DISSERTATION

SAMPLING AND ESTIMATION TECHNIQUES FOR SURVEILLANCE AND MONITORING

Submitted by

Bruce Alan Wagner

Department of Clinical Sciences

In partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring, 2003

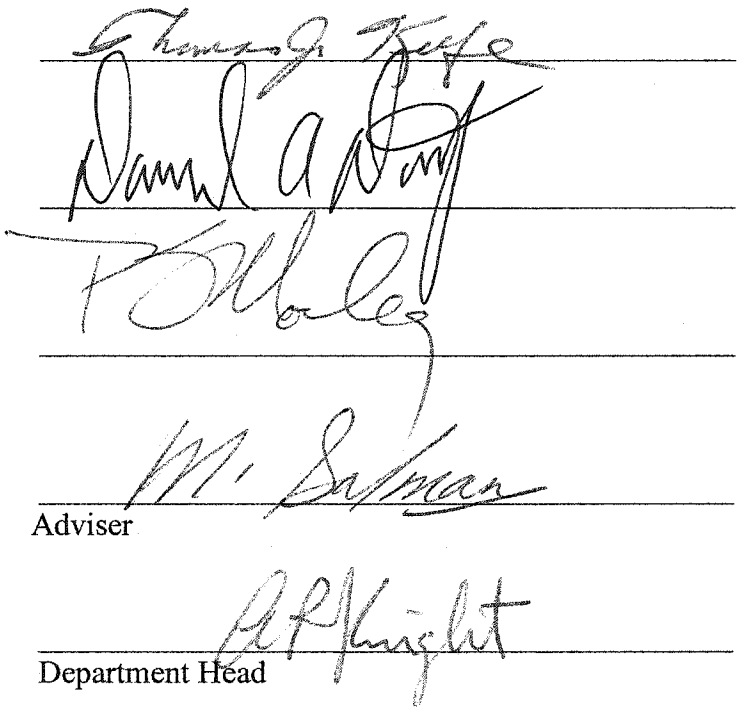
ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

March 5, 2003

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY BRUCE ALAN WAGNER ENTITLED SAMPLING AND ESTIMATION TECHNIQUES FOR SURVEILLANCE AND MONITORING BE ACCEPTED AS FULFILLING IN PART THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work



Adviser

Department Head

ABSTRACT OF DISSERTATION

SAMPLING AND ESTIMATION TECHNIQUES FOR
SURVEILLANCE AND MONITORING

The complexity of issues facing animal agriculture and veterinary and human medicine necessitate the use of appropriate scientific methods for obtaining and analyzing data so that objective inferences can be made and practical solutions can be sought. An overall objective of these studies was to examine the validity of underlying assumptions of sampling methodologies in the context of antimicrobial susceptibility and surveillance and monitoring for animal diseases. A secondary objective was to examine analytical methods that can be used to address the intricacies of antimicrobial susceptibility data.

Evaluation of sampling strategies for measuring antimicrobial susceptibility patterns in isolates taken from the feces of feedlot cattle was the subject of two studies. In the first study, rectal and pen floor fecal samples were collected. Individual samples were used to create pools of 5 and 10 fecal samples. Five *Escherichia coli* isolates were obtained from each individual and pooled sample. The susceptibility patterns were compared among the collection methods and the individual/pooled samples.

Antimicrobial resistance patterns from isolates obtained from rectal samples did not differ from isolates obtained from pen floor fecal samples. Little pen-to-pen variation in

resistance prevalence was observed but clustering of resistance phenotypes within pens and samples was detected. Pooling of fecal samples yielded resistance patterns that were consistent with those of single fecal samples when the prevalence of resistance to an antimicrobial was > 2 percent. Pooling may be a practical alternative when investigating patterns of resistance that are not rare.

The second study of antimicrobial resistance in *E. coli* isolated from feedlot cattle focused on the short-term repeatability of antimicrobial susceptibility patterns. Fecal samples were collected from the floors of six pens on two sampling occasions separated by 48 hours. Resistance to individual antimicrobials was consistent across the periods and individual/pooled samples by all analytical measures when the prevalence of resistance was at least 2 percent. Inconsistent results were obtained for antimicrobials to which resistance rarely occurred. The apparent inconsistencies did not appear to be related to external factors but rather to sampling intensity. Short-term stability is a plausible assumption under sampling strategies designed to detect specific prevalence levels. However, when resistance levels are low there likely will be fluctuations in the occurrence, prevalence and central tendency measures of rare resistance phenotypes.

Factor analysis was used to explore resistance and susceptibility patterns of the minimum inhibitory concentration for the 17 antimicrobials tested on *E. coli* isolates. New generation cephalosporins, older generation beta-lactams, fluoroquinolones and aminoglycosides grouped separately as classes of antimicrobials on four of the six factors. One of the remaining factors was a grouping of antimicrobials that had been

identified as being related in previous feedlot studies. The last factor was a grouping of three of the five antimicrobials that comprise the antimicrobials found in penta-resistant strains of *Salmonella* Typhimurium. The factor analysis provided patterns in the MIC data that would not have been apparent if the antimicrobial-resistance data had been analyzed merely according to the susceptible/resistance categories.

Two-stage sampling designs are appropriate for disease surveys. A sample size formula for estimating herd-level prevalence was proposed that depends on herd-level sensitivity and specificity. The impact of the distribution of the within-herd prevalence as determined by animal-level prevalence and the intracluster correlation coefficient was modeled using a Monte Carlo simulation. At low prevalence, herd-level sensitivity increased with increasing intracluster correlation, but sensitivity was less affected at higher prevalence. Also at low prevalence, many herds were being classified as positive based only on false positive test results. Positive predictive values dropped sharply with increasing intracluster correlation. A hypothetical and two real life two-stage sampling designs were used as examples to evaluate the model and the sample size formula. The use of a distribution for within-herd prevalence resulted in a conservative estimate of herd-level test characteristics. The model allows researchers to trade off between the number of herds and the number of animals sampled by manipulating herd-level test characteristics.

Bruce Alan Wagner
Department of Clinical Sciences
Colorado State University
Fort Collins, CO 80523
Spring 2003

ACKNOWLEDGEMENTS

The pursuit of this doctorate degree has been an intellectually rewarding opportunity as well as a challenge to push myself to grow in many ways. But this dissertation and my whole program would not have been possible without the intellectual, emotional and physical input from many important people.

