methods for tracking withdrawal, monthly culling frequency,
357 and Salmonella vaccine were found to be important predictive features with high SHAP values
358 (Figure 6). The exploration into partial dependence of these features gave insights into
359 relationships of feature values with classifying a cow as shedding MDR phenotype *Salmonella*
360 (Figure 7). Increasing rolling herd average milk production and monthly culling percentage was
361 positively associated with MDR phenotypic *Salmonella* in cow feces. Similarly, cows sampled in
362 spring were more likely to be classified as shedding MDR *Salmonella*.
363
364 Models predicting phenotypes in commensals *E. coli*, and *Enterococcus* spp.; the number
365 of culled animals in the previous year was always in the top ten most important features to
366 classify cows as shedding MDR bacteria (Figure S5- S8). Rolling herd average milk production
367 features describing culling practices, and herd size, consistently featured as important in
368 classifying cows as shedding MDR phenotypic bacteria for all the other three models
369 (Commensals, *E. coli,* and *Enterococcus* spp.). The frequency of extra-label drug use was an
370 important feature for models separately predicting MDR in *Enterococcus* spp. (figure S7) and *E.*
371 *coli* shed by cows.
372

373 Discussion

374 The current study investigated antimicrobial resistance phenotypes between bacteria shed in
375 dairy cattle using decision tree classification algorithms. A simple decision tree model does tend

to find the best fit for the training data, but the splitting process rarely is generalizable to other data. A random forest model, which is bagging of decision trees, and a boosting classification model, which is boosting decision trees, tend to perform better on testing data and can help us identify the generalizable conclusion. In this analysis we explored these models step by step from a simple decision tree to generalized boosting trees to find important management-related factors that might affect the distribution of multidrug resistance in dairy cattle herds.

A unique aspect of the current study is use of the aforementioned algorithms to distinguish between the resistance profiles (no resistance, antimicrobial resistance and multidrug resistance) of a pathogenic bacteria, *Salmonella*, and commensal bacteria (*E. coli* and *Enterococcus* spp.) isolates from feces of cull dairy cows. Previous investigations were restricted to understanding herd and cow level characteristics with the resistant phenotypes in *Salmonella* shed by cull dairy cows (Pereira et al. 2019). We explored if considering the resistance phenotypes for these three bacteria together, and separately, reveal any associations between herd management practices and prevalence of AMR phenotypes. Although, none of the isolates had shared phenotype resistance the six cows that had tetracycline resistance in at least two of the three bacteria studies should be explored further with whole genome sequencing and follow up studies that employ metagenomic analyses on the microbiome.

Algorithms indicated that the overall distribution of three bacterial resistance phenotypes classified in this study as MDR, AMR, or no resistance was mainly governed by resistance in *Enterococcus* spp., which showed the highest prevalence of MDR and AMR phenotypes compared to *Salmonella* and *E. coli*. The latter may be explained by *Enterococcus* spp. that are known to have frequent MDR phenotypes such as *E. faecium*. Comparative feature importance plots for all random forest models developed for these bacterial groups indicated the same, where

similar features are found to be important for the model predicting MDR in all bacteria together, in commensals together, and in *Enterococcus* spp..

Herd size has been already identified as associated with higher odds of detecting *Salmonella* resistant to tetracycline (Pereira et al. 2019). In the current study, we showed that herd size was also positively associated with detecting MDR phenotypes in *Salmonella* as well as the commensal bacteria *E. coli* and *Enterococcus* spp. Salmonellosis is known to be associated with poor production and reproduction and hence increased risk for diseases, such as mastitis and infertility, and AMD treatments which may explain the 12% MDR in *Salmonella* isolates from the study cows (Lanzas et al. 2010). Similarly, *Salmonella* has been associated with clinical disease in both adult and young dairy cattle and beef cows (Divers & Peek 2007; Pender 2003; Roy et al. 2001; Smith 2014). Percent cull and rate would reflect the increase in culling due to diseases better than actual numbers culled, this is evident from the importance of all three in the Random forest model for *Salmonella* MDR phenotypes. In contrast, the number culled on any day may reflect economic decisions that prioritize the culling of cows based on milk or beef prices (with respect to dairy beef) where a producer may elect to cull dairy cows if the price of dairy beef is favorable. It is worthy to note that the random forest algorithm identified diseases of relative importance for MDR in *Salmonella* but not commensals. However, caution should be exercised with interpreting findings from this specific analysis due to the inability to ascertain that such diseases preceded the *Salmonella* shedding and specifically MDR status. The combined analysis of all 3 species MDR random forest however did not show diseases as correlated with MDR, which may be due to the overall effect of the commensals in the dataset.

SHAP values ranked variables by importance for classifying resistance type (either MDR, AMR, or no resistance), however, in the case of RHA prediction of resistance, type for *Salmonella* and

Enterococcus was fully allocated to either MDR or AMR only with zero importance in predicting no resistance. The RHA is hence more important in terms of identifying any resistance type (MDR or AMR) versus no resistance. In contrast, for *E. coli*, RHA was important in identifying all resistance classifications (AMR, MDR, no resistance).

XGBoost results show that season was correlated with MDR in all 3 species. Specifically, that Spring and Fall had a greater correlation with MDR compared to Summer and Winter, with Winter being the least correlated. Similar findings have been observed with a risk of disease in calves in Spring and Fall with bovine respiratory disease (Cummings et al. 2019; Dubrovsky et al. 2019; Maier et al. 2019).

Conclusions

The current study characterized dairy cattle herd management practices that were associated with fecal shedding of multi-drug resistant bacteria. We identified generalizable and distant associations between pathogenic *Salmonella* and commensal bacteria. The data-driven suite of machine learning algorithms used here can help develop data-informed tools for better decision making, and risk assessment related to antibacterial resistant shedding by cows.

Acknowledgments

The authors acknowledge the study dairies' owners, herd managers, and staff from California dairies participating in the study. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

Funding

Funding for this research was provided by the U.S. Department of Agriculture Project number CA-V-PHR-4707-AH407 and CA-V-PHR-4048-H. This study was also partially supported by the USDA National Institute of Food and Agriculture (NIFA) project number 201567030238928. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

USDA Project number: CA-V-PHR-4707-AH407 and CA-V-PHR-4048-H.

USDA National Institute of Food and Agriculture (NIFA) project number: 201567030238928.

Competing interests

The authors declare that they have no competing interests.

Animal Ethics

The study was approved by the University of California, Davis's Institutional Animal Care and Use Committee (protocol number 18019).

Data Availability

The de-identified data was shared for peer review only as the dairy owners did not consent to publish it alongside the article.

References

Abu Aboud OA, Adaska JM, Williams DR, Rossitto PV, Champagne JD, Lehenbauer TW, Atwill R, Li X, and Aly SS. 2016. Epidemiology of *Salmonella* sp. in California cull dairy cattle: prevalence of fecal shedding and diagnostic accuracy of pooled enriched broth culture of fecal samples. *PeerJ* 4:e2386. 10.7717/peerj.2386

Adaska JM, Ekong PS, Clothier KA, Williams DR, Rossitto PV, Lehenbauer TW, Atwill ER, Li X, and Aly SS. 2020. Bayesian estimation of diagnostic accuracy of fecal culture and PCR-based tests for the detection of *Salmonella enterica* in California cull dairy cattle. *PeerJ* 8:e8310. 10.7717/peerj.8310

Aidara-Kane A, Angulo FJ, Conly JM, Minato Y, Silbergeld EK, McEwen SA, and Collignon PJ. 2018. World Health Organization (WHO) guidelines on use of medically important antimicrobials in food-producing animals. *Antimicrobial Resistance & Infection Control* 7:7.

