Dairy management practices associated with multi-drug resistant fecal commensals and *Salmonella* in cull cows: a machine learning approach

Pranav S Pandit [1], **Deniece R. Williams** [2], **Paul Rossitto** [2], **John M Adaska** [3], **Richard Pereira** [4], **Terry W. Lehenbauer** [2, 4], **Barbara A. Byrne** [5], **Xunde Li** [6], **Edward R 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 Author: Sharif S. Aly
Email address: 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 producer decisions to prioritize the culling of otherwise healthy but low-producing cows based on milk or beef prices (with respect to dairy beef), amongst other factors. Using a 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.

Dairy management practices associated with multi-drug resistant fecal commensals and *Salmonella* in cull cows: a machine learning approach

4 Pranav S. Pandit[1], Deniece R. Williams[2], Paul Rossitto[2], John Adaska[3], Richard Pereira[4], Terry

5 W. Lehenbauer[2,4], Barbara A. Byrne [5], Xunde Li[6], Robert Atwill[6], Sharif S. Aly[2,4]*

7 [1]EpiCenter for Disease Dynamics, School of Veterinary Medicine, University of California
8 Davis, Davis, CA, USA

10 [2] Veterinary Medicine Teaching and Research Center, University of California, Davis, Tulare,
11 CA, USA
12 [3]California Animal Health and Food Safety Laboratory System, University of California, Davis,
13 Tulare, CA, USA

15 [4]Department of Population Health and Reproduction, School of Veterinary Medicine, University
16 of California, Davis, Davis, CA, USA

18 [5]Department of Pathology, Microbiology & Immunology, School of Veterinary Medicine,
19 University of California, Davis, California

21 [6]Western Institute for Food Safety and Security, University of California, Davis, Davis, CA,
22 USA

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

25 **Abstract**

26 **Background**: Understanding the effects of herd management practices on the prevalence of

27 multidrug-resistant pathogenic *Salmonella* and commensals *Enterococcus* spp. *and Escherichia*

28 *coli* in dairy cattle is key in reducing antibacterial resistant infections in humans originating from

29 food animals. Our objective was to explore the herd and cow level features associated with the

30 multi-drug resistant, and resistance phenotype shared between *Salmonella*, *E. coli,* and

31 *Enterococcus* spp. using machine learning algorithms.

32 **Methods:** Randomly collected fecal samples from cull dairy cows from six dairy farms in central

33 California were tested for multi-drug resistance phenotypes of *Salmonella, E. coli,* and

34 *Enterococcus* spp. Using data on herd management practices collected from a questionnaire, we

35 built three machine learning algorithms (decision tree classifier, random forest, and gradient

36 boosting decision trees) to predict the cows shedding multidrug-resistant *Salmonella* and

37 commensal bacteria.

38 **Results:** The decision tree classifier identified rolling herd average milk production as an

39 important feature for predicting fecal shedding of multi-drug resistance in *Salmonella* or

40 commensal bacteria. The number of culled animals, monthly culling frequency and percentage,

41 herd size, and proportion of Holstein cows in the herd were found to be influential herd

42 characteristics predicting fecal shedding of multidrug-resistant phenotypes based on random

43 forest models for *Salmonella* and commensal bacteria. Gradient boosting models showed that

44 higher culling frequency and monthly culling percentages were associated with fecal shedding of

45 multidrug resistant *Salmonella* or commensal bacteria. In contrast, an overall increase in the

46 number of culled animals on a culling day showed a negative trend with classifying a cow as

47 shedding multidrug-resistant bacteria. Increasing rolling herd average milk production and spring

48   season were positively associated with fecal shedding of multidrug-resistant *Salmonella*. Only 6

49   individual cows were detected sharing tetracycline resistance phenotypes between *Salmonella*

50   and either of the commensal bacteria.

51   **Discussion:** Percent culled and culling rate reflect the increase in culling over time adjusting for

52   herd size and were associated with shedding multidrug resistant bacteria. In contrast, number

53   culled was negatively associated with shedding multidrug resistant bacteria which may reflect

54   producer decisions to prioritize the culling of otherwise healthy but low-producing cows based

55   on milk or beef prices (with respect to dairy beef), amongst other factors. Using a data-driven

56   suite of machine learning algorithms we identified generalizable and distant associations

57   between antimicrobial resistance in *Salmonella* and fecal commensal bacteria, that can help

58   develop a producer-friendly and data-informed risk assessment tool to reduce shedding of

59   multidrug-resistant bacteria in cull dairy cows.

Introduction

60

61 The Centers for Disease Control and Prevention (CDC) estimates that more than 2.8 million

62 antimicrobial resistant infections occur in the U.S. with more than 35,000 deaths annually

63 (Control & Prevention 2019). Amongst the resistant bacteria, the CDC classifies nontyphoidal

64 *Salmonella enterica* as a serious public health threat (Control & Prevention 2019). *Salmonella* is

65 an important foodborne zoonotic agent in the U.S. (Scallan et al. 2011) and several studies

66 reported on its prevalence in cull cattle. Troutt et al found that the prevalence of *Salmonella* in

67 cecal contents of dairy cattle at slaughter in the Western U.S. ranged between 9.6% and 35.6% in

68 the winter, and between 32.3% and 93% in the summer (Troutt et al. 2001). More recent studies

69 reported on the associations between herd management and seasonal differences on the

70 prevalence of *Salmonella* in fecal samples of cull dairy cows collected quarterly from seven

71 California dairies with an overall *Salmonella* shedding prevalence of 3.42% (95% CI 1.28, 5.56)

72 (Abu Aboud et al. 2016). Pereira et al. (Pereira et al. 2019) followed six of the same California

73 dairies for a second year showing an increase in *Salmonella* shedding prevalence (30.6%; 95%

74 CI 262-350). The increase in the prevalence was speculated to be due to increased rainfall and

75 drier summer season and herd changes that occurred during the latter study (Adaska et al. 2020).

76 The latter study explored how the study herd and cow level features were associated with

77 shedding of antimicrobial resistant *Salmonella* and although 60% of the isolates were pan-

78 susceptible, the remaining isolates were found to be resistant to different medically important

79 antimicrobial drugs (MIAD) defined as antimicrobial drugs (AMD) that are important for

80 therapeutic use in humans, with 12% of the isolates being multidrug resistance to more than two

81 drug classes.

82 Fecal commensals such as *E. coli* and *Enterococcus* spp. can acquire mobile gene elements that

83 encode antimicrobial resistance to these species (Aidara-Kane et al. 2018). Given the

84 documented antimicrobial resistance in *Salmonella* isolated from cull dairy cows, further

85 research on the similarity between resistance patterns of fecal commensals and *Salmonella* shed

86 in feces of dairy cattle is needed.

87   Traditional risk factor approaches (mixed-effects modeling) can often have limitations

88 due to high-dimensional, imbalanced, and non-linear data and can perform poorly in cases of

89 large number predictive variables. To overcome these, we used a suite of classification tree

90 models, a group of supervised machine learning models that can handle various types of data and

91 handle interactions between predictive variables. Classification tree models like random forest,

92 gradient boosting have been found useful in investigating the prevalence and associated risk

93 factors of bovine viral diarrhea virus (Machado et al. 2015), swine pneumonia (Mollenhorst et al.

94 2019), and mastitis in dairy cattle (Hyde et al. 2020). In the study reported here, our objective

95 was to explore the herd and cow level features associated with the multi-drug resistance and

96 resistance phenotype shared between *Salmonella*, *E. coli,* or *Enterococcus* spp. using machine

97 learning algorithms.

98 **Materials & Methods**

99

100 **Farms surveys and sampling**

101   The study was approved by the University of California, Davis's Institutional Animal

102 Care and Use Committee (protocol number 18019). Six dairy farms located in the San Joaquin

103 Valley of California were reenrolled as a convenience sample and followed up for a second year

104 after being part of an earlier study (Abu Aboud et al. 2016). Briefly, cull cows were identified for

105 fecal sampling once during each season between 2015 and 2016, specifically during summer

106 (July 1–September 30, 2015), fall (October 1–December 31, 2015), winter (January 1–March 31,

107 2016) and spring (April 1–June 30, 2016). The choice of sampling week to collect fecal samples

108 from the cull cows during any of the four seasons was also by convenience. From the list of cows

109 selected by the dairy farms for culling, 10 cows were randomly selected for fecal sampling on

110 the day of their removal from the herd using a random number generator (Excel; Microsoft

111 Corp., Redmond, WA, USA). Several lists of random numbers were prepared in multiples of 10

112 ranging from 11–20 to 91–100 cows and the respective list was selected depending on the

113 number of cows presented for culling on the day of sampling. If a producer had less than 11

114 cows presented for culling, all the cows were sampled. An individual disposable sleeve was used

115 to manually collect fecal samples from the rectum of each selected cow. The fecal samples were

116 transported to the Aly Lab (Dairy Epi Lab) on wet ice for processing on the same day.