I would like to thank my adviser, Dr. Mo Salman for his contributions to this work and to my education. I remember back when I embarked on this process and Dr. Salman expressed his feelings that learning should be a two-way street. These were not hollow words. We spent numerous hours sharing ideas in very stimulating discussions. I emphasize sharing because I always felt that Dr. Salman listened to my ideas, appreciated them and took the worthwhile ones to heart. Similarly, I learned a great deal from those discussions. Dr. Salman has the ability to convey the big picture while still being able to discuss the mechanics of simulation models. Throughout the program, I have been thankful for guidance that Dr. Salman provided on the many decisions both within my program and in my professional career. His energy and enthusiasm are a wonder.

All my committee members made important contributions to my education. Sometimes the meetings were informal sharing of ideas. I met with Dr. Paul Morley for lunch one day. From that lunch grew the concept for a large portion of this dissertation. Dr.

Morley perspectives on epidemiology will stay with me forever. He emphasized the basic, philosophical context of the science of epidemiology through his teaching and through our many discussions. He is just as comfortable working with the analytical side of epidemiology and we traded many ideas over the years. I would also like to thank him for critical review of my research writing as well as his gentle suggestions for improving my work.

Dr. David Dargatz is an inspiration and a model for me. He has a seemingly endless desire to explore ideas and pursue opportunities for gaining knowledge that is infectious. Dr. Dargatz is at once a source of questions that need to be addressed and a source of well thought out perspectives on important issues. His practical guidance as well as his encouragement throughout my studies was very valuable.

I started the wheels of pursuing a doctorate by taking two biostatistics classes from Dr. Tom Keefe. I remember asking him, after one of his lectures, about sampling and estimation when an imperfect diagnostic test was being applied. That was the first of many discussions we had on this and many other statistical and epidemiological analysis issues. Dr. Keefe was a sounding board for new ideas and application of statistical methodologies.

Many other people contributed substantially to this research. Dr. Tom Wittum and his staff at the Ohio State University performed the *Escherichia coli* isolation and antimicrobial susceptibility testing. Alison Hurwitch, Dr. Zach Mills, Kirsten

Chmielewski, and Dr. Ann Hillberg-Seitzinger assisted with the collection of samples from feedlots.

I also would like to acknowledge the support for these projects that was received from the USDA:Cooperative State Research, Education, and Extension Service through the Colorado State University - Center for Economically Important Infectious Animal Diseases and from the Research Council of the College of Veterinary Medicine and Biomedical Sciences, Colorado State University.

Lastly I would like to thank my wife, Nancy, for encouragement that she provided, the extra burdens that she accepted, and for her timely assurances of confidence in me that I could accomplish my goals. My two sons, Benjamin and Max, sacrificed a great deal over the duration of this degree. I hope they understand the importance of their role in my attaining this degree and in my life. Many times I felt I was living an example for them so that they can pursue and attain whatever they set as their goals if they work hard and have the desire.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter provides an overview of the structure of this dissertation. The broad theme of the dissertation involves the design and analysis of epidemiological studies of animal agriculture. The dissertation is divided into two major topics: sampling and analyzing antimicrobial susceptibility; and herd-level testing for disease/infection.

Chapter 2 is focused on the statistical methods for analyzing antimicrobial resistance outcomes. The chapter includes a review of methods that have been used in both human and animal studies. Also, the chapter offers alternative analytical approaches using either existing methodology or novel approaches. This chapter is formatted for the *American Journal of Veterinary Research*.

Chapter 3 is the first of three papers that report the results of a study investigating antimicrobial susceptibility non-type specific *Escherichia coli* in feedlot cattle. This paper, which has been published in the *American Journal of Veterinary Research* (Wagner et al., 2002), compares the feasibility of sampling approaches including collecting rectal versus pen floor samples and individual fecal samples versus pooled samples.

Chapter 4 examines the short-term repeatability of antimicrobial resistance patterns obtained from *E. coli*. The antimicrobial resistance patterns of isolates obtained from individual and pooled fecal samples taken two days apart are compared. This paper has been accepted by the *Journal of Veterinary Diagnostic Investigation*.

Chapter 5 presents a factor analysis of the minimum inhibitory concentration outcomes for *E. coli* isolated from feedlot cattle feces. This multivariate approach explores the pattern of susceptibility outcomes from testing individual isolates against multiple antimicrobials. This chapter has been published in *Preventive Veterinary Medicine* (Wagner et al., 2003).

Chapter 6 is a literature review of theoretical and applied approaches to herd-level diagnostic testing.

In Chapter 7, a two-stage sampling approach is developed by first investigating the impact of clustering on determination of herd-level disease status. A first stage sample size calculation is proposed, and both hypothetical and real life design protocols are developed.

Chapter 8 provides a summary of the conclusions from Chapters 2 through 7.

References

Wagner BA, DA Dargatz, MD Salman, PS Morley, TE Wittum and TJ Keefe. 2002. Sampling Techniques for Antimicrobial Susceptibility of Enteric *Escherichia coli* Recovered From Feedlot Cattle. *Am J Vet Res* 63:1662-1670.

Wagner BA, MD Salman, DA Dargatz, PS Morley, TE Wittum and TJ Keefe. 2003. Factor analysis of minimum inhibitory concentrations for *Escherichia coli* isolated from feedlot cattle to model relationships among antimicrobial outcomes. *Prev Vet Med* 53(3):127-139.

CHAPTER 2

ANALYSIS METHODS FOR BACTERIAL ANTIMICROBIAL RESISTANCE OUTCOMES

Summary

The objective of this study was to delineate statistical methods that have been used to analyze antimicrobial resistance data and to identify additional methods that may have application. A review of recent veterinary and human antimicrobial resistance literature was conducted. Statistical methodologies implemented in the literature were identified and categorized into broad analytical technique categories. Additional statistical procedures, both established and newly derived, were incorporated into the review where methodological gaps or analytical needs were identified.

Descriptive statistical techniques are widely used in the literature. A cumulative distribution of minimum inhibitory concentrations, standardized for resistance breakpoints, was proposed for visually comparing multiple antimicrobials. Simple indices have been used to summarize antimicrobial resistance data and other indices may have application. Categorical and nonparametric methods have been used primarily to analyze susceptible/resistance categorized data. Multivariate and cluster analytical techniques have not been widely used but have potential application.