Chen T, and Guestrin C. 2016. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining: ACM. p 785-794.

CLSI. 2014. *Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals: approved standard*: NCCLC.

CLSI. 2018. *M100 Performance standards for antimicrobial susceptibility testing*: Clinical and Laboratory Standards Institute Wayne, PA.

Control CfD, and Prevention. 2019. *Antibiotic resistance threats in the United States, 2019*: Centres for Disease Control and Prevention, US Department of Health and Services, CDC.

Cummings KJ, Rodriguez-Rivera LD, Capel MB, Rankin SC, and Nydam DV. 2019. Oral and intranasal administration of a modified-live *Salmonella Dublin* vaccine in dairy calves: Clinical efficacy and serologic response. *Journal of dairy science* 102:3474-3479.

Divers TJ, and Peek SF. 2007. *Rebhun's diseases of dairy cattle*: Elsevier Health Sciences.

Dubrovsky S, Van Eenennaam A, Karle B, Rossitto P, Lehenbauer TW, and Aly SS. 2019. Epidemiology of bovine respiratory disease (BRD) in preweaned calves on California dairies: The BRD 10K study. *Journal of dairy science* 102:7306-7319.

Elith J, Leathwick JR, and Hastie T. 2008. A working guide to boosted regression trees. *J Anim Ecol* 77:802-813. 10.1111/j.1365-2656.2008.01390.x

Friedman JH. 2001. Greedy function approximation: A gradient boosting machine. *Annals of statistics* 29:1189-1232. DOI 10.1214/aos/1013203451

Grimont PA, and Weill F-X. 2007. Antigenic formulae of the Salmonella serovars. *WHO collaborating centre for reference and research on Salmonella* 9:1-166.

Hyde RM, Down PM, Bradley AJ, Breen JE, Hudson C, Leach KA, and Green MJ. 2020. Automated prediction of mastitis infection patterns in dairy herds using machine learning. *Scientific reports* 10:1-8.

Lanzas C, Warnick LD, James KL, Wright EM, Wiedmann M, and Grohn YT. 2010. Transmission dynamics of a multidrug-resistant *Salmonella typhimurium* outbreak in a dairy farm. *Foodborne Pathog Dis* 7:467-474. 10.1089/fpd.2009.0411

Li X, Aly S, Su Z, Pereira R, Williams D, Rossitto P, Champagne J, Chase J, Nguyen T, and Atwill E. 2018. Phenotypic antimicrobial resistance profiles of *E. coli* and *Enterococcus* from dairy cattle in different management units on a central California dairy. *Clin Microbiol* 7:1-7.

Li X, Watanabe N, Xiao C, Harter T, McCowan B, Liu Y, and Atwill ER. 2014. Antibiotic-resistant *E. coli* in surface water and groundwater in dairy operations in Northern California. *Environmental monitoring and assessment* 186:1253-1260.

510 Lundberg SM, and Lee S-I. 2017. A unified approach to interpreting model predictions.
511 Advances in Neural Information Processing Systems. p 4765-4774.
512 Machado G, Mendoza MR, and Corbellini LG. 2015. What variables are important in predicting
513 bovine viral diarrhea virus? A random forest approach. *Veterinary research* 46:85.
514 Maier G, Love W, Karle B, Dubrovsky S, Williams D, Champagne J, Anderson R, Rowe JD,
515 Lehenbauer TW, and Van Eenennaam A. 2019. Management factors associated with
516 bovine respiratory disease in preweaned calves on California dairies: The BRD 100
517 study. *Journal of dairy science* 102:7288-7305.
518 Mollenhorst H, Ducro BJ, De Greef KH, Hulsegge I, and Kamphuis C. 2019. Boosted trees to
519 predict pneumonia, growth, and meat percentage of growing-finishing pigs. *Journal of*
520 *animal science* 97:4152-4159.
521 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer
522 P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M,
523 and Duchesnay E. 2011. Scikit-learn: Machine Learning in Python. *Journal of machine*
524 *learning research* 12:2825-2830.
525 Pender AB. 2003. Salmonellosis in a herd of beef cows. *The Canadian Veterinary Journal*
526 44:319.
527 Pereira R, Williams DR, Rossitto P, Adaska J, Okello E, Champagne J, Lehenbauer TW, Li X,
528 Chase J, Nguyen T, Pires AFA, Atwill ER, and Aly SS. 2019. Association between herd
529 management practices and antimicrobial resistance in *Salmonella spp*. from cull dairy
530 cattle in Central California. *PeerJ* 7:e6546. 10.7717/peerj.6546
531 Quinn P, Markey BK, Carter M, Donnelly W, and Leonard F. 2002. *Veterinary microbiology and*
532 *microbial disease*: Blackwell science.
533 Roy R, Higgins R, Fortin M, and Tardif S. 2001. *Salmonella* Give infection in 2 dairy herds. *The*
534 *Canadian Veterinary Journal* 42:468.
535 Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, and Griffin
536 PM. 2011. Foodborne illness acquired in the United States—major pathogens. *Emerging*
537 *infectious diseases* 17:7.
538 Smith BP. 2014. *Large animal internal medicine-E-Book*: Elsevier Health Sciences.
539 Troutt HF, Galland JC, Osburn BI, Brewer RL, Braun RK, Schmitz JA, Sears P, Childers AB,
540 Richey E, and Mather E. 2001. Prevalence of *Salmonella spp* in cull (market) dairy cows
541 at slaughter. *Journal of the American Veterinary Medical Association* 219:1212-1215.
542

Figure 1

Seasonal variation in the prevalence of multidrug antimicrobial resistance

Seasonal variation in the prevalence of multidrug antimicrobial resistance (MDR; resistance to 3 or more drug classes), and antimicrobial resistance (AMR; resistance to 1 or 2 drug classes only) in Salmonella (a), E. coli (b), and Enterococcus spp. (c) isolates from six California dairy herds. Orange and green dashed lines show the annual average prevalence of MDR and AMR in all six herds respectively. The proportion of cows that did not show any resistance are the inverse of the sum of MDR and AMR proportions and not shown in the figure. Point estimates and single standard error deviation are represented by circles and whiskers respectively.