117   A survey was also completed with the help of the herd manager on the same sampling

118 day. The survey questions were described in an earlier report (Pereira et al., 2019). Briefly, the

119 questions targeted management of the herd in the previous 4 months and collected information

120 on herd size, breed, rolling herd average, cull rate, frequency of culling per month, the proportion

121 of cows sold for beef (compared to as dairy), proportion and reasons for condemnation of culled

122 cows. The survey also collected information on the proportion of culled cows that received

123 injectable medical treatments in the 3 weeks prior to culling, the role of dairy staff allowed to

124 treat cows on the dairy, practices to avoid drug residue violations (use of specific drug types,

125 following withdrawal times, milk and/or urine testing prior to the cow being culling, or other

126 practices), tracking of drug withdrawal intervals, having a drug inventory system in place, and

127 extra-label drug use (ELDU, frequency and familiarity). Finally, a backup of the herd record

128   software was obtained to collect information on the culled cows' milk production and health

129   events. A relational database was used to house and merge data from the surveys, dairy records,

130   and test results (Access; Microsoft Corp., Redmond, WA, USA).

131

132   **Bacteriological culture**

133   The California Animal Health and Food Safety (CAHFS) laboratory conducted all the

134   study sample testing for *Salmonella* as described by Adaska et al. (Adaska et al. 2020). Briefly,

135   1 gram of feces was inoculated into tetrathionate selective enrichment broth and incubated at 37

136   ± 2 °C. The next morning a cotton swab was used to inoculate the overnight broth onto XLD and

137   XLT-4 plates and these were incubated overnight at 37 ± 2 °C. H_2S positive, *Salmonella* suspect

138   colonies from each set of plates were subcultured onto individual bi-plates (5% sheep blood

139   agar-MacConkey agar) and incubated overnight at 37 ± 2 °C. One colony from each bi-plate was

140   used for biochemical testing which included triple sugar iron, urea, motility indole ornithine,

141   citrate, O-nitrophenyl-beta-D-galactopyranoside, and lysine iron agar slants. Serogrouping and

142   serotyping were performed, using the White–Kauffmann–Le Minor scheme (Grimont & Weill

143   2007) on colonies with biochemical test results consistent with *Salmonella* (Quinn et al. 2002)

144   Fecal samples were also cultured for *E. coli* and *Enterococcus* spp. isolation as described

145   previously (Li et al. 2014). Briefly, a 40 mL solution of buffered peptone water containing 5 g of

146   feces in a 50 mL polypropylene tube is homogenized using a mechanical shaker for 15 min

147   before filtering using gauze. 1000, 100, and 10 µl of the filtered solution were then streaked on

148   CHROMAgar ECC (Chromagar, Paris) and Enterococcus Indoxyl-β-D-Glucoside agar (mEI)

149   plates (Becton, Dickinson and Company, Franklin Lakes, NJ) both incubated at 37°C for 24 h.

150   Reference strains ATCC 25922 (*E. coli*) and ATCC 29212 (*Enterococcus faecalis*) were plated

151   on agar plates as positive controls. Two pure colonies were isolated of each species after

152   presumptive colonies were confirmed using biochemical tests (*E. coli* were confirmed using

153   urea, indole, triple sugar iron, Methyl Red–Voges–Proskauer and Simmons citrate; *Enterococcus*

154   were confirmed using bile esculin, brain heart infusion agar, and growth in broth with and

155   without 6.5% NaCl).

156

157   **Antimicrobial susceptibility testing**

158   *Salmonella* and *E. coli* bacterial resistance was evaluated with a broth microdilution

159   method using a Gram-negative Sensititre plate (CMV2AGNF) and *Enterococcus* spp. evaluated

160   on gram-positive Sensititre plate (CMV3AGPF) (Sensititre, Thermo Fisher Scientific, MA,

161   USA) according to the manufacturer's instructions and as described in previous studies (Li et al.

162   2018; Pereira et al. 2019). The minimum inhibitory concentration (MIC) values were the lowest

163   concentrations of AMD that inhibited visible growth of bacteria. Interpretations of antimicrobial

164   resistance were based on breakpoints recommended by the National Antimicrobial Resistance

165   Monitoring System (https://www.cdc.gov/narms/antibiotics-tested.html; and

166   https://www.ars.usda.gov/ARSUserFiles/60400520/NARMS/ABXEntero.pdf) and the Clinical

167   and Laboratory Standards Institute (CLSI 2014; CLSI 2018). Due to the inherent resistance of

168   *Salmonella* to cephalosporins, aminoglycosides, lincosamides, oxazolidinones, and

169   glycopeptides were excluded from the antimicrobial resistance analysis. In addition, the

170   following drug classes to which *E. coli* is inherently resistant were excluded from the analysis:

171   lincosamides, oxazolidinones, penicillins, streptogramins, glycopeptides. Similarly, the drug

172   classes exclude due to the inherent resistance of *Enterococcus* spp. included cephalosporins,

173   lincosamides, fluoroquinolones, aminoglycosides, aminocyclitols, sulfonamides, folate pathway

174 antagonists. Isolates from any of the three species (*Salmonella*, *E. coli*, *Enterococcus* spp.) were

175 identified as multi-drug resistant if resistance to at least one AMD in each of three or more drug

176 classes was observed (Magiorakos et al., 2012).

Development of classification algorithms

178 For cows that were found positive for *Salmonella*, classification tree models were

179 developed to test if herd management practices and features related to dairy cows can predict

180 multi-drug resistance phenotype in *Salmonella* and fecal commensal *E. coli*, and *Enterococcus*

181 spp. isolated from the same cows. A cow was considered to shed MDR bacteria if any of its

182 *Salmonella, Enterococcus* spp. and/or *E. coli* isolates showed resistance for three or more

183 antimicrobial classes (regardless of whether the species with resistance was *Salmonella,*

184 *Enterococcus* spp. and/or *E. coli*), in which case the cow was labeled as 'shedding bacteria with

185 multi-drug resistance (MDR) phenotype' (numerical label = 2). In contrast, if a cow shed

186 bacteria that were resistant to only one or two antimicrobial classes (regardless of species) the

187 bacteria isolated were labeled as 'shedding bacteria with antimicrobial resistance (AMR)

188 phenotype' (numerical label = 1). If the bacteria isolated from a cow did not show any resistance

189 across all three bacterial species to any AMD, the bacteria were labeled as non-resistant

190 (numerical label = 0). The mutually exclusive definitions were necessary to develop a single

191 model that predicts one of three resistance states MDR, AMR, or no resistance. Similarly, cows

192 were also classified based on resistance phenotypes separately observed in each bacterial species

193 isolated (three separate classification labels based on resistance phenotypes of Salmonella,

194 *Enterococcus* spp. and *E. coli* isolates). Finally, a classification was generated based on resistant

195 phenotypes observed in either commensal bacteria (*Enterococcus* spp. and *E. coli*) shed in feces

196   collectively. Using features from herd surveys, classification tree models were trained to predict

197   the MDR phenotype of bacteria shed in the feces of the study cows.

198   Three machine learning algorithms, specifically decision tree classifier (DTC), random

199   forest (RF), and gradient boosting (GB) were developed to explore the risk factors for resistance

200   phenotypes in the study isolates using herd survey data, specifically to predict the multilabel

201   outcome based on the resistance phenotype of bacteria shed by cows (non-resistant, AMR, and

202   MDR). For each algorithm, data from the entire study cohort described by Pereira et al. (238

203   cows) were used, except models specific to predicting AMR phenotypes in *Salmonella* were

204   restricted to the cohort of 58 *Salmonella* positive cows. Table 1 describes all classification

205   algorithms trained and developed for various bacteria-specific outcomes.