Many descriptive and inferential statistical techniques are available to researchers. The advantages and disadvantages of the methods must be considered when selecting analytical techniques. Use of minimum inhibitory concentration data is encouraged because susceptible/resistance data can result in a loss of information. Analytical methods need to account for the lack of independence of observations (clustering) in appropriate situations. Multiple drug resistance or multiple outcomes can be analyzed using multivariate statistical methods.

Introduction

The volume and scope of research in the area of antimicrobial resistance of human and veterinary bacterial isolates belies the critical importance of the issue. Ultimately, the concern is focused on the potential for treatment failure and the selection of bacteria that no longer respond to currently available antimicrobial agents. The debate on the roles of prescription practices in humans, antimicrobial usage in animals (production and companion) and plants, and a declining infection control infrastructure is underway[14;28;29;46;53;56;67;70;75]. Regardless of the debate, the general goal of research should be to create a better understanding of the emergence, dissemination, and maintenance of resistance in human and animal populations and the environment.

These general goals entail a number of specific research questions and objectives. An initial research objective is to quantify the amount of resistance (prevalence) or susceptibility that a bacterial species has to a single antimicrobial or to a panel of antimicrobials, usually within a specific source population. Detection of changes in

prevalence, such as trend identification, is an important extension of point-in-time prevalence estimation. Identification of emerging resistance is a second objective of this type of research. This objective is essentially an evaluation of trend where the focus is detecting the initial increase in resistance prevalence above some minimum threshold or, perhaps, a substantial shift in the distribution of the minimum inhibitory concentration (MIC) values. The purpose of surveillance activities, establishing resistance base levels and identifying changes in resistance patterns over time[59], is a combination of the first two research objectives. A third research objective is to identify risk factors that are associated with identified resistance patterns.

A variety of epidemiological research approaches to address these objectives have been implemented including case-studies[28;34;68], experimental designs[20;26], cohort studies [5;23;41;48] and cross-sectional studies[18;44;71]. Many investigations have relied on evaluation of human and animal clinical isolates[3;29].

Predictably, a diverse battery of statistical methods has been used to analyze resistance data. The objective of this paper is to delineate methods that have been used and to identify additional methods that may have application. For presentation, these methods will be grouped into six general categories: 1) descriptive, 2) indices, 3) categorical, 4) nonparametric, 5) multivariate, and 6) analysis of clustered data.

Descriptive methods

Descriptive statistical methods have been widely applied to data categorized into susceptible/intermediate/resistant (SIR) and to MIC data. SIR data are usually presented in tables, graphs (histograms and line charts) or in text as the proportion of isolates that are resistant (susceptible). Antimicrobial resistance data presented in this manner occur commonly in both the human and veterinary medical literature to describe resistance to specific antimicrobials[7;9;10;27;50;57;58;60;61]. Antibiograms are a tabulation of resistance phenotypes of bacterial isolates. Isolates are grouped based on shared phenotypic resistance patterns as a way of descriptively presenting multiple resistance proportions[9;18;54;61;66;70].

The use of MIC data to describe patterns of resistance is pervasive in the literature. Commonly, MIC values are summarized in a frequency cross-tabulation of the antimicrobials and their concentration levels[1;47;79]. Cumulative distributions have also been presented in cross-tabulations[40]. Also, frequency distributions and cumulative distributions have been combined into single tables[58;69].

Minimum inhibitory concentration values are summarized in other formats besides frequency tabulation. An alternative format is to represent the MIC data with summary descriptive statistics[40]. Often the 50^{th} and 90^{th} percentiles, MIC_{50} and MIC_{90} respectively, are presented along with the maximum and minimum values which are sometimes erroneously referred to as the range[4;13;80]. The geometric mean is used as a measure of central tendency in the MIC data[15]. Another alternative is to graphically depict the MIC data, typically with histograms[11]. Sahm et al.[59] presented a histogram of MIC data but

also demarcated the SIR boundaries. Cumulative distribution plots, also called a "Finland-o-gram", have been suggested as a simple alternative that is easy to understand and is useful as long as numerous distributions are not being depicted in a single figure[40].

The cumulative distribution plots may have use in comparing different antimicrobials in relation to their respective breakpoint values. MIC values for individual antimicrobials can be standardized by using the following formula:

$$MIC_{standard} = \frac{\ln(MIC/breakpoint)}{\ln(2)}$$

The National Antimicrobial Resistance Monitoring System – Enteric Bacteria (NARMS-EB) publishes information regarding the occurrence of resistance in isolates of *Salmonella* spp. from a variety of animal species which can be used to demonstrate this standardization method[52]. In 1997, NARMS-EB reported MIC results for clinical cattle isolates. Four antimicrobials, cefoxitin, cephalothin, nalidixic acid, and streptomycin, were selected for a graphical demonstration of this method. The nonstandardized MIC values for ceftiofur, nalidixic acid, and streptomycin define fairly distinct peaks (Figure 2.1). Cephalothin MIC values tend to be more broadly distributed. When the MIC values are standardized relative to their respective breakpoints, the graphical representation changes substantially (Figure 2.2). The cumulative distributions of three antimicrobials, ceftiofur, cephalothin, and nalidixic acid, now appear very similar even though the broad distribution of cephalothin is still reflected in its slightly less steep increase to 100 percent. The distribution of standardized MIC values for streptomycin is shifted to the right, still reflecting the presence of resistance. The percent susceptible for

Figure 2.1. Nonstandardized minimum inhibitory concentration values for four antimicrobials. Log breakpoints based on NCCLS guidelines for each antimicrobial is ceftiofur ≥3.47; cephalothin ≥3.47; nalidixic acid ≥3.47; streptomycin ≥4.16