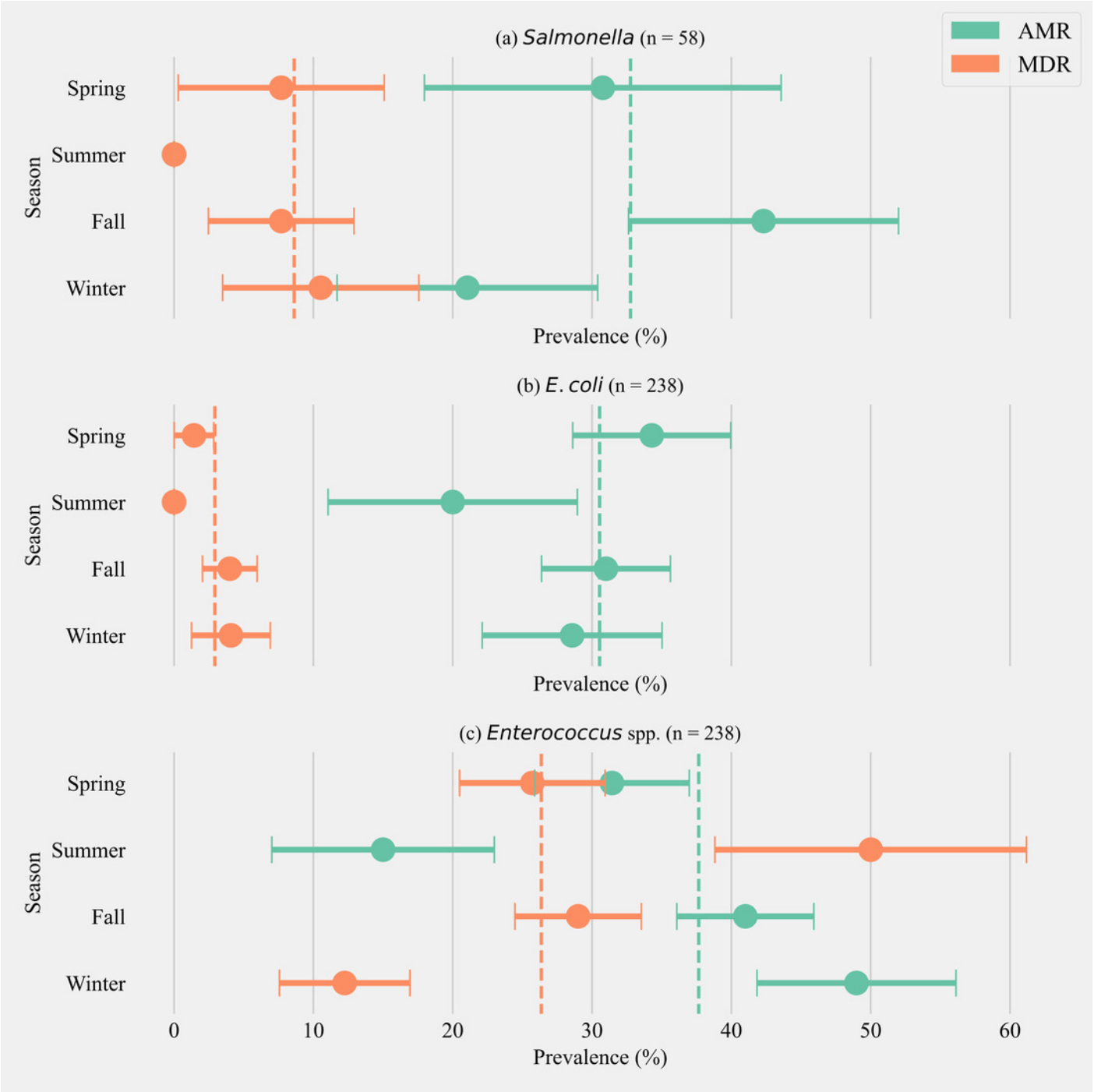
AMR
MDR
(a) Salmonella (n = 58)
Spring
Summer
Fall
Winter
Season
Prevalence (%)
(b) E. coli (n = 238)
Spring
Summer
Fall
Winter
Season
Prevalence (%)
(c) Enterococcus spp. (n = 238)
Spring
Summer
Fall
Winter
Season
0
10
20
30
40
50
60
Prevalence (%)

Figure 2

Optimum decision tree to classify cows shedding multi-drug resistant (MDR), antimicrobial-resistant (AMR), and non-resistant *Salmonella*, *Enterococcus spp.*, *E. coli* based on management practices observed in Californian dairy herds.

Node boxes describe the decision point based on management features, followed by Gini impurity at the node, and the number of samples being classified at the node into three classes (value). Left arrows indicate true for the Boolean condition while the right arrow indicates false for the Boolean condition. Factor acronym definitions are described in Table 2.

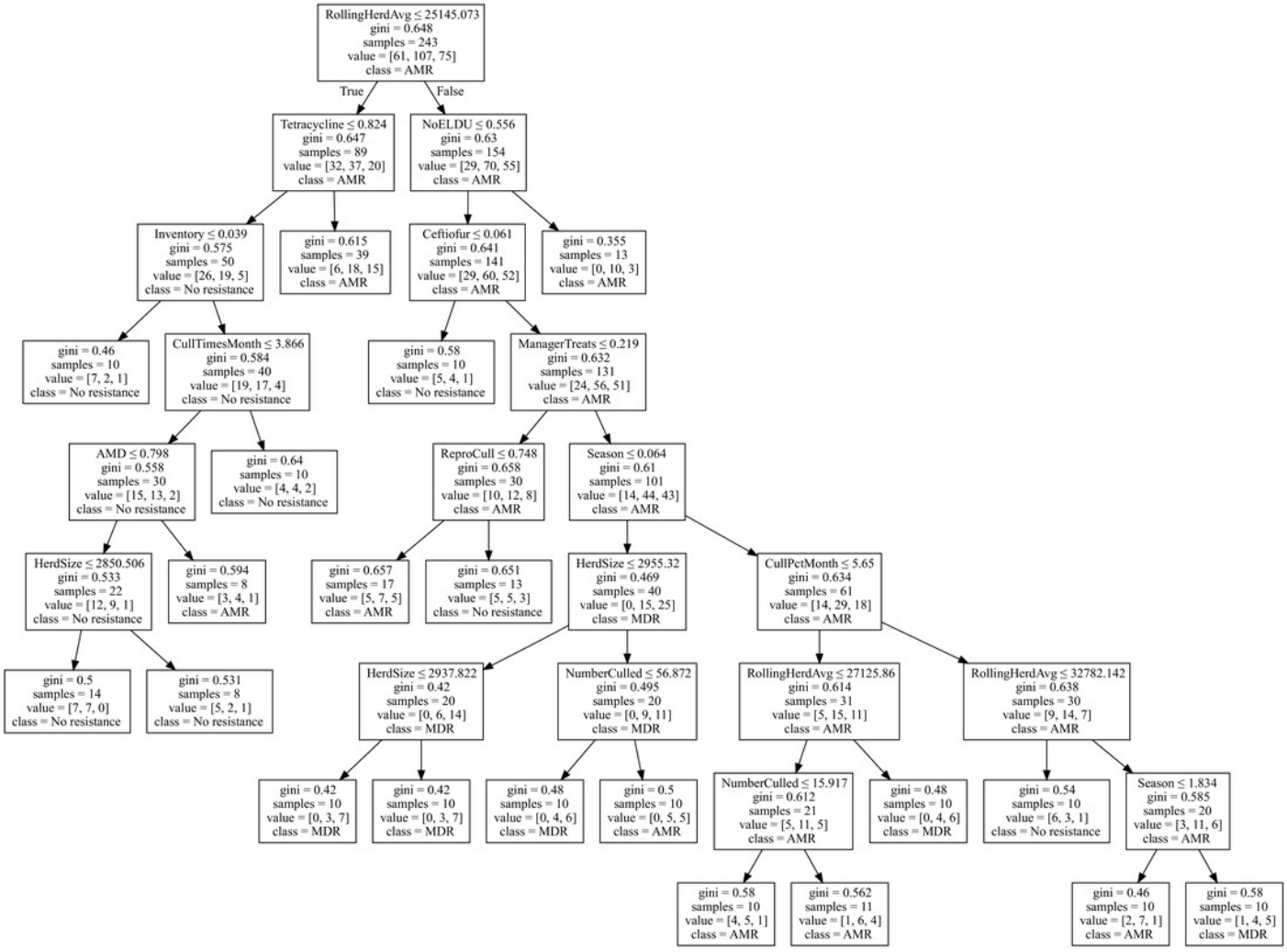

Figure 3

Top ten herd management practices based on variable importance (Gini coefficient) in classifying cows shedding multi-drug resistant (MDR), antimicrobial-resistant (AMR), and non-resistant for *Salmonella*, *Enterococcus spp.*, *E. coli* in Cal

Factor acronym definitions described in Table 2.