206

207   For all three classification algorithms, 25 features related to herd management and 12

208   features related to individual cows collected from the survey were used as predictive features

209   (Table 2). The DTC generates an optimum tree based on attributes by recursive selections to split

210   data into classes and it was only used to visualize the optimum tree and as a contrast to the

211   remaining classification tree models (RF and GB) that prevent overfitting, unlike DTC. The RF

212   and GB algorithms both generate a series of recursive trees of binary splits for randomly sampled

213   predictor variables. While all tree classification algorithms handle interaction effects between

214   predictors, within GB, boosting builds and combines collective models improving the predictive

215   performance of many weak models substantially, and fits complex nonlinear relationships (Elith

216   et al. 2008). For the validation of each algorithm, the data was split into training and validation

217   datasets. To identify the best hyperparameters of the classification algorithms (hyper-tuning), a

218   grid search was implemented on the training dataset. Grid search is a tuning technique that

219   computes optimum values of model parameters automatically by an exhaustive search performed

220   on a set of parameter values. Training datasets, composed of 80% of the data, were randomly

221   selected for gird search, maintaining the outcome proportional to the original dataset, with three-

222   fold cross-validation (internal validation). Model parameters tested for hyper-tuning of RF and

223   GB are given in Table 1. The best performing model parameters were chosen based on the

224   accuracy of the model. For each algorithm, the external validation of the best performing model

225   was done on the validation dataset (20% remaining random sample of the original dataset) to

226   quantify the performance of the model on the completely independent validation dataset. The

227   DTC and RF models were implemented using the Scikit-learn machine learning package

228   (Pedregosa et al. 2011) and GB was implanted using XGBoost in python (Chen & Guestrin

229   2016).

230   Validated models were eventually fit on the complete dataset to produce the final

231   predictions. The relative influence (importance) of features for the random forest model was

232   estimated using average gini importance, permutation, and feature drop methods. The gini

233   importance for a feature is defined as the sum over the number of splits (across all tress) that

234   include the feature, proportionally to the number of samples it splits (Breiman 2001). The

235   gradient boosting model with the XGboost platform was evaluated using Shapely Additive

236   Explanations (SHAP) that assigned each predictive feature an additive feature unifying six

237   existing methods (Lundberg & Lee 2017). Partial dependence of gradient boosted model

238   prediction on model features (expected output response trend as a function of feature) was

239   explored to understand the associations of the herd and cow-related features with predictions

240   (Friedman 2001). The Python code used for pre-processing the data, training and validating

241   models and generating figures can be found here in the Zenodo repository,

242   https://doi.org/10.5281/zenodo.4387017.

243

244   Results

245   *Salmonella* isolates were detected in the feces of 58 cows (24.4% ± 2.8 out of 238 cows)

246   on the six dairy herds. Two herds had no *Salmonella* positive samples throughout the study

247   period while for others the prevalence ranged from 12.5% (±5.2, n=40) to 70.0% (±7.2, n=40).

248   The most common reason reported for culling was low milk production (65.1%, ±3.1)

249   followed by poor reproduction (31.1%, ±3.0). Lameness (10.5%, ±1.9) and mastitis (10.1%,

250   ±1.9) were the other reasons reported in the survey. Of the sampled cull cows, 15.5% (±2.3)

251   were reported as having received AMD as part of a treatment protocol for the condition resulting

252   in their culling decision. In contrast, 5.9% (±1.5) received anti-inflammatory drugs for the

253   condition resulting in their culling decision.

254   **Distribution of antimicrobial resistance**

255   ***Predominant resistance phenotypes detected***

256   The prevalence of AMR and MDR phenotypes in *Salmonella* was reported previously by

257   Pereira et al. (Pereira et al. 2019). Briefly, tetracyclines were the most prevalent drug class for

258   which *Salmonella* were resistant and 12% of the study isolates tested positive for MDR

259   *Salmonella* (Pereira et al. 2019). Within *Enterococcus* spp. isolates, the most common AMR

260   phenotype detected was for nitrofuran antimicrobials, (10.83%, ± 2.47). The most common type

261   of MDR phenotype in *Enterococcus* spp. isolates was resistance to oxazolidinones, nitrofuran

262   antibacterials, and macrolides (5.73%, ±1.85). Frequencies of all resistance phenotypes observed

263   in *Enterococcus* spp. are presented in Table 3.

264   Tetracyclines resistance was the most prevalent phenotype in *E. coli* isolates (30.86%,

265   ±5.13). *E. coli* isolates from seven cows (8.64%, ±3.12), showed MDR phenotypes, and only one

266   MDR phenotype was detected more than once (aminoglycosides, tetracyclines, amphenicols,

267   2.47%, ±1.72). Frequencies of all resistance phenotypes observed in *E. coli* are presented in

268   Table 4.

269   ***Seasonality of multi-drug resistance***

270   The overall prevalence of shedding MDR bacteria (*Salmonella*, *E. coli,* and *Enterococcus*

271   spp.) in cows was 30.54%, (±2.97, n = 238). The prevalence of cows shedding bacteria that are

272   resistant to <=2 antimicrobial drug classes (AMR) was 43.93% (±3.21, n = 238). The highest

273   prevalence of MDR was seen in the summer (50.00% ±11.18) followed by fall (34.00%, ±4.73).

274   The highest prevalence of shedding MDR bacteria was seen in herd 4 (52.00%, ± 7.89) and the

275   lowest was seen in herd 3 (10.00%± 4.74). Herds 1 and 2 each showed the lowest prevalence of

276   shedding AMR bacteria (37.50% ± 7.65), while herd 4 showed the highest prevalence of

277   shedding AMR (50.00%, ±7.90). The seasonal prevalence of MDR and AMR resistance across

278   three bacterial species is presented in Supplementary Figure 1.

279   The annual prevalence of cows shedding MDR *Salmonella* was 8.62% (±3.65, n = 58).

280   The highest prevalence for MDR was detected in the winter season (10.53%, ±7.04, n = 19) and

281   the highest AMR prevalence was detected in fall (42.31%±9.68, n = 26). *Salmonella* isolated

282   from cows in the summer season did not show any AMR or MDR phenotypes (Figure 1a). The

283   annual prevalence for MDR phenotypes within *E. coli* isolates was 2.92% (±1.09, n = 238). The

284   highest prevalence for MDR phenotypes in *E. coli* was detected in winter (4.08%±2.82, n = 49)

285   and the highest AMR prevalence in *E. coli* was detected in spring (34.28%±5.67, n = 70, Figure

286   1b). The annual prevalence for MDR phenotypes within *Enterococcus* spp. isolates was 26.35%

287   (±2.84, n = 238). The highest prevalence for MDR *Enterococcus* spp. was detected in the

288   summer season (50.00%±11.18, n = 20) and the highest AMR prevalence in *Enterococcus* spp.

289   was detected in the winter season (48.97%±7.14, n =49, Figure 1c).

290

291   *Antimicrobial resistance phenotype shared between bacterial isolates*:

292   Study found no perfectly shared resistance phenotype between the study isolates. Within the

293   commensal bacteria (*E. coli or Enterococcus* spp.), resistance to tetracycline was the most

294   prevalent (11 cows), while shared resistance to each of the drugs kanamycin, streptomycin, and

295   chloramphenicol was observed once. However, 6 individual cows were detected sharing

296   tetracycline resistance phenotypes between *Salmonella* and one or both commensal bacteria (*E.*

297   *coli or Enterococcus* spp.). Of the 6 cows, five were from a single herd (herd 4) and three shed

298   MDR bacteria (Table 5). However, Salmonella, and one or both commensal species, were found

299   to share resistance to only 1 drug, tetracycline, in 6 cows across the study. Of the 6 cows, only 3

300   had bacterial species with MDR phenotypes detected and in 2 of these cows the MDR profile

301   included tetracycline.

302

303   **Model performance and tuning results:** Grid search of model parameters with 3-fold cross-

304   validation yielded satisfactory results in classifying MDR cows based on herd management

305   practices and cow-related features. The sensitivity of models ranged from 0.47 to 0.74 with

306   precision ranging from 0.46 to 0.66. Details related to hyperparameters, best performing decision

307   tree classifiers, random forest, and XGboost models and the performance of the selected models

308   are given in Table 1.

309

310   **Association between herd management practices and cow-related features with shedding**

311   **multi-drug resistant bacteria**

312   *Decision tree classification models:*

313   For all DTC models, the impurity (gini), a measure of the heterogeneity of the outcome in

314   a subset of samples resulting from a split in a decision tree, was reduced in samples the most by

315   the rolling herd average milk production, denoting its highest position in decision trees generated

316   by all five models (Figure1, Figure S2-S5). Other management features that formed nodes

317   showing high measures of split quality (gini) including the proportion of Jersey cows in the herd,

318   administration of tetracyclines, monthly culling percentage, culling frequency, and the number of

319   culled individuals. Of the five DCT models, all but the model for overall resistance across

320   *Salmonella* and commensal bacterial species were able to generate nodes that classify cows

321   based on their fecal shedding of bacteria to a single class of AMD, or not resistant

322   (Supplementary figures 2-5) with 100% purity (homogenous subsets). While only the DCT

323   model for commensal bacteria, resulted in pure nodes for MDR cows, none of the other models

324   were able to classify cows into pure MDR nodes (Supplementary figures 2). Figure 2 shows the

325   optimum decision tree for classifying cows into MDR, AMR, or not resistant based on

326   phenotypes of all bacteria (*Salmonella, Enterococcus* spp*., E. coli*).