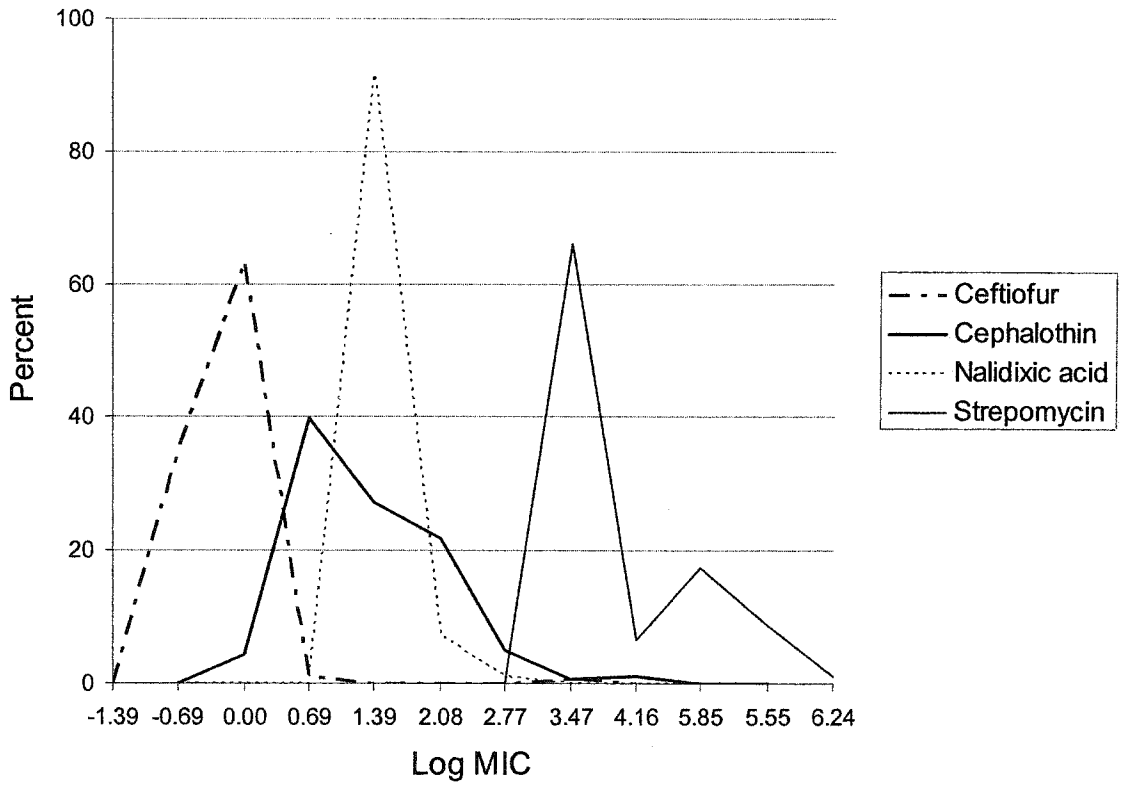
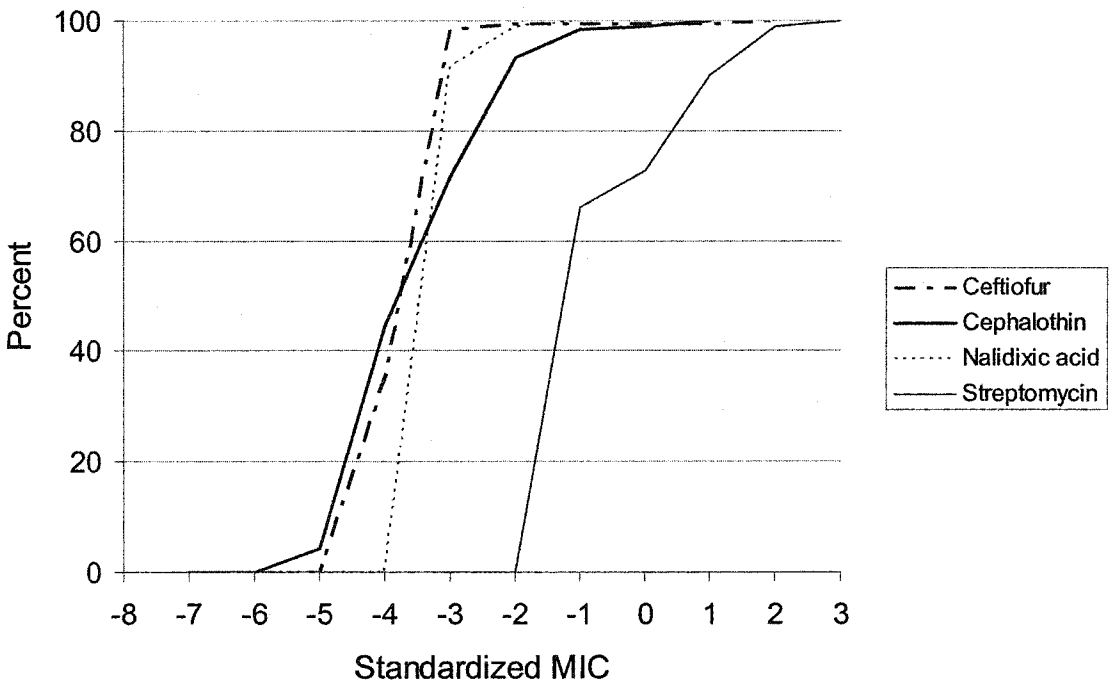


Figure 2.2 Standardized cumulative minimum inhibitory concentrations MIC distributions for four antimicrobials



streptomycin, 72.7 percent, is the value of the cumulative distribution when the standardized MIC value is equal to zero. Correspondingly, the percent resistance is 27.3 percent.

The advantage of descriptive data presentation is that large amounts of data can be summarized and displayed in a manner that is understandable by a wide range of interested groups. SIR data have additional advantages of being relatively easily understood, compact for presentation, and more readily related to the in vitro interpretation of resistance. Antibiograms have the advantage of allowing for examination of the multivariate nature of the SIR data including looking for groupings and multiple resistance, as well as identification of rare or high prevalence resistance phenotypes.

There are also disadvantages of descriptive data presentations. The categorization of the MIC data (or zone diameter data) data into SIR data can result in the loss of some information. Subtle shifts in distributions of MIC values, which are below the breakpoint, are not detectable. SIR data would not necessarily be the most appropriate data structure when the objective of testing is to serve as an early warning when slight progressions, perhaps all below the resistance breakpoint, in MIC values are important to monitor. Jones[40] noted that the dynamic nature of breakpoints over time may lead to interpretability problems with SIR data. He suggested reliance on MIC data distribution to allow for consistency over time. The use of MIC values does remove the subjectivity associated with the choice of breakpoints and the SIR categorization. However, MIC

values, determined by standard microbroth dilution methods, do not form a truly continuous distribution because of the scale (doubling of dilutions) and the potential for truncation of values at both extremes. Analyses that assume continuous level data should be used cautiously. Antibiograms serve as a good example of the problems with relying solely on SIR data. Suppose in one isolate the MIC value is one level below the breakpoint and in another isolate the MIC value is at the breakpoint. These two isolates would appear as different phenotypes because of the SIR categorization although biologically the difference may not be important. Another concern with the use of antibiograms is that they can be difficult to interpret if relationships of more than two antimicrobials are being considered at one time.