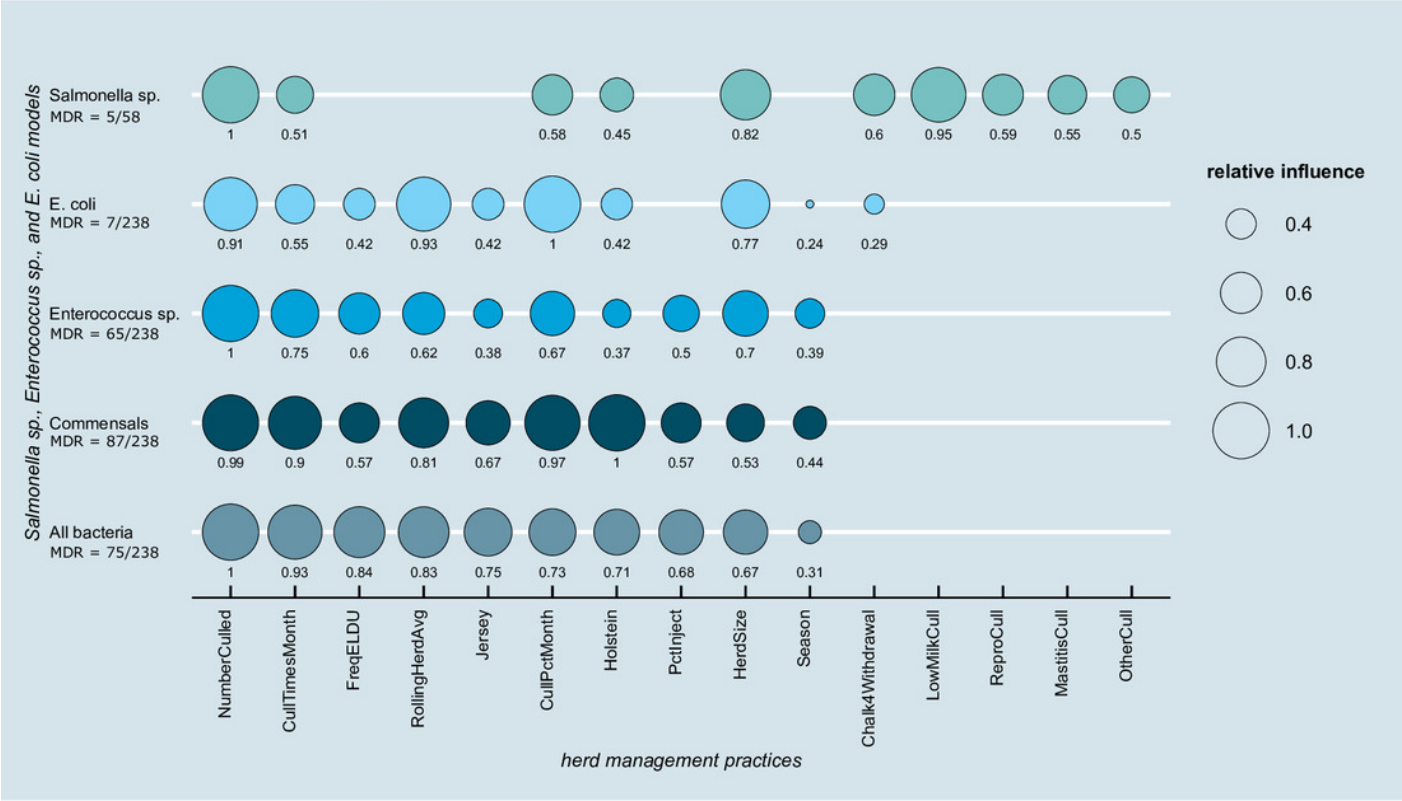

Figure 4

Mean SHAP values depicting the impact of herd management practices on predicting multi-drug resistant phenotype in either *Salmonella*, *Enterococcus spp.*, and *E. coli* shed in dairy cows for Gradient boosting classification (XGboost) model.

Factor acronym definitions described in Table 2.

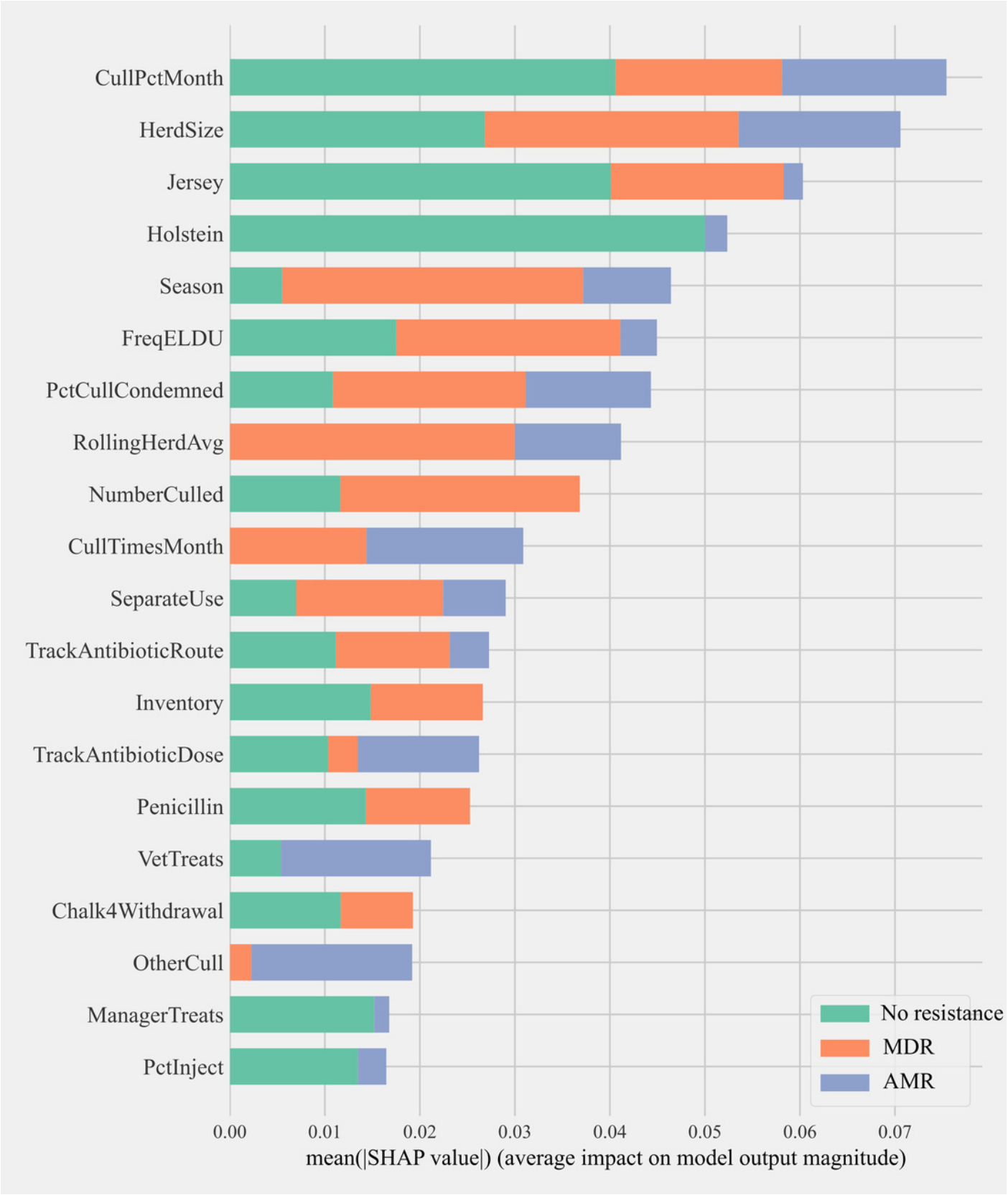
CullPctMonth
HerdSize
Jersey
Holstein
Season
FreqELDU
PctCullCondemned
RollingHerdAvg
NumberCulled
CullTimesMonth
SeparateUse
TrackAntibioticRoute
Inventory
TrackAntibioticDose
Penicillin
VetTreats
Chalk4Withdrawal
OtherCull
ManagerTreats
PctInject
No resistance
MDR
AMR
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
mean(|SHAP value|) (average impact on model output magnitude)