327   ***Random forest models:***

328   Results of random forest models indicated common herd management practices that

329   influence the shedding of MDR and AMR phenotypes of *Salmonella, Enterococcus* spp*., and E.*

330   *coli* collectively, as well as individually (Figure 3). The number of culled animals, monthly

331   culling frequency and percentage, herd size, and proportion of Holstein cows in the herd were

332   found to be influential herd characteristics (top ten features by relative influence) predicting

333   MDR phenotypes in all algorithms. Random forest algorithms for predicting AMR phenotype in

334   *Enterococcus* spp., commensal bacterial species combined, and in all bacterial species showed

335   the same top ten most influential herd management features. Individual-level features such as

336 culling due to milk production, mastitis, reproductive and other reasons appeared important for

337 predicting resistance phenotypes in *Salmonella*. The use of the chalk method for withdrawal

338 determination was in the top ten most influential features and for the *E. coli* and *Salmonella*

339 models.

340

341 ***Gradient boosting classification models:***

342   Herd characteristics that showed higher variable importance in SHAP values for

343 predicting cows shedding MDR resistance bacteria based on all bacterial species were herd size,

344 the proportion of Jersey cows, sampling season, the frequency of extra-label drug use, rolling

345 herd average, and culling related features. These features also showed high marginal

346 contributions in predicting AMR phenotypes. For predicting AMR phenotypes, features such as

347 the number of monthly veterinary treatments, antibiotic dose tracking, and the proportion of

348 Holstein cows also showed higher SHAP values (Figure 4).

349

350 Partial dependence of gradient boosting model prediction on important features showed the

351 possible relationships of these herd and cow-related features in predicting the AMR of their fecal

352 bacteria as MDR. Higher culling frequency and monthly culling percentages were associated

353 with cows with MDR phenotypes from all bacteria, whereas an overall increase in the number of

354 culled animals from the herd showed a negative trend with classifying a cow as shedding MDR

355 bacteria. Winter season was negatively associated with MDR phenotype bacteria shed by cows

356 compared to cows sampled in Spring. Similarly, herds with more than 10,432 kgs of rolling

357 average milk production showed a positive trend with MDR positive cows. Cows from herds

358 with a higher proportion of Jersey cows were negatively associated with being classified as

359   shedding MDR bacteria by the gradient boosting algorithm (Figure 5). Within other herd

360   characteristics that were identified as important, herd size showed a varying trend with

361   classifying cows shedding MDR bacteria, with some herd sizes showing a positive association of

362   classifying cows as shedding MDR bacteria (Figure 5).

363   For predicting MDR phenotypes in *Salmonella* shed by cows, rolling herd average milk

364   production, sampling season, chalk methods for tracking withdrawal, monthly culling frequency,

365   and Salmonella vaccine were found to be important predictive features with high SHAP values

366   (Figure 6). The exploration into partial dependence of these features gave insights into

367   relationships of feature values with classifying a cow as shedding MDR phenotype *Salmonella*

368   (Figure 7). Increasing rolling herd average milk production and monthly culling percentage was

369   positively associated with MDR phenotypic *Salmonella* in cow feces. Similarly, cows sampled in

370   spring were more likely to be classified as shedding MDR *Salmonella*.

371

372   Models predicting phenotypes in commensals *E. coli*, and *Enterococcus* spp.; the number

373   of culled animals in the previous year was always in the top ten most important features to

374   classify cows as shedding MDR bacteria (Figure S5- S8). Rolling herd average milk production

375   features describing culling practices, and herd size, consistently featured as important in

376   classifying cows as shedding MDR phenotypic bacteria for all the other three models

377   (Commensals, *E. coli,* and *Enterococcus* spp.). The frequency of extra-label drug use was an

378   important feature for models separately predicting MDR in *Enterococcus* spp. (figure S7) and *E.*

379   *coli* shed by cows.

380

381   **Discussion**

382   The current study investigated antimicrobial resistance phenotypes between bacteria shed in

383   dairy cattle using decision tree classification algorithms. A simple decision tree model does tend

384   to find the best fit for the training data, but the splitting process rarely is generalizable to other

385   data. A random forest model, which is bagging of decision trees, and a boosting classification

386   model, which is boosting decision trees, tend to perform better on testing data and can help us

387   identify the generalizable conclusion. In this analysis we explored these models step by step

388   from a simple decision tree to generalized boosting trees to find important management-related

389   factors that might affect the distribution of multidrug resistance in dairy cattle herds.

390   A unique aspect of the current study is use of the aforementioned algorithms to distinguish

391   between the resistance profiles (no resistance, antimicrobial resistance and multidrug resistance)

392   of a pathogenic bacteria, *Salmonella*, and commensal bacteria (*E. coli* and *Enterococcus* spp.)

393   isolates from feces of cull dairy cows. Previous investigations were restricted to understanding

394   herd and cow level characteristics with the resistant phenotypes in *Salmonella* shed by cull dairy

395   cows (Pereira et al. 2019). We explored weather considering the resistance phenotypes for these

396   three bacterial species together, and separately, to identify associations between herd

397   management practices and prevalence of resistance phenotypes. Although, none of the isolates

398   had shared phenotype resistance the six cows that had tetracycline resistance in at least two of

399   the three bacteria studies should be explored further with whole genome sequencing and follow

400   up studies that employ metagenomic analyses on the microbiome.

401   The three machine learning algorithms tested in this study indicated that the overall distribution

402   of three resistance phenotypes classified in this study as MDR, AMR, or no resistance was

403   mainly governed by resistance in *Enterococcus* spp., which showed the highest prevalence of

404   MDR and AMR phenotypes compared to *Salmonella* and *E. coli*. The latter may be explained by

405 *Enterococcus* spp. that are known to have frequent MDR phenotypes such as *E. faecium*.

406 Comparative feature importance plots for all random forest models developed for these bacterial

407 groups indicated the same, where similar features are found to be important for the model

408 predicting MDR in all bacteria together, in commensals together, and in *Enterococcus* spp..

409 Herd size has been already identified as associated with higher odds of detecting *Salmonella*

410 resistant to tetracycline (Pereira et al. 2019). In the current study, we showed that herd size was

411 also positively associated with detecting MDR phenotypes in *Salmonella* as well as the

412 commensal bacteria *E. coli* and *Enterococcus* spp. Salmonellosis is known to be associated with

413 poor milk production and reproduction and hence increased risk for diseases, such as mastitis

414 and infertility, and AMD treatments which may explain the 12% MDR in *Salmonella* isolates

415 from the study cows (Lanzas et al. 2010). Similarly, *Salmonella* has been associated with clinical

416 disease in both adult and young dairy cattle and beef cows (Divers & Peek 2007; Pender 2003;

417 Roy et al. 2001; Smith 2014). Percent cull and rate would reflect the increase in culling due to

418 diseases better than actual numbers culled. This is evident from the importance of all three in the

419 random forest model for *Salmonella* MDR phenotypes. In contrast, the total number culled was

420 negatively associated with shedding multidrug resistant bacteria which may reflect producer

421 decisions to prioritize the culling of otherwise healthy but low-producing cows based on milk or

422 beef market value (with respect to dairy beef), amongst other factors. Other than market

423 dynamics including milk or beef demand, underlying herd health or management reasons may

424 explain the opposing trends of both percent and number culled, and the outcome MDR in cull

425 cow fecal bacteria. Herds with MDR in their cull cow fecal bacteria hence need to be explored

426 further to identify mechanisms that eventually increase culling rates. It is worthy to note that the

427 random forest algorithm identified diseases of relative importance for MDR in *Salmonella* but

428   not commensals. However, caution should be exercised with interpreting findings from this

429   specific analysis due to the inability to ascertain that such diseases preceded the *Salmonella*

430   shedding and specifically MDR status. The combined analysis of all 3 species MDR random

431   forest however did not show diseases as correlated with MDR, which may be due to the overall

432   effect of the commensals in the dataset.

433   SHAP values ranked variables by importance for classifying resistance type (either MDR, AMR,

434   or no resistance). However, in the case of rolling herd average milk production (RHA), SHAP

435   values were allocated only for predicting MDR or AMR in models for *Salmonella* and

436   *Enterococcus* and commensals, showing zero importance in predicting absence of any resistance.