Summary descriptive analyses also have advantages and disadvantages. Cumulative distribution functions (CDF) allow for visual representation of quantiles, including MIC_{50} and MIC_{90}, as well as comparison of several antimicrobials or bacterial species. If only a few MIC categories are represented in the data then the CDFs and MIC quantiles may not be very informative. Also, the CDF will not be very meaningful if only a few isolates are being represented.

Indices

Ecologists have long sought indices to represent a single quantitative measure of species diversity[38]. In ecology, species diversity is considered to be a function of the number of species present (species richness) and the evenness with which individuals of the different species are distributed[43]. The translation of the concept of a single

comprehensive value which represents animal species diversity to obtaining a single value which represents the diversity of sensitivity/resistance for a panel of antimicrobials has intuitive appeal.

Perhaps the simplest index is species richness, which has been adopted somewhat indirectly as an index of antimicrobial resistance. In the context of antimicrobial susceptibility, species richness would be interpreted as the number of antimicrobials to which a single isolate is resistant. Penta-resistance, although often applied to specific clones, essentially is an example of antimicrobial resistance richness or an index of multi-resistance. Sherley et al.[62] used the multi-resistance index, which they called multiple antibiotic resistance, to compare patterns across a number of bacterial species. They constructed histograms of multi-resistance for eight species to allow for visual comparison of the frequency distributions. Seyfarth et al.[61] compared the distribution of the number of antimicrobials to which bacterial isolates were resistant among humans and three species of production animal sources.

The number and selection of antimicrobials in the test panel can impact the multi-resistance index. The maximum value of the index could be limited by selection of a small test panel. Also, if related antimicrobials are placed on the panel, the index value could be artificially manipulated depending on the resistance status of the antimicrobials. The multiple antibiotic resistance (MAR) index, defined by Krumperman[44] as the proportion of antimicrobials to which an isolate is resistant or the number of antimicrobials to which an isolate is resistant divided by the number of antimicrobials

that the isolate was tested against, adjusts for panel size but not for the diversity of the antimicrobials selected for the panel. The MAR index, also referred to as the antibiotic resistance index (ARI)[33], can be used to summarize the proportion of resistance among several isolates. The index has been used to identify sources of fecal contamination, both in food[44] and water[42], and to track changes in resistance patterns over time in calves [33].

With the exception of richness and MAR/ARI, it appears that antimicrobial susceptibility studies have not drawn on the many indices available from the ecological sciences. Simpson's index[63], a measure of evenness, may have application as a index of diversity for analysis of antimicrobial resistance patterns. The index has been used to evaluate the discriminatory ability of bacterial typing systems[37]. Use of the index can be demonstrated with a comparison of two hypothetical bacterial samples, each with a size of 100 isolates, which are individually tested for susceptibility. Further suppose that 10 antimicrobial resistance phenotypes are represented in each of the samples. Of the 100 isolates in the first sample, suppose that phenotypes 1 through 4 each had 10 isolates, phenotypes 5 through 9 each had one isolate and the last phenotype had 55 isolates. Also suppose that the distribution of the second 100 isolates was equal in all phenotypes (10 per phenotype). Simpson's index is represented as:

$$D = 1 - \sum_{i=1}^{k} \frac{n_i(n_i - 1)}{N(N-1)}$$

where, in the context of the hypothetical analysis of the antibiogram above, n_i = the number of isolates in the i^{th} phenotype (k phenotypes) of the antibiogram and N = the total number of isolates used to construct the antibiogram. Simpson's index would yield a value of 0.664 for the first isolate sample and 0.909 for the second isolate sample. The evenness of the distribution of the phenotypes in the second sample is reflected in the higher index value as compared to the first sample.

The principal advantage of using indices is they provide a mechanism for incorporating multiple outcomes into a single measure. However there are disadvantages to consolidating all of the outcomes into a single measure. First, specific values for the indices do not necessarily represent a unique or comparable set of conditions. For instance, the value of three for richness can arise in two isolates based on two different sets of resistant antimicrobials. This problem points to the second disadvantage of indices, which is the loss of information when the data are distilled into a single number. If SIR data are used to calculate the index, then there will likely be an additional loss of information. Specific trends or changes in individual antimicrobial patterns are not readily discernable. Lastly, results from indices can be dependent on the panel selected. A panel of 5 antimicrobials, for example, has a maximum richness of 5 while a panel of 17 antimicrobials can readily have richness values in excess of 5. Despite these limitations, indices may have use in conjunction with other descriptive measures and inferential analysis to help paint patterns of antimicrobial resistance.

Categorical and nonparametric methods

Antimicrobial resistance data, whether in the form of SIR or MIC data, have been analyzed using a wide range of both parametric and nonparametric categorical statistical methods. Categorical methods that have or may be used range from the univariate estimation of a single proportion with its associated confidence interval to multivariable techniques such as logistic regression. Unlike the descriptive methods and indices, the goal of categorical analysis of antimicrobial resistance data is to be able to make inference to the population of interest.

Minimally, for inference, point estimates should be presented with an associated measure or indicator of error, such as the standard error or a confidence interval. However, the proportion of resistant (susceptible) bacteria in a sample is frequently presented in the literature with neither an associated measure of error nor a confidence interval. If the assumptions of the binomial probability distribution are appropriate and the sample size is presented, the standard error and the confidence interval can be calculated by the reader. The standard error for a proportion, given the binomial probability distribution assumption is appropriate and the population size from which the sample was taken is very large, is:

$$SE(p) = \sqrt{\frac{p*q}{n}}$$

where p is the proportion of interest, q=1-p and n=sample size[65]. For example, if 12 isolates from a sample of 85 isolates are resistant to an antimicrobial drug then the point

estimate is 0.141. The standard error estimate is 0.038. If the normal approximation is used, the 95 percent confidence interval is (0.066, 0.216). Exact confidence intervals can also be calculated by SAS[a] and other statistical software packages. In this example, the exact confidence interval is (0.075, 0.234). Note that the exact confidence interval is not symmetric.