Figure 5

Partial dependence indicating the association of top ten predictive herd management practices in classifying cows as multi-drug resistant phenotype in either *Salmonella*, *Enterococcus spp.* and *E. coli* shed in dairy cows for Gradient boost

Partial dependence plots are generated for values presented in the data resulting in the non-linear x-axis.

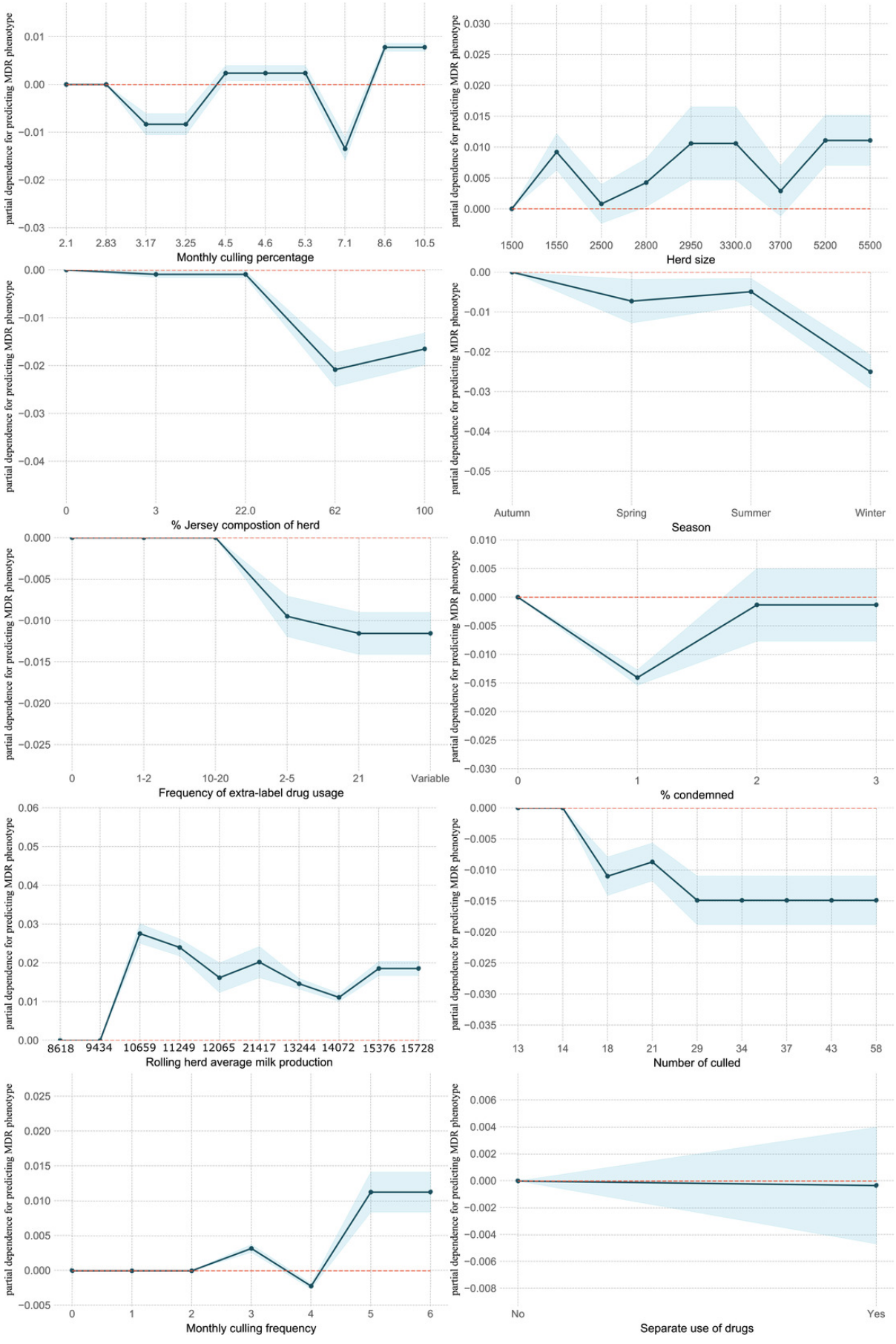
partial dependence for predicting MDR phenotype
Monthly culling percentage
Herd size
% Jersey compostion of herd
Season
Frequency of extra-label drug usage
% condemned
Rolling herd average milk production
Number of culled
Monthly culling frequency
Separate use of drugs

Figure 6

Mean SHAP values depicting the impact of herd management practices on predicting multi-drug resistant phenotype in *Salmonella* shed in dairy cows for Gradient boosting classification (XGboost) model.

Factor acronym definitions described in Table 2.