437   The RHA is hence more important in terms of identifying any resistance type (MDR or AMR)

438   versus no resistance. All three algorithms (DCT, RF and XGboost) indicated a high importance

439   of RHA in predicting MDR in fecal bacteria of cull cows. RHA is a dairy performance indicator

440   affected by multiple herd characteristics such as age and breed structure. Studies have indicated

441   stress related to higher production which may increase the chance of certain health conditions

442   subsequently increasing the risk of antimicrobial drug use and hence bacterial resistance

443   (Robbins et al. 2016). In addition, the association between RHA and MDR presented by all the

444   models here is conditioned upon other features that follow the splits further down the

445   classification tree. In contrast, for *E. coli*, RHA was important in identifying all resistance

446   classifications (AMR, MDR, no resistance).

447   XGBoost results show that season was correlated with MDR in all 3 species. Specifically, that

448   Spring and Fall had a greater correlation with MDR compared to Summer and Winter, with

449   Winter being the least correlated. Similar findings have been observed with a risk of disease in

450   calves in Spring and Fall with bovine respiratory disease (Cummings et al. 2019; Dubrovsky et

451   al. 2019; Maier et al. 2019). The current study's bacterial species, specifically Salmonella, *E. coli*

452   and *Enterococcus* spp., shared no specific MDR profiles; however, shared tetracycline resistance

453   was detected. Further studies employing whole genome sequencing and metagenomics on the

454   microbiome may explore factors that explain such shared resistance.

455

456

457   **Conclusions**

458   The current study characterized dairy cattle herd management practices that were associated with

459   fecal shedding of multi-drug resistant bacteria. We identified generalizable and distant

460   associations between pathogenic *Salmonella* and commensal bacteria. The data-driven suite of

461   machine learning algorithms used here can help develop data-informed tools for better decision

462   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.

Breiman L. 2001. Random forests. *Machine learning* 45:5-32.

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.

533   Li X, Watanabe N, Xiao C, Harter T, McCowan B, Liu Y, and Atwill ER. 2014. Antibiotic-resistant
534   E. coli in surface water and groundwater in dairy operations in Northern California.
535   *Environmental monitoring and assessment* 186:1253-1260.
536   Lundberg SM, and Lee S-I. 2017. A unified approach to interpreting model predictions.
537   Advances in Neural Information Processing Systems. p 4765-4774.
538   Machado G, Mendoza MR, and Corbellini LG. 2015. What variables are important in predicting
539   bovine viral diarrhea virus? A random forest approach. *Veterinary research* 46:85.
540   Maier G, Love W, Karle B, Dubrovsky S, Williams D, Champagne J, Anderson R, Rowe JD,
541   Lehenbauer TW, and Van Eenennaam A. 2019. Management factors associated with
542   bovine respiratory disease in preweaned calves on California dairies: The BRD 100
543   study. *Journal of dairy science* 102:7288-7305.
544   Mollenhorst H, Ducro BJ, De Greef KH, Hulsegge I, and Kamphuis C. 2019. Boosted trees to
545   predict pneumonia, growth, and meat percentage of growing-finishing pigs. *Journal of*
546   *animal science* 97:4152-4159.
547   Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer
548   P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M,
549   and Duchesnay E. 2011. Scikit-learn: Machine Learning in Python. *Journal of machine*
550   *learning research* 12:2825-2830.
551   Pender AB. 2003. Salmonellosis in a herd of beef cows. *The Canadian Veterinary Journal*
552   44:319.
553   Pereira R, Williams DR, Rossitto P, Adaska J, Okello E, Champagne J, Lehenbauer TW, Li X,
554   Chase J, Nguyen T, Pires AFA, Atwill ER, and Aly SS. 2019. Association between herd
555   management practices and antimicrobial resistance in Salmonella spp. from cull dairy
556   cattle in Central California. *PeerJ* 7:e6546. 10.7717/peerj.6546
557   Quinn P, Markey BK, Carter M, Donnelly W, and Leonard F. 2002. *Veterinary microbiology and*
558   *microbial disease*: Blackwell science.
559   Robbins J, Von Keyserlingk M, Fraser D, and Weary D. 2016. Invited review: Farm size and
560   animal welfare. *Journal of animal science* 94:5439-5455.
561   Roy R, Higgins R, Fortin M, and Tardif S. 2001. Salmonella Give infection in 2 dairy herds. *The*
562   *Canadian Veterinary Journal* 42:468.
563   Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, Jones JL, and Griffin
564   PM. 2011. Foodborne illness acquired in the United States—major pathogens. *Emerging*
565   *infectious diseases* 17:7.
566   Smith BP. 2014. *Large animal internal medicine-E-Book*: Elsevier Health Sciences.
567   Troutt HF, Galland JC, Osburn BI, Brewer RL, Braun RK, Schmitz JA, Sears P, Childers AB,
568   Richey E, and Mather E. 2001. Prevalence of Salmonella spp in cull (market) dairy cows
569   at slaughter. *Journal of the American Veterinary Medical Association* 219:1212-1215.
570

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.
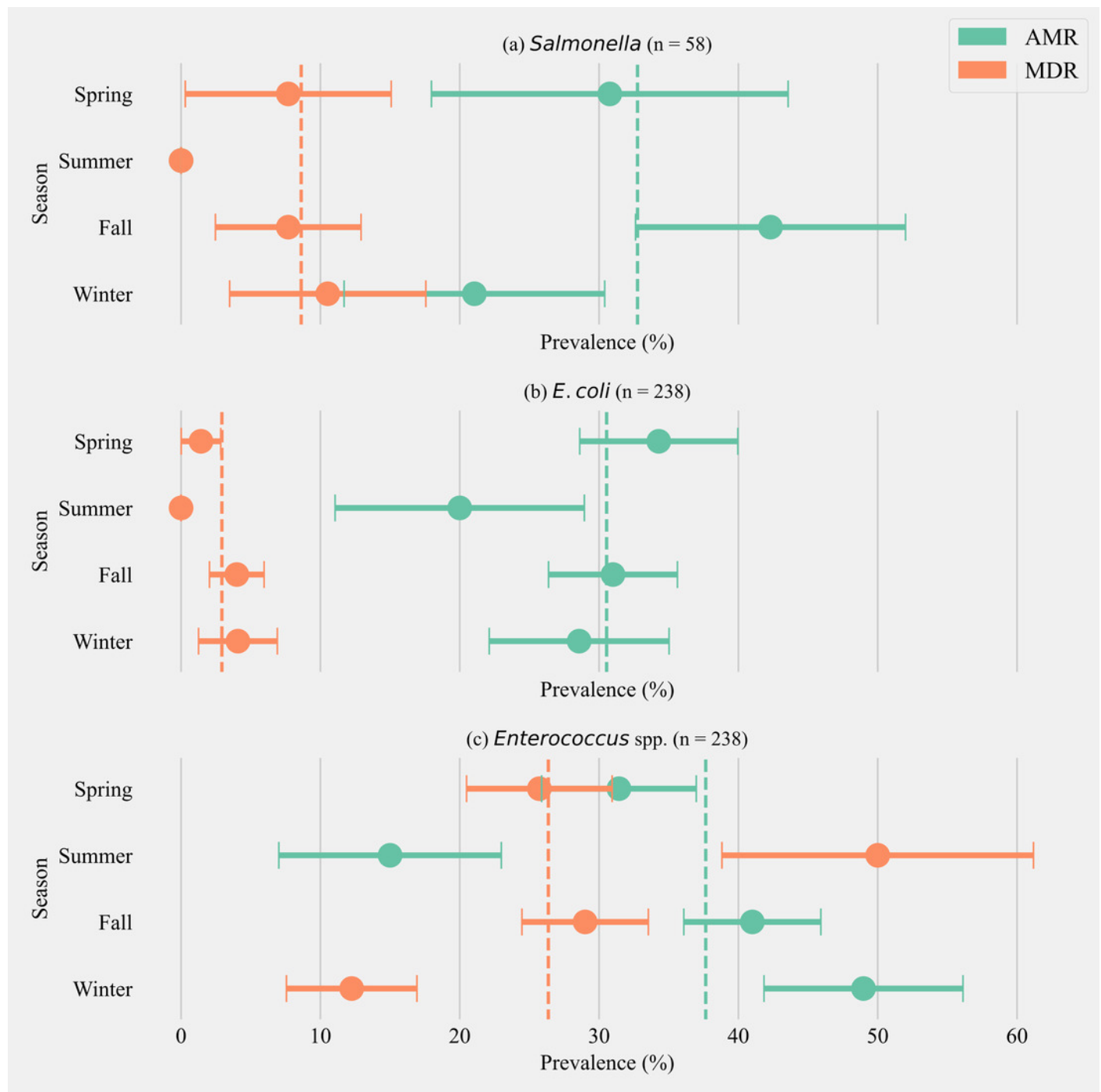
Figure 2

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.