When comparison of two proportions becomes important, the chi-square test probably is the most common statistical test applied. The proportions that have been compared include the frequency of a bacterial species[6] or frequency of resistance to a specific antimicrobial in two populations[4;58], changes in frequency of resistance between two time periods[19;59], and frequency of multiple resistant isolates versus non-multiple resistant isolates[45]. Many studies have relied on Fisher's exact test for similar types of comparisons when expected counts in the cells of the tables were low[13;48;65;71]. When the data are paired or matched more appropriate tests are available such as McNemar's test[30] or the Mantel-Haenszel chi-square test[64]. Bager et al.[5] used an adjusted chi-square test, described by Donner[22], to account for the lack of independence of isolates originating from the same poultry flock.

An extension of the chi-square test, the chi-square test for trend or linearity[30] , has been used primarily to evaluate changes in prevalence of resistance over time[19;53;64]. However, the test could be applied to other quantitatively ordered variables, such as age group, when one is interested in looking for linear trends in proportions. Whitney et al.[76] looked for trends in the proportion of resistant isolates individually to 13 antimicrobial agents

across the SIR categories for penicillin. Resistance to 12 of the 13 antimicrobials increased linearly with increasing resistance (susceptible to intermediate to resistance) to penicillin.

The chi-square and Fisher's exact tests are intended to measure the significance of association. A number of measures of association, not just significance of association, have been derived to describe the degree of association[30]. Two measures of association that have been used to analyze antimicrobial data include the kappa statistic and odds ratios. The kappa measures the amount of agreement after correcting for chance agreement. Dargatz et al.[18] used kappa to univariately compare the association of resistance to pairs of antimicrobials among *Salmonella* spp. isolates.

Odds ratios, the second measure of association, are typically used to assess the relationship of resistance relative to an exposure of interest (i.e., identifying risk factors). Typically, odds ratios are derived from multivariable logistic regression models, which allow for adjustment for other variables in the model as well as confounding variables and potential interactions. Whitney et al.[76] used logistic regression to examine risk factors for death associated with invasive disease due to *Streptococcus pneumoniae* that were resistant to penicillin among all patients with invasive pneumococcal disease. Their multivariable model included age group, race, state, and hospitalization status. Spika et al.[65] compared the odds of having chloramphenicol-resistant salmonella detected on a dairy relative to exposure to chloramphenicol during the recent past. Odds ratios also can be used for matched sampling designs. Smith et al.[64] used exact conditional logistic

regression to investigate the risk factors, foreign travel and recent use of quinolone, for infection with quinolone-resistant *Campylobacter jejuni* as compared to quinolone-sensitive *C. jejuni*. Relative risk, the measure that the odds ratio is usually intended to approximate, was used to examine the relationship between avoparcin use and the occurrence of vancomycin-resistant *Enterococcus faecium* on swine farms in Denmark[5]. Relative risk and the associated confidence interval, unlike the chi-square test above, were not adjusted for the lack of independence of samples.

A measure that, to our knowledge, has not been applied to antimicrobial resistance data is population attributable fraction (PAF). Population attributable fraction can be defined as "the fraction of the overall occurrence of the outcome that can be attributed uniquely to the presence of the risk factor"[74]. PAF allows for combining prevalence information with odds ratios (or relative risk) into a single measure. The interpretation of PAF results has intuitive appeal. For example, PAF results may suggest that eliminating the effects associated with a specific practice may prevent a portion of the existing antimicrobial resistance. Univariable and multivariable methods for calculating PAF are available and will not be discussed further here [12;30].

Nonparametric statistics have been applied primarily to bacterial counts associated with antimicrobial resistance. Eady et al.[25] determined total viable counts of bacteria on coryneform agar with furazolidone (to inhibit staphylococci) and differential counts with the same media with the addition of either erythromycin (5 mg/L) or tetracycline (10mg/L). They compared bacterial counts for erythromycin-resistant bacteria versus

tetracycline-resistant bacteria from a single location using Wilcoxon's test for matched pairs. They used a Wilcoxon-Mann-Whitney test to compare proportion of resistant bacteria between locations.

Minimum inhibitory concentration data have been handled either descriptively (mean, quantiles, CDFs) or categorically because of the semi-continuous nature of the data. Randall et al.[55] used a Wilcoxon-Mann-Whitney two-sided exact test to compare MIC values between groups of isolates that were either cyclohexane-resistant or cyclohexane-susceptible. A second nonparametric test, the Kolmogorov-Smirnov test, can be used to compare the CDFs of MIC values. Wagner et al.[72] compared the CDFs of isolates obtained from different sampling protocols using the Kolomogorov-Smirnov test.

Categorical and nonparametric methods described above have the principal advantage of making inferences to the sampled populations. The availability of the common statistical procedures including proportion estimation, chi-square tests, and Fisher's exact test on many statistical software packages is another important advantage. Fisher's exact test is appropriate for sparse data in that it uses an exact calculation of the probability-value[30].

Logistic regression is an important analytical method in epidemiology and has several important advantages: 1) allowing for comparison of a dichotomous outcome with one or more independent variables, 2) conversion of regression coefficients to odds ratios that are adjusted for other independent variables in the model, and 3) allowing for examination of relationships between independent variables (e.g. confounding, effect

modification)[35]. Nonparametric tests have the advantage of relaxing the assumption of an underlying assumed probability distribution for the data[16].

An assumption that is common to all of these inferential statistical methods is that of independence of observations. However, independence of observations may be the assumption that is most likely to be violated in studies of antimicrobial susceptibility. For example, bacterial isolates from production animals within herds or flocks, or from the same animal, may have more similar antimicrobial resistance patterns than single isolates from a randomly selected group of animals. This similarity within a group tends to be reflected in smaller variance than might be expected if the grouping had not occurred[49]. Thus, failing to account for the lack of independence can result in smaller standard errors than the true standard errors; this could increase the probability of a type I error (concluding statistical significance is present when, in fact, it was not present). The methods for adjusting for the lack of independence (clustering) will be covered in a later section.

Individually, the procedures have other disadvantages that should also be noted. Categorization of MIC data into SIR data for the purpose of resistance prevalence estimation or chi-square tests may result in loss of information contained in the data. The chi-square test evaluates SIR data for individual antimicrobials. A panel of 17 antimicrobials would require 17 separate tests. Fisher's exact test works well with sparse data in a 2x2 contingency table, but it can be a difficult procedure to implement when contingency tables are larger than 2x2. The standard test for trend (Cochran-Armitage

trend test) considers only a linear trend in the proportion in relation to a quantitative variable and thus does not test for curvilinear trends[30]. The test can be modified to test for the significance of polynomial (curvilinear) trends in the data but we are unaware of this approach being applied to antimicrobial resistance data. Also, tests for trends can only be interpreted when the data collection, testing methodologies, and sampled population are the same over the time frame of interest. The Kappa statistic, a measure of agreement, is affected by the prevalence of the response being measured[30]. McNemar's test is used for bivariate matched data[16]. If there are more than two levels in variables or more than paired observations then alternative methods such as Stuart's chi-square test or Cochran's Q test may be valid alternatives[30].