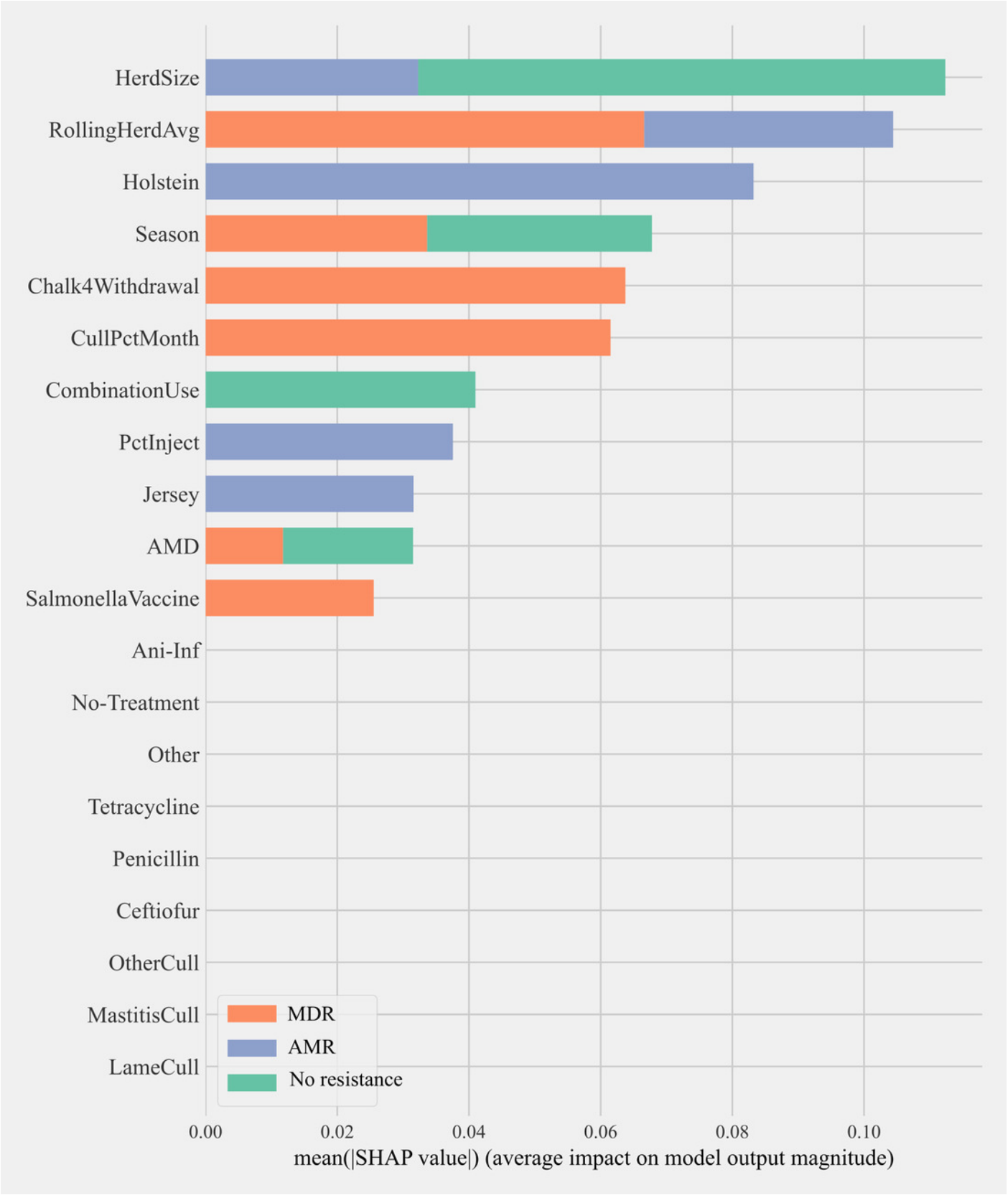
HerdSize
RollingHerdAvg
Holstein
Season
Chalk4Withdrawal
CullPctMonth
CombinationUse
PctInject
Jersey
AMD
SalmonellaVaccine
Ani-Inf
No-Treatment
Other
Tetracycline
Penicillin
Ceftiofur
OtherCull
MastitisCull
LameCull
MDR
AMR
No resistance
0.00
0.02
0.04
0.06
0.08
0.10
mean(|SHAP value|) (average impact on model output magnitude)

Figure 7

Partial dependence indicating the association of top-six predictive herd management practices in classifying cows as multi-drug resistant phenotype in *Salmonella* shed in dairy cows for Gradient boosting classification (XGboost) model.

Partial dependence plots are generated for values presented in the data resulting in the non-linear x-axis.

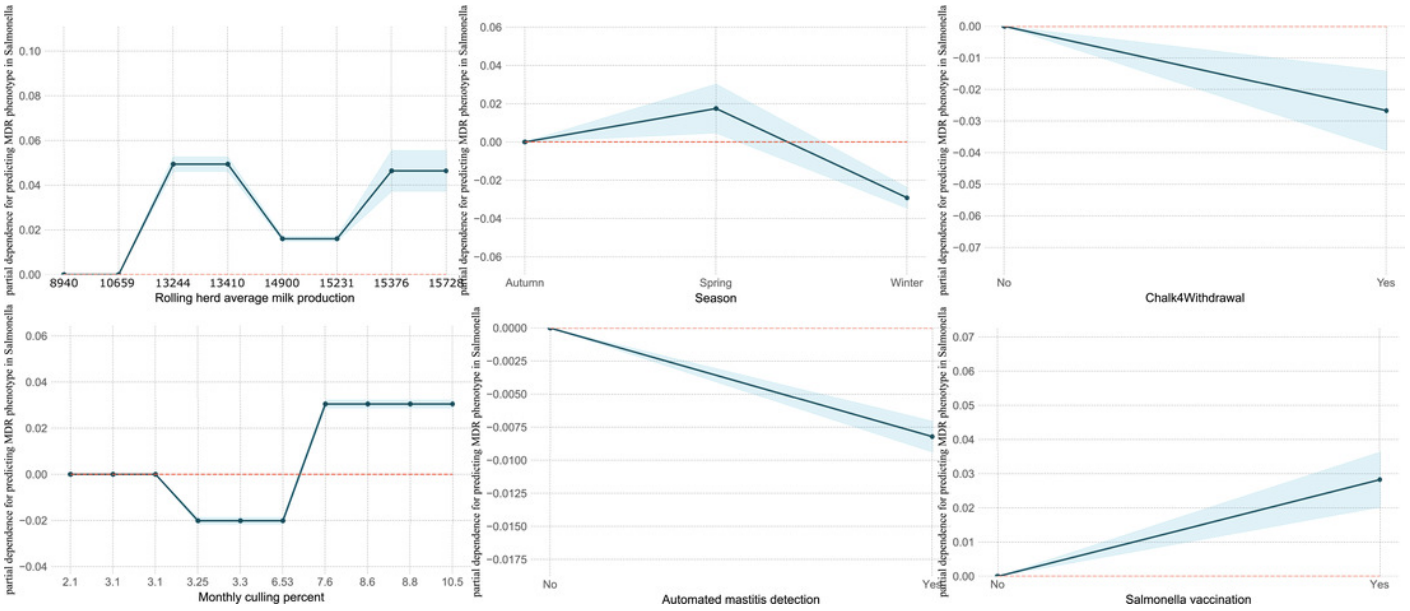

Table 1(on next page)

Hypertuning of model parameters and validation

Classification algorithms trained and tested to predict multidrug resistance phenotypes from bacterial isolates for bacterial species and groups of bacteria. Parenthesis (n) sample size of the number of cows for each model. For each bacterial and bacterial group model, hyper-tuning of the decision tree classifier, random forest, and XGBoost models is presented. The table shows parameters tuned, values tested for tuning models, best model parameters, and the performance of the selected model in terms of precision, recall and f1-score for the holdout dataset.

Table 1: Classification algorithms trained and tested to predict multidrug resistance phenotypes from bacterial isolates for bacterial species and groups of bacteria. Parenthesis (n) sample size of the number of cows for each model. For each bacterial and bacterial group model, hyper-tuning of the decision tree classifier, random forest, and XGBoost models is presented. The table shows parameters tuned, values tested for tuning models, best model parameters, and the performance of the selected model in terms of precision, recall and f1-score for the holdout dataset.