Nodes are represented by stacked histograms depicting distribution samples in the data (AMR, MDR, no resistance), followed by optimum decision point (pointer on the x-axis). Arrows on the left and right indicate lesser and greater than the decision point respectively. Final nodes are represented by pie chart with distribution of samples. Factor acronym definitions are described in Table 2.
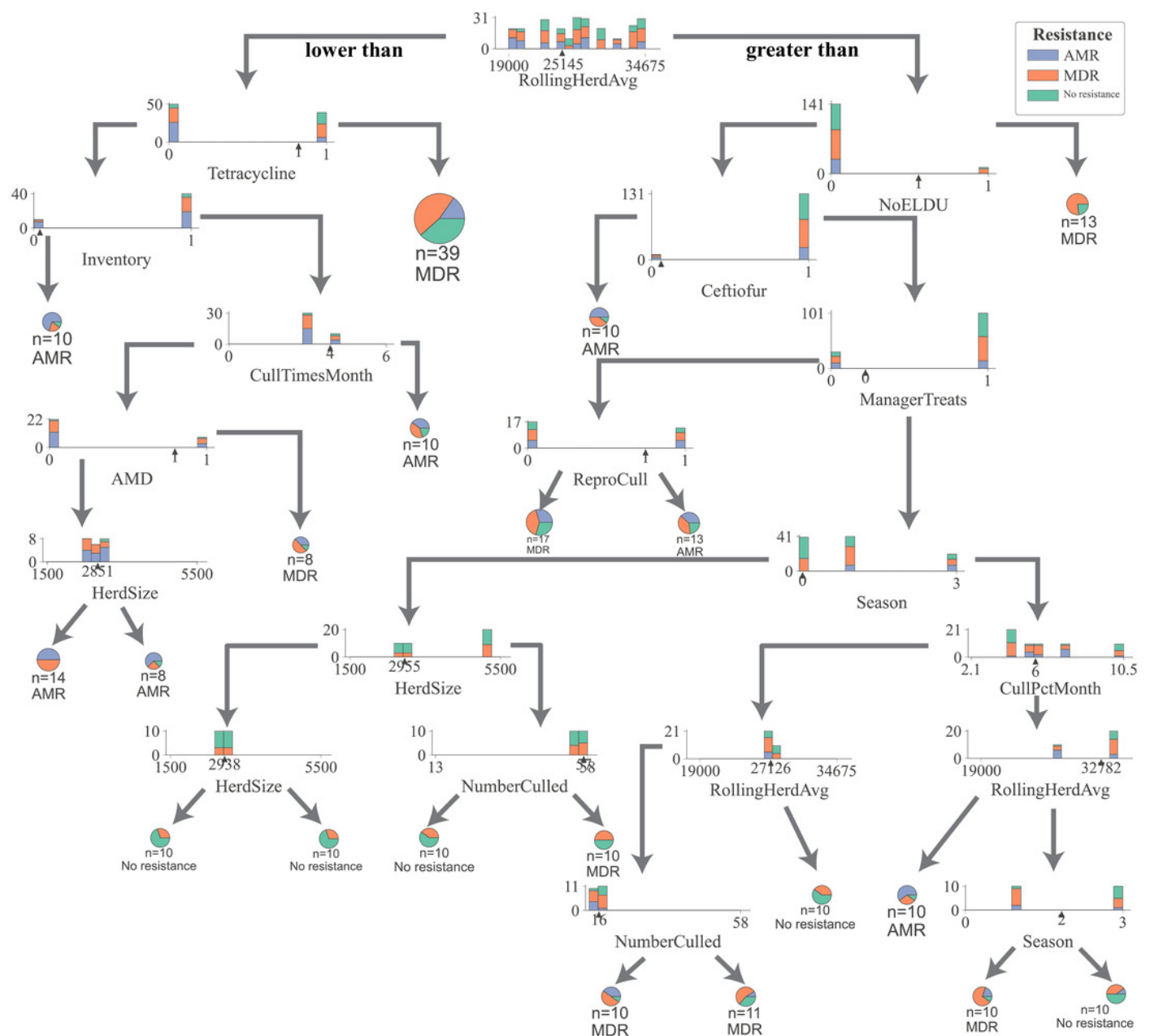
Figure 3

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.
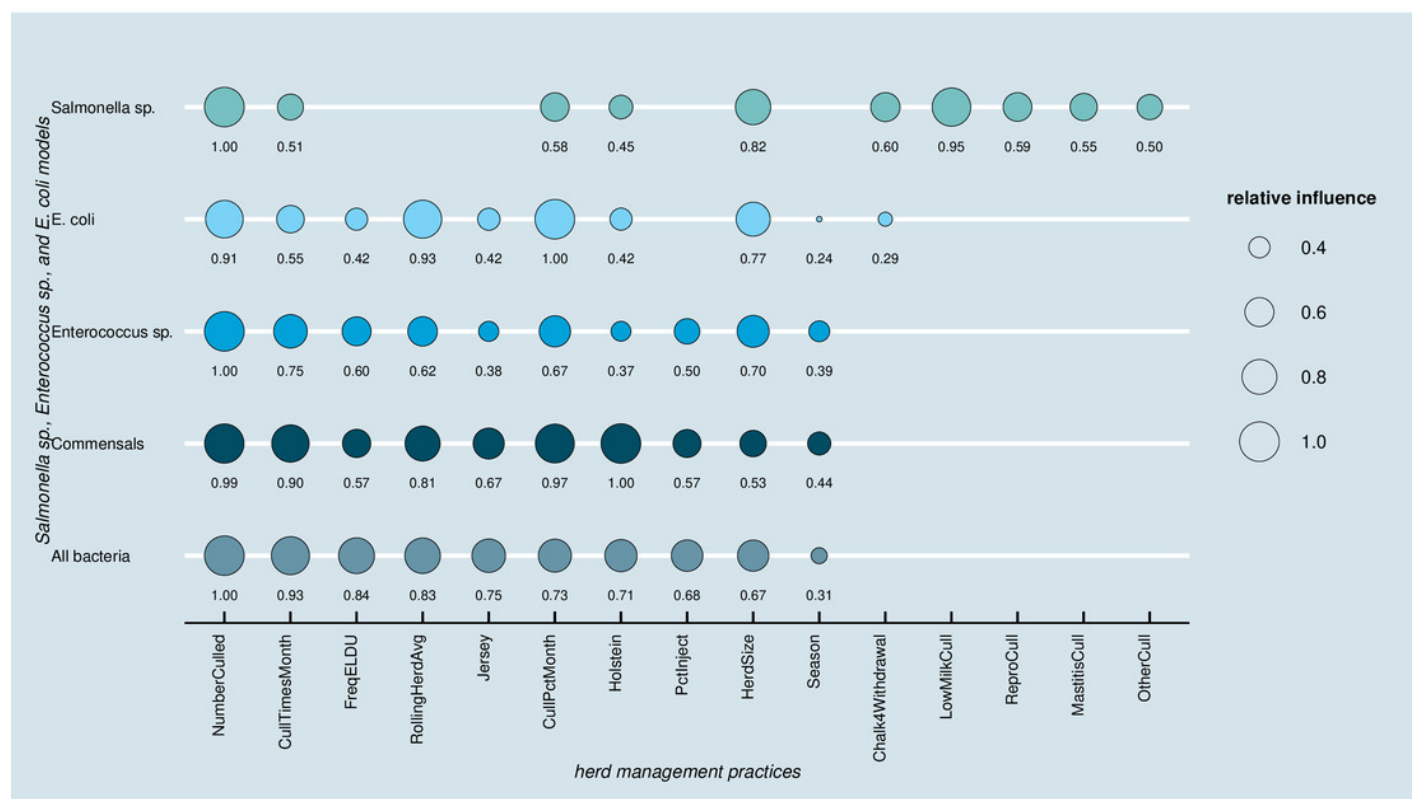
Figure 4