Multivariate statistics

The use of multivariate statistics in analyzing antimicrobial resistance has been limited. However, the testing of multiple antimicrobials against individual isolates results in multiple outcomes per isolate, which makes multivariate statistics a possible alternative for analysis. Multivariate methods that have been applied include cluster, correspondence, discriminant, and factor analyses.

A hierarchical cluster analysis was used by London et al.[45] based on average linkage (unweighted pair-group method of analysis -UPGMA) and squared Euclidean distances, apparently based on using MIC data. Besides distances, similarities can also be used as a basis for cluster analysis. Johnson and Wichern[39] list a number of similarity coefficients for cluster analysis of binary outcomes. To date, we are unaware of use of these other

similarity measures in the antimicrobial resistance literature. Typically, cluster analysis has been used to group isolates based on genetic testing, such as the presence or absence of specific bands from ribotyping analysis. Boerlin et al.[8] used a Dice coefficient (an ecological index measure) to construct a dendogram using UPGMA. After the dendogram was constructed they attached additional information to each of the isolates including their resistance profile, geographic and animal species origin, and erythromycin resistant methylase B RFLP category.

Correspondence analysis is an exploratory analytical method for describing complex relationships among qualitative outcome variables[21]. Principally, the objective is to summarize the variables in fewer dimensions called ranks and represent the data graphically. Sherley et al.[62] used correspondence analysis to examine host and geographic influences on antimicrobial resistance patterns. The first 2 ranks from their correspondence analysis explained 26.8 percent of the observed variation in resistance from the original panel of 13 antimicrobials. They graphically displayed the rank 1 and rank 2 means and standard deviations for each host species and geographical area of interest. Additionally, they used the values from rank 1 as the dependent variable and host and geographical categories as independent variables in an analysis of variance model.

Discriminant analysis is a multivariate classification procedure where the primary objective is separating groups of objects and allocating new observations to the existing groups[39]. In the context of antimicrobial resistance, the objective of discriminant analysis

has been to examine the potential for creating a discriminant function based on antimicrobial resistance data which can reliably distinguish groups of isolates by sources. In three studies[31;77;78], the objective of discriminant analysis was to identify the source (human or animal) of fecal streptococci based on antimicrobial resistance patterns. The binary outcome of growth/no growth in specified concentrations of the antimicrobials of interest were used to construct the discriminant function.

Factor analysis, unlike discriminant analysis, is oriented toward understanding underlying patterns in data. The method assumes a set of interpretable factors that can be computed as a function of the original variables based on the underlying covariance or correlation structure[21;39]. Wagner et al.[73] used factor analysis to describe underlying patterns in MIC data. They identified six factors, which could be interpreted based on class of antimicrobial, prevalence of susceptibility/resistance, and previously described associations.

When panels of antimicrobials are tested against a group of isolates, it can be very difficult to summarize patterns of association using either SIR or MIC data. Multivariate methods are appropriate for analyzing data with multiple outcomes without the pitfall of over summarizing the data as when indices are implemented. Multivariate methods can be used to reduce the number of outcomes being analyzed. In this sense, the output from multivariate approaches is used in further analysis. Sherley et al.[62] demonstrated this with the use of correspondence results in an analysis of variance model.

There are a number of disadvantages inherent with using multivariate methods to analyze antimicrobial resistance data. First, the methods have not been widely used to analyze antimicrobial data and the audience may have some difficulty in understanding the methods, results, and interpretations. Second, multivariate normality is an assumption of most multivariate techniques[39]. This assumption may not be plausible when MIC data are being used unless the multivariate equivalent of the central limit theorem is invoked[51]. Multivariate measures of central tendency, such as multivariate analysis of variance, may be difficult to interpret in the context of antimicrobial resistance.

Methods for clustered data

Mixed model analysis of variance and other methods for analyzing clustered data have not been widely used to analyze antimicrobial resistance data despite the potential utility of the techniques. A detailed discussion of the need for appropriate design and analytic methods is presented by McDermott et al.[49]. Bager et al.[5] used a chi-square test that was adjusted for within-flock correlation when measuring the association between vancomycin resistance in *Enterococcus faecium* and avoparcin use in poultry. Herrero et al.[32] used a multistage sampling design to investigate the prevalence of vancomycin resistance in *Enterococcus faecium* in pigs. They calculated the design effect associated with their sampling design. The design effect is a measure of the change in the variance due to the sampling design relative to a what the variance would have been under a simple random sampling design[17]. Confidence intervals that they presented for farm-level prevalence were adjusted for the design effect. Dunlop et al.[24] used a logistic regression model to estimate room, pen, and gender effects on the proportion of resistant

E. coli on swine farms. They corrected for clustering, or the overdispersion, by using Finney's correction factor to adjust the likelihood ratio test statistics of two models, one for tetracycline and another for gentamicin, by factors of 47.26 and 13.55. They also adjusted the standard errors associated with their models by factors of 6.87 and 3.681. In both models the differences in the proportion of resistance between rooms of pigs and between genders were not significant when clustering was accounted for by the correction.

Dargatz et al.[18] used a Taylor series expansion to account for clustering when they estimated the variance associated with the proportion of antimicrobial resistance.

Mixed models are another method for accounting, analytically, for the sampling design that has potential for use in analyzing either MIC or SIR data. Mixed models allow for inclusion of both random and fixed effects. Additionally, variance components can be evaluated to assess the importance of sources of variation in the data. Dunlop et al.[24] used a random effects analysis of variance model to examine the proportion of *E. coli* concentrations, both total and tetracycline-resistant, that were attributable to the hierarchical levels associated with their sampling technique. For the proportion of tetracycline-resistant *E. coli*, they showed that the majority of the variance was in the between-pig, within pen component while between-pen, within-room, and between-room components were relatively small[24].