Model parameters	**Output definition based on resistance for number of antimicrobial classes**		***Salmonella*** **(n=58)**	***E. coli*** **(n=238)**	***Enterococcus*** **spp. (n=238)**	Combined resistance in *E. coli* and *Enterococcus* spp. **(n=238)**	Resistance in antimicrobial classes in either *Salmonella* or *E. coli* or *Enterococcus* spp. **(n=238)**
	Parameter explanation	**Parameter values tested**	**Best performing model parameters**				
Decision tree classifier							
Criterion	The function to measure the quality of a split	gini, entropy	gini	entropy	entropy	gini	gini
Splitter	The strategy used to choose the split at each node.	best, random	best	best	best	random	random
Maximum depth	The maximum depth of the tree	10, 20, 30, 40, 45, 50, 70	10	45	50	45	10
Minimum split	The minimum number of samples required to split an internal node	2, 3, 4, 6	6	6	6	3	3
Maximum features	The number of features to consider when looking for the best split	auto, sqrt	auto	auto	auto	sqrt	sqrt
Minimum leaf	The minimum number of samples required to be at a leaf node	1, 3, 4, 6, 7, 8	4	4	6	1	7
			Model performance for holdout dataset				
n	number of samples in holdout dataset		19	73	73	73	73
Precision	positive predictive value		0.72	0.65	0.51	0.44	0.47
Recall	sensitivity		0.79	0.68	0.52	0.44	0.47
F1-score	harmonic mean of PPV and		0.75	0.66	0.50	0.44	0.47

	sensitivity						
Random forest							
Bootstrap	Whether bootstrap samples are used when building trees	True, False	FALSE	TRUE	TRUE	TRUE	TRUE
Criterion	The function to measure the quality of a split	gini, entropy	gini	entropy	entropy	entropy	entropy
Maximum depth	The maximum depth of the tree	10, 20, 30, 40, 45, 50, 70	30	50	30	40	10
Maximum features	The number of features to consider when looking for the best split	auto, sqrt	sqrt	sqrt	sqrt	auto	auto
Minimum leaf	The minimum number of samples required to be at a leaf node	1, 3, 4, 6, 7, 8	1	8	4	8	8
Minimum split	The minimum number of samples required to split an internal node	2, 3, 4, 6	3	3	2	4	4
Number of estimators	The number of trees in the forest	100, 200, 300, 500	200	100	100	100	100
			Model performance for holdout dataset				
n	number of samples in holdout dataset		19	73	73	73	73
Precision	positive predictive value		0.66	0.62	0.46	0.47	0.41
Recall	sensitivity		0.74	0.67	0.47	0.48	0.41
F1-score	harmonic mean of PPV and sensitivity		0.69	0.61	0.46	0.47	0.38
Gradient boosting (XGBoost framework)							
Column sample	Subsample ratio of columns when constructing each tree	0.2, 0.1,0.15, 0.4, 0.7	0.1	0.2	0.2	0.4	0.1
Gamma	Minimum loss reduction required to make a further partition on a leaf node of the tree	0.0, 0.1, 0.2, 0.4, 0.45, 0.5, 0.6, 0.7	0.1	0.1	0.2	0.6	0.5
Learning rate	Boosting learning rate	0.001, 0.002, 0.005, 0.008, 0.01, 0.02, 0.05, 0.10, 0.25, 0.5	0.25	0.5	0.008	0.5	0.05
Maximum depth	Maximum tree depth for base learners	3, 4, 7, 8, 9, 10, 15, 20	3	8	8	7	7
Minimum child weight	Minimum sum of instance weight (hessian) needed in a child	1, 3, 5, 7	3	1	1	1	7
Number of estimators	Number of boosting rounds	3, 5, 10, 30, 40, 50, 100	5	30	10	3	30

Objective	Learning task, binary, multiple, etc.	multi:softprob	multi:softprob	multi:softprob	multi:softprob	multi:softprob	multi:softprob
Model performance for holdout dataset							
n	number of samples in holdout dataset		19	73	73	73	73
Precision	positive predictive value		0.61	0.59	0.51	0.49	0.46
Recall	sensitivity		0.68	0.6	0.51	0.48	0.44
F1-score	harmonic mean of PPV and sensitivity		0.65	0.59	0.51	0.48	0.38

5

Table 2(on next page)

Predictive features

Dairy cattle and herd related features used as predictors in classification models.

1 *Table 1: Dairy cattle and herd related features used as predictors in classification models*

Cow related features	Herd related features
Low milk production cull (LowMilkCull)	Milking herd size (HerdSize)
Reproduction cull (ReproCull)	Milk production level (RollingHerdAvg)
Lameness cull (LameCull)	Holstein Breed (Holstein)
Mastitis cull (MastitisCull)	Jersey Breed (Jersey)
Other reasons cull (OtherCull)	Percent culled monthly (CullPctMonth)
Antimicrobial Drug Use for cull condition (AMD)	Times culled monthly (CullTimesMonth)
Anti-inflammatory treatment for condition (Ani-Inf)	Main cull reason disease
No-Treatment for condition (No-Treatment)	Percent culled sold for beef (PctCullBeef)
Other treatment for condition (Other)	Percent carcasses condemned (PctCullCondemned)
	Percent culled injected within 2 ~ 3 weeks (PctInject)
	Veterinarian gives sick cow treatments (VetTreats)
	Dairy manager gives sick cow treatments (ManagerTreats)
	Staff gives sick cow treatments (StaffTreats)
	Prevent Residue by avoiding specific drugs (ResiduePrevent)
	Chalk on cows to track drug withdrawal (Chalk4Withdrawal)
	Keep drug inventory (Inventory)
	Penicillin
	Ceftiofur
	Tetracycline
	Antibiotics used separately (SeparateUse)
	Antibiotics combinations used (CombinationUse)
	Track antibiotic dose used (TrackAntibioticDose)
	Track antibiotic route used (TrackAntibioticRoute)
	Familiarity with ELDU (FamiliarELDU)
	Frequency of ELDU (FreqELDU)
	No ELDU (NoELDU)
	Number of cull cows culled today (NumberCulled)
	Use of *Salmonella* vaccine (*Salmonella*Vaccine)
	Sampling Season

2

3

Table 3(on next page)

Resistance phenotypes detected in *Enterococcus* spp. isolates.