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.
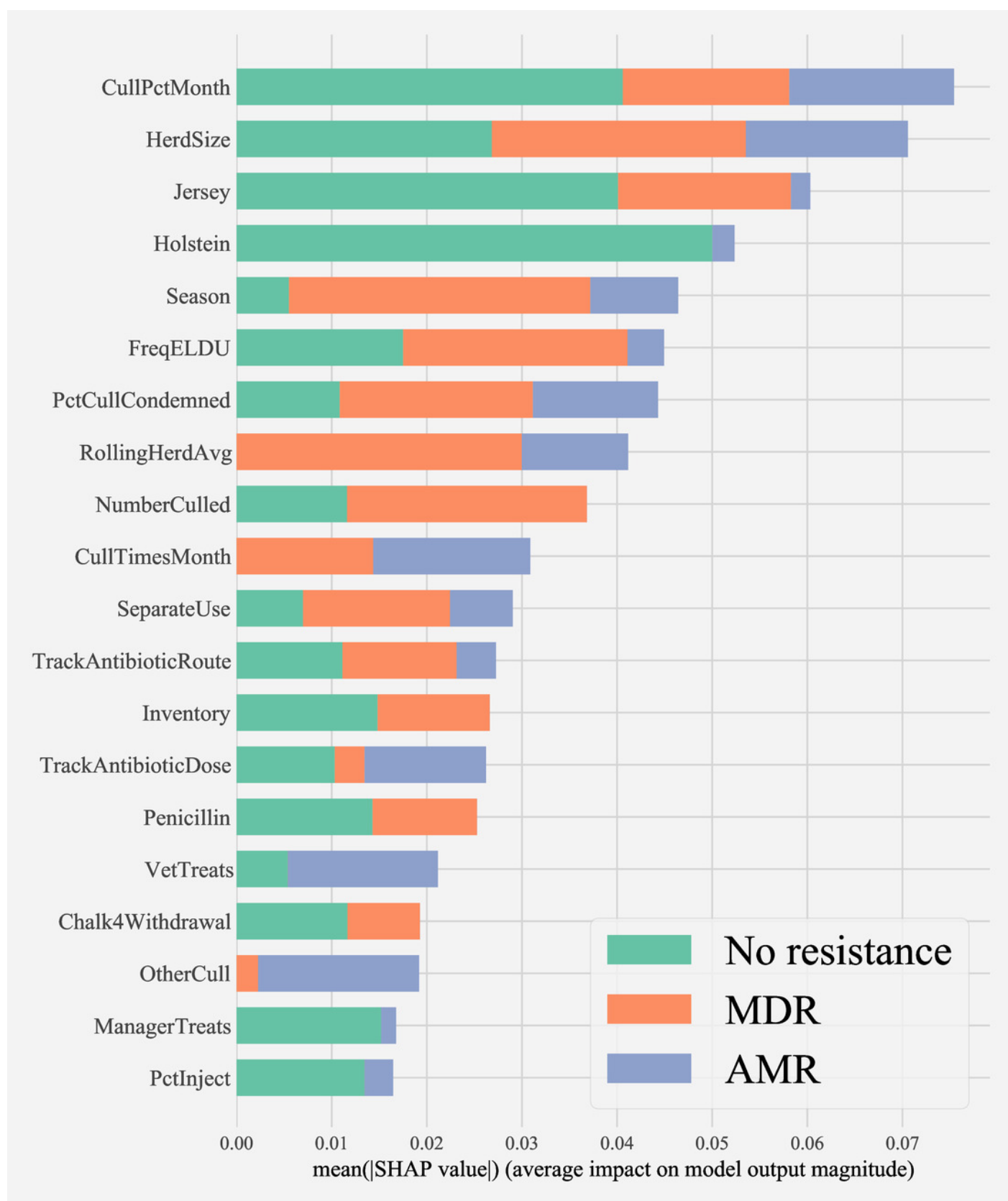
Figure 5

Partial dependence indicating the association of top six 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. Blue shaded region and error bars represents standard deviation of partial dependence (n=238).
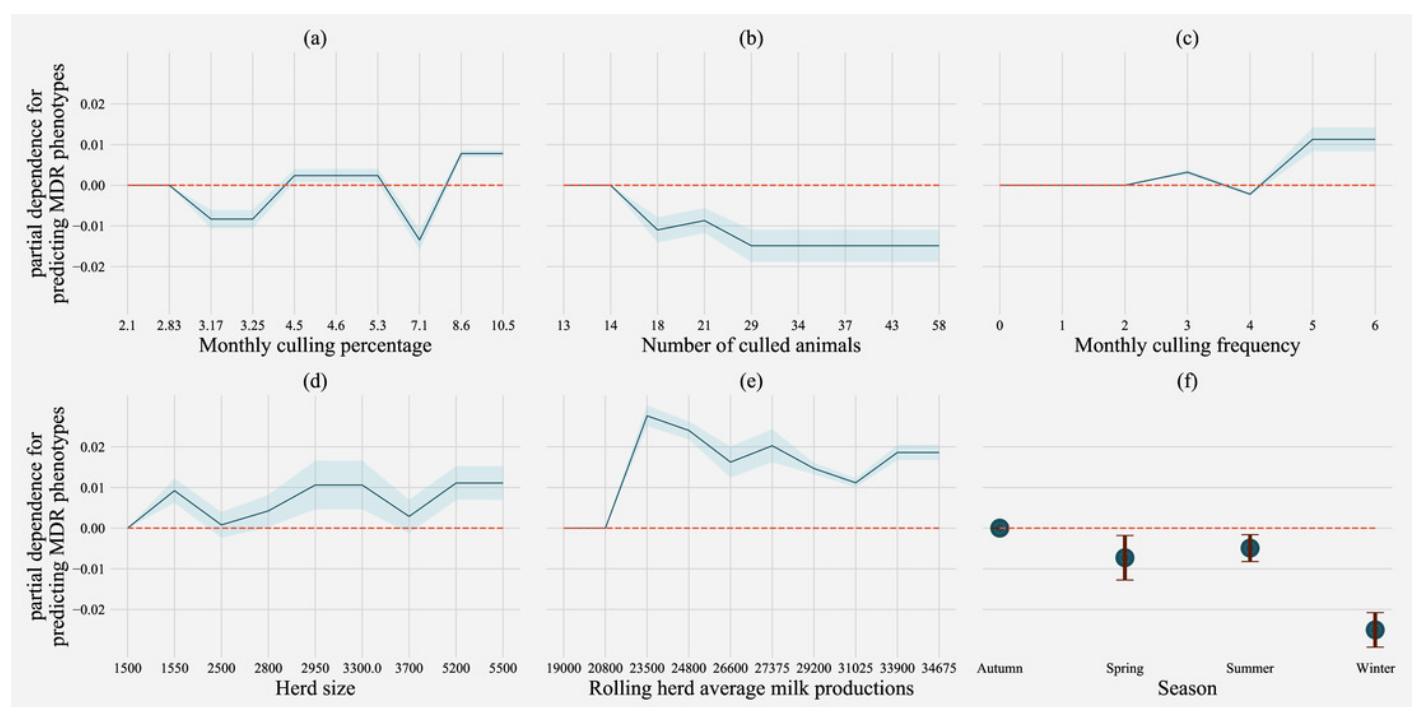
Figure 6

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.
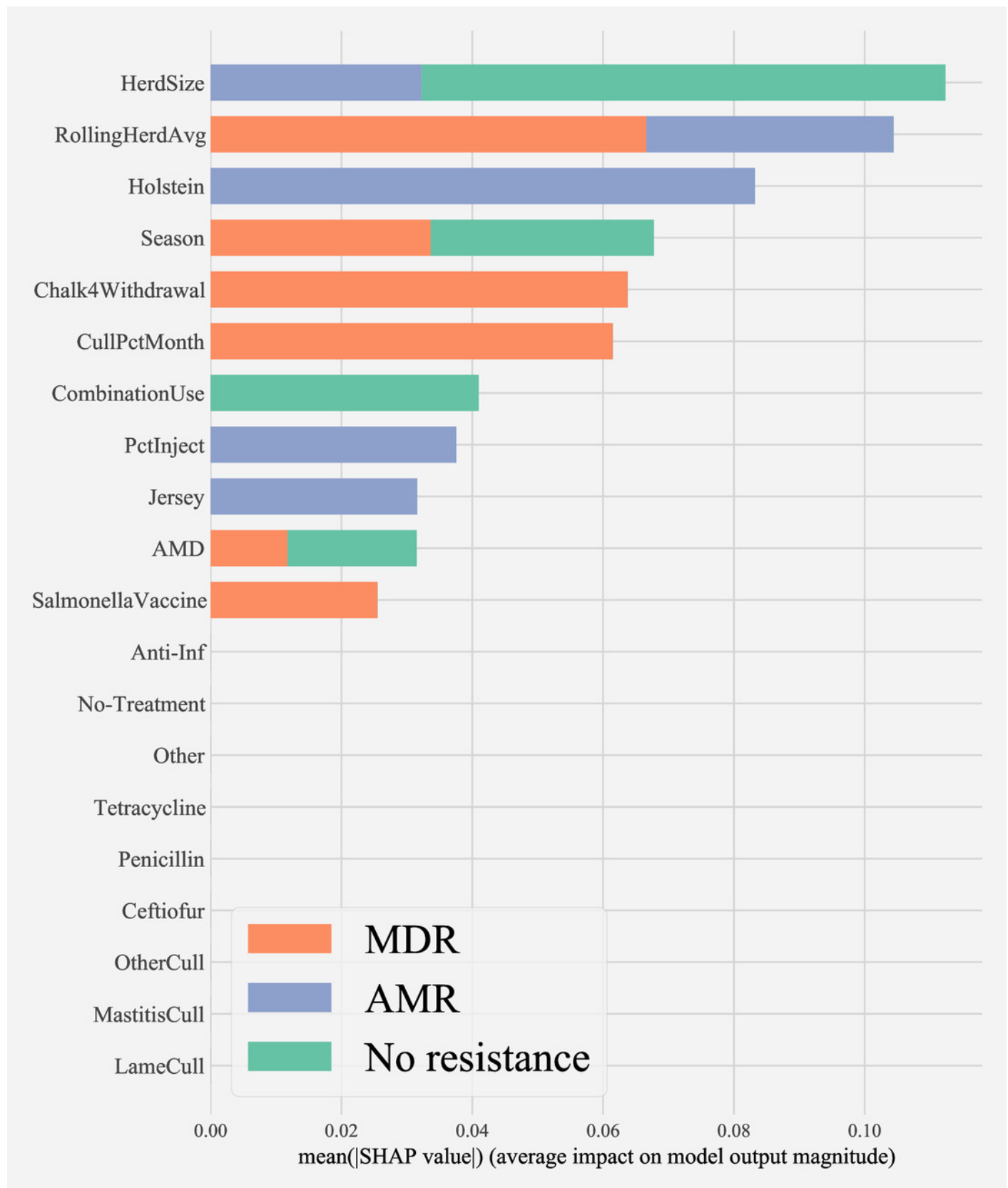
Figure 7