1 Table 1: Resistance phenotypes detected in *Enterococcus* spp. isolates.

Resistance Phenotypes observed in *Enterococcus* spp. (Antimicrobial class)	**Number of Cows (Total N = 157)**	**The proportion of cows (%)**	**95%CI**
Nitrofuran antibacterial	17	10.83	5.967-15.689
Macrolides	15	9.55	4.956-14.152
Nitrofuran antibacterial, Macrolides	15	9.55	4.956-14.152
Oxazolidinones, Nitrofuran antibacterial, Macrolides†	9	5.73	2.096-9.369
Oxazolidinones, Nitrofuran antibacterial	8	5.1	1.656-8.535
Tetracyclines	7	4.46	1.23-7.687
Oxazolidinones	6	3.82	0.823-6.821
Tetracyclines, Nitrofuran antibacterial	6	3.82	0.823-6.821
Tetracyclines, Nitrofuran antibacterial, Macrolides†	5	3.18	0.438-5.931
Streptogramin, Oxazolidinones, Nitrofuran antibacterial, Macrolides†	4	2.55	0.083-5.013
Tetracyclines, Amphenicols, Oxazolidinones, Nitrofuran antibacterial†	4	2.55	0.083-5.013
Streptogramin, Nitrofuran antibacterial	4	2.55	0.083-5.013
Tetracyclines, Amphenicols, Nitrofuran antibacterial, Macrolides†	3	1.91	0.0-4.052
Amphenicols, Oxazolidinones, Nitrofuran antibacterial, Macrolides†	3	1.91	0.0-4.052

Tetracyclines, Macrolides	3	1.91	0.0-4.052
Amphenicols, Oxazolidinones, Nitrofuran antibacterial†	3	1.91	0.0-4.052
Oxazolidinones, Macrolides	3	1.91	0.0-4.052
Amphenicols, Streptogramin, Oxazolidinones, Nitrofuran antibacterial, Macrolides†	3	1.91	0.0-4.052
Streptogramin	2	1.27	0.0-3.028
Tetracyclines, Oxazolidinones	2	1.27	0.0-3.028
Tetracyclines, Amphenicols, Nitrofuran antibacterial†	2	1.27	0.0-3.028
Tetracyclines, Amphenicols, Macrolides, Streptogramin, Oxazolidinones, Nitrofuran antibacterial, Macrolides†	2	1.27	0.0-3.028
Amphenicols, Streptogramin, Oxazolidinones, Nitrofuran antibacterial†	2	1.27	0.0-3.028
Tetracyclines,Oxazolidinones, Nitrofuran antibacterial, Macrolides†	2	1.27	0.0-3.028
Other single isolates of MDR phenotypes**	23	14.72	9.118-20.180
Other single isolates of AMR phenotypes*	4	2.55	0.083-5.013

†Represents phenotypes that are multi-drug resistant. **Other single isolates of MDR phenotypes from *Enterococcus* spp. isolates: 1) tetracyclines, amphenicols, oxazolidinones, nitrofuran antibacterial, macrolides 2) oxazolidinones, nitrofuran antibacterial, macrolides, glycopeptides 3) tetracyclines, amphenicols, oxazolidinones, macrolides 4) streptogramin, oxazolidinones, nitrofuran antibacterial 5) tetracyclines, streptogramin, nitrofuran antibacterial, macrolides 6) tetracyclines, amphenicols, streptogramin, oxazolidinones, nitrofuran antibacterial,

macrolides 7) tetracyclines, amphenicols, macrolides, streptogramin, oxazolidinones, macrolides 8) amphenicols, macrolides, streptogramin, oxazolidinones, nitrofuran antibacterial 9) macrolides, streptogramin, oxazolidinones, nitrofuran antibacterials, macrolides 10) amphenicols, nitrofuran antibacterial, macrolides 11) oxazolidinones, nitrofuran antibacterial, glycopeptides 12) amphenicols, oxazolidinones, macrolides 13) tetracyclines, macrolides, oxazolidinones, nitrofuran antibacterial, macrolides 14) tetracyclines, streptogramin, nitrofuran antibacterial 15) tetracyclines, macrolides, nitrofuran antibacterial, macrolides 16) amphenicols, streptogramin, macrolides 17) tetracyclines, streptogramin, oxazolidinones 18) streptogramin, nitrofuran antibacterial, macrolides 19) tetracyclines, macrolides, oxazolidinones, nitrofuran antibacterial 20) tetracyclines, oxazolidinones, nitrofuran antibacterial 21) tetracyclines, amphenicols, macrolides, oxazolidinones, nitrofuran antibacterial 22) tetracyclines, amphenicols, macrolides, nitrofuran antibacterial 23) amphenicols, macrolides, nitrofuran antibacterial. *other single isolates of amr phenotypes from *enterococcus* sp. isolates: 1) amphenicols, nitrofuran antibacterial 2) streptogramin, oxazolidinones 3) amphenicols, macrolides 4) macrolides, oxazolidinone

Table 4(on next page)

Resistant phenotypes detected in *E. coli* isolates.

Table 1: Resistant phenotypes detected in E. coli isolates.

Resistant Phenotypes observed in *E. coli* (Antimicrobial class)	Number of Cows (Total N = 81)	The proportion of cows (%)	95%CI
Tetracyclines	25	30.86	20.805-40.924
Aminoglycosides	11	13.58	6.12-21.041
Cephalosporins	9	11.11	4.267-17.955
Aminoglycosides, Tetracyclines	7	8.64	2.523-14.761
Folate pathway antagonist	5	6.17	0.932-11.414
Amphenicols	5	6.17	0.932-11.414
Tetracyclines, Cephalosporins	2	2.47	0.0-5.849
Aminoglycosides, Tetracyclines, Amphenicols†	2	2.47	0.0-5.849
Other single isolates of AMR phenotypes*	5	6.17	0.932-11.374
Other single isolates of MDR phenotypes**	5	6.17	0.932-11.414

†Represents phenotypes that are multi-drug resistant. *Other single isolates of AMR phenotypes from *E. coli*: 1) quinolones, aminoglycosides 2) amphenicols, tetracyclines 3) tetracyclines, folate pathway antagonist 6) cephalosporins, fluoroquinolones 7) amphenicols, folate pathway antagonist. ** other single isolates of mdr phenotypes from *e. coli*: 1) amphenicols, folate pathway antagonist, aminoglycosides 2) amphenicols, tetracyclines, cephalosporins, aminoglycosides 3) amphenicols, tetracyclines, folate pathway antagonist 4) amphenicols, tetracyclines, folate pathway antagonist, aminoglycosides 5) tetracyclines, cephalosporins, fluoroquinolones, quinolones, aminoglycosides

Table 5(on next page)

Tetracycline antimicrobial resistance phenotype shared (highlighted in grey) between commensal bacteria (*Enterococcus spp*., *E. coli*) and *Salmonella*.

Each row represents bacterial resistance phenotypes of bacterial isolates from a single culled dairy cow.

Table 1: Antimicrobial resistance phenotype shared (highlighted in grey) between commensal bacteria (Enterococcus spp.*, E. coli) and Salmonella. Each row represents bacterial resistance phenotypes of bacterial isolates from a single culled dairy cow.*

Resistance Phenotypes observed				
Salmonella	***Enterococcus spp.***	***E. coli***	Herd ID	Resistance
Tetracyclines	Nitrofuran antibacterial	Aminoglycosides, Tetracyclines	4	AMR
Tetracyclines	Macrolides, Oxazolidinones, Nitrofuran antibacterial†	Tetracyclines, Amphenicols	4	MDR
Tetracyclines, Penicillins	Macrolides, Nitrofuran antibacterial	Aminoglycosides, Tetracyclines, Folate pathway antagonist, Amphenicols†	4	MDR
Tetracyclines, Folate pathway antagonist	Macrolides, Amphenicols	Tetracyclines	4	AMR
Tetracyclines	Tetracyclines, Nitrofuran antibacterial	Tetracyclines, Folate pathway antagonist, Amphenicols†	4	MDR
Tetracyclines	Streptogramin, Oxazolidinones	Tetracyclines	6	AMR

†Represents phenotypes that are multi-drug phenotypes