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. Blue shaded region and error bars represent the standard deviation of partial dependence (n=58).
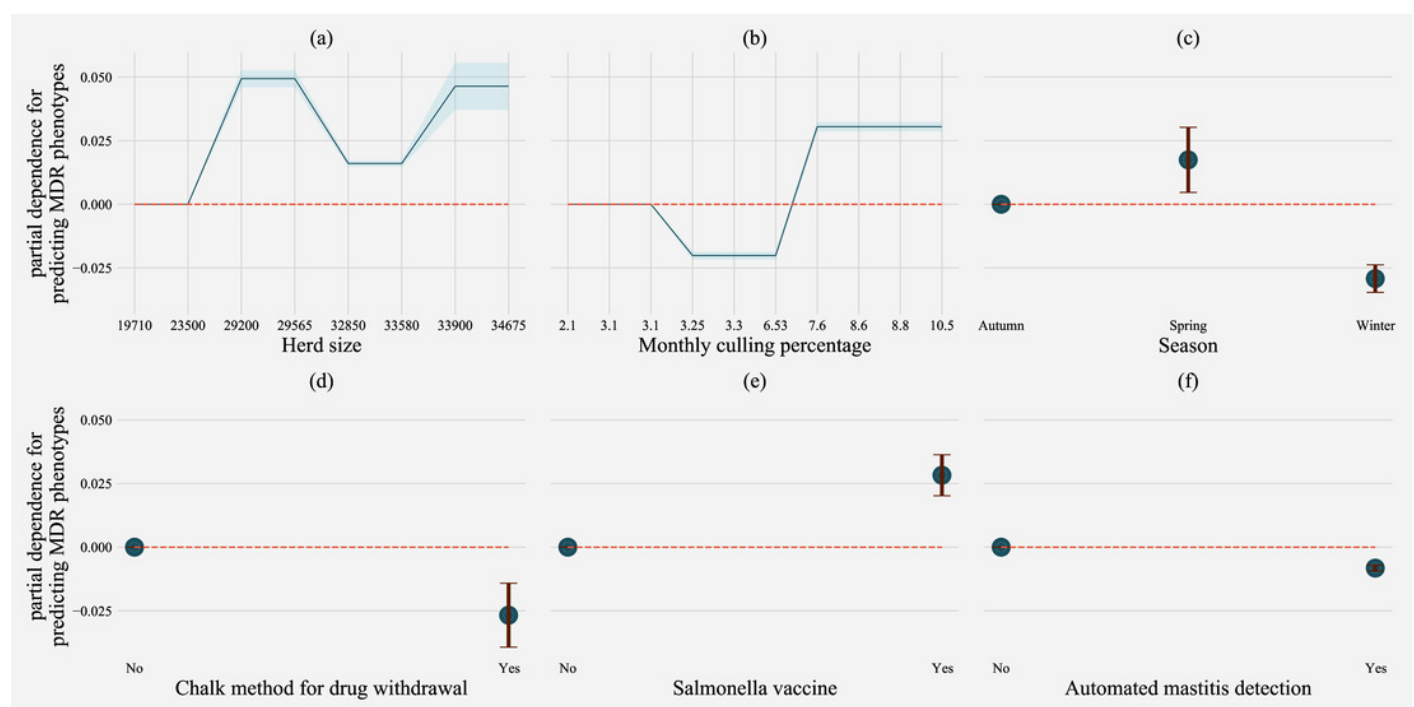
Table 1_(on next page)

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.

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 sensitivity		0.75	0.66	0.50	0.44	0.47
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		0.65	0.59	0.51	0.48	0.38

sensitivity	

Predictive features

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

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

1   †Represents phenotypes that are multi-drug resistant. **Other single isolates of MDR phenotypes from *Enterococcus* spp. isolates: 1)

2   tetracyclines, amphenicols, oxazolidinones, nitrofuran antibacterial, macrolides 2) oxazolidinones, nitrofuran antibacterial, macrolides,

3   glycopeptides 3) tetracyclines, amphenicols, oxazolidinones, macrolides 4) streptogramin, oxazolidinones, nitrofuran antibacterial 5) tetracyclines,

4   streptogramin, nitrofuran antibacterial, macrolides 6) tetracyclines, amphenicols, streptogramin, oxazolidinones, nitrofuran antibacterial,

5   macrolides 7) tetracyclines, amphenicols, macrolides, streptogramin, oxazolidinones, macrolides 8) amphenicols, macrolides, streptogramin,

6   oxazolidinones, nitrofuran antibacterial 9) macrolides, streptogramin, oxazolidinones, nitrofuran antibacterials, macrolides 10) amphenicols,

7   nitrofuran antibacterial, macrolides 11) oxazolidinones, nitrofuran antibacterial, glycopeptides 12) amphenicols, oxazolidinones, macrolides 13)

8   tetracyclines, macrolides, oxazolidinones, nitrofuran antibacterial, macrolides 14) tetracyclines, streptogramin, nitrofuran antibacterial 15)

9   tetracyclines, macrolides, nitrofuran antibacterial, macrolides 16) amphenicols, streptogramin, macrolides 17) tetracyclines, streptogramin,

10   oxazolidinones 18) streptogramin, nitrofuran antibacterial, macrolides 19) tetracyclines, macrolides, oxazolidinones, nitrofuran antibacterial 20)

11   tetracyclines, oxazolidinones, nitrofuran antibacterial 21) tetracyclines, amphenicols, macrolides, oxazolidinones, nitrofuran antibacterial 22)

12   tetracyclines, amphenicols, macrolides, nitrofuran antibacterial 23) amphenicols, macrolides, nitrofuran antibacterial. *other single isolates of amr

13   phenotypes from *enterococcus* sp. isolates: 1) amphenicols, nitrofuran antibacterial 2) streptogramin, oxazolidinones 3) amphenicols, macrolides

14   4) macrolides, oxazolidinone

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

1 †Represents phenotypes that are multi-drug resistant. *Other single isolates of AMR phenotypes from *E. coli*: 1) quinolones, aminoglycosides 2) amphenicols,

2 tetracyclines 3) tetracyclines, folate pathway antagonist 6) cephalosporins, fluoroquinolones 7) amphenicols, folate pathway antagonist. ** other single isolates

3 of mdr phenotypes from *e. coli*: 1) amphenicols, folate pathway antagonist, aminoglycosides 2) amphenicols, tetracyclines, cephalosporins, aminoglycosides 3)

4 amphenicols, tetracyclines, folate pathway antagonist 4) amphenicols, tetracyclines, folate pathway antagonist, aminoglycosides 5) tetracyclines, cephalosporins,

5 fluoroquinolones, quinolones, aminoglycosides

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.

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

1 †Represents phenotypes that are multi-drug phenotypes