The design-based structure of antimicrobial resistance data is often overlooked during the analysis phase of studies. Methods for analyzing clustered data, principally multilevel models, can account for the structure of the data by accounting for dependencies, random effects and hierarchical nesting[36]. Mixed models can be used to evaluate the contribution of variance from different sources (e.g. herd, flock, pen, and individual). Knowledge of the important sources of variance can be important in designing further studies and considering interventions[24;52;72].

Mixed models, however, can be relatively difficult to understand, implement, and interpret. Many available software packages will allow for more than two levels of nesting in the sampling design when the response is dichotomous but will not produce variance components for more than two levels of data even though, in practice, three or more levels may be of interest (e.g. feedlot, pen, animal, within-animal).

Discussion

Many methods have been used to analyze antimicrobial resistance data. This paper demonstrates the variety of the methods that have been applied to this type of data and discusses some of the advantages and disadvantages. In summary, we believe that there are several issues that should be considered when analyzing antimicrobial resistance data.

First, the conversion of MIC or disk diffusion zone size to SIR data must be considered carefully. The categorization of the data has the advantage of distilling the information into an easier form to be both presented to and assimilated by the intended audience.

However, the summarization of the MIC or disk diffusion zone size can result in the loss of information. We recommend that MIC and zone size data be presented whenever possible. Jones et al.[40] made a similar recommendation because the lack of internationally uniform resistance breakpoints hinders interpretation. They emphasized that presentation of data, such as MIC or zone size, will allow for comparison of data in the future with ease and accuracy. The presentation of MIC and zone size data does not have to exclude the analysis or presentation of SIR data.

A second issue for consideration relates to the assumption of independent observations that is typical of inferential statistical techniques reviewed in this paper[16;30;65]. When multiple isolates are obtained from the same individual or from the same source, such as a hospital or a feedlot cattle pen, the assumption of independence may be inappropriate. For example, Aksaray et al.[2] obtained gram-negative isolates from intensive care units in hospitals in Turkey. A total of 749 isolates were obtained from 473 patients. Of these isolates 17.1 percent were derived from polymicrobial growths on the same occasion and 21.3 percent were obtained from repeat cultures, which were interpreted as persistent colonization. If observations are assumed to be independent but actually have some correlation, then the variance estimates are likely to be too small[49]. Underestimating variance may lead to statistically accepting the research hypothesis in cases where it should not be accepted (i.e., rejecting the null hypothesis). The degree to which the lack of independence can alter the results is evident by the size of the adjustment factors that were employed by Dunlop et al.[24]. Methods used for inferential analysis of antimicrobial resistance data should account for the lack of independence or correlated data structures,

when present, to avoid erroneous conclusions. Many of the common analytical methods, including prevalence estimation and chi-square tests, can be implemented to account for the clustering either through post-hoc variance adjustments or through alternative variance estimation procedures.

An important issue in antimicrobial resistance is the occurrence of multiple resistant bacterial strains. Almost exclusively, patterns of multiple drug resistance have been examined using SIR data, which is usually collapsed into resistant/non-resistant to create antibiograms or multi-resistant phenotypic descriptions. Multivariate methods provide an alternative method to assess patterns of either resistance or susceptibility using MIC data. We recommend that analysts consider use of multivariate techniques for exploring and describing potentially complex relationships among antimicrobials. The use of multivariate techniques should not preclude other methods of presenting multi-resistant results that may be more widely understood.

The last issue for consideration is the selection of the analytical methodology. The selection of analytical methods should reflect clearly stated research goals, be appropriate for the study design and data structure, and be appropriate for the intended audience. These criteria make it implausible that a single analytical method will be most appropriate for all situations. Similarly, there likely will be a need to use several analytical methods to display, summarize, and test hypothesis to meet these criteria.

---

[a] SAS, version 8.1, SAS Institute Inc, Cary, NC.

Reference List

1. Aarestrup FM, Agerso Y, Ahrens P, et al. Antimicrobial susceptibility and presence of resistance genes in staphylococci from poultry. *Vet Microbiol* 2000;74:353-64.

2. Aksaray S, DokuzoGuz B, Guvener E, et al. Surveillance of antimicrobial resistance among gram-negative isolates from intensive care units in eight hospitals in Turkey. *J Antimicrob Chemother* 2000;45:695-9.

3. Aminov RI, Garrigues-Jeanjean N, Mackie RI. Molecular ecology of tetracycline resistance: Development and validation of primers for detection of tetracycline resistance genes encoding ribosomal protection proteins. *Appl. Environ. Microbiol.* 2001;67:22-32.

4. Arvand M, Hoeck M, Hahn H, et al. Antimicrobial resistance in Streptococcus pyogenes isolates in Berlin. *J. Antimicrob. Chemother.* 2000;46:621-623.

5. Bager F, Madsen M, Christensen J, et al. Avoparcin used as a growth promoter is associated with the occurrence of vancomycin-resistant Enterococcus faecium on Danish poultry and pig farms. *Prev Vet Med* 1997;31:95-112.

6. Barisic Z, Punda-Polic V. Antibiotic resistance among enterococcal strains isolated from clinical specimens. *Int. J. Antimicrob. Agents* 2000;16:65-68.

7. Benbachir M, Benredjeb S, Boye CS, et al. Two-year surveillance of antibiotic resistance in Streptococcus pneumoniae in four African cities. *Antimicrob Agents Chemother* 2001;45:627-9.

8. Boerlin P, Burnens AP, Frey J, et al. Molecular epidemiology and genetic linkage of macrolide and aminoglycoside resistance in Staphylococcus intermedius of canine origin. *Vet Microbiol* 2001;79:155-69.

9. Breuil J, Brisabois A, Casin I, et al. Antibiotic resistance in salmonellae isolated from humans and animals in France: comparative data from 1994 and 1997. *J Antimicrob Chemother* 2000;46:965-71.

10. Brophy PO, Caffrey PJ, Collins JD. Sensitivity patterns of Escherichia coli isolated from calves during and following prophylactic chlortetracycline therapy. *Br Vet J* 1977;133:340-5.

11. Bruun MS, Schmidt AS, Madsen L, et al. Antimicrobial resistance patterns in Danish isolates of Flavobacterium psychrophilum. *Aquaculture* 2000;187:201-212.

12. Bruzzi P, Green SB, Byar DP, et al. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol* 1985;122:904-914.

13. Butaye P, Devriese LA, Haesebrouck F. Differences in Antibiotic Resistance Patterns of Enterococcus faecalis and Enterococcus faecium Strains Isolated from Farm and Pet Animals. *Antimicrob Agents Chemother* 2001;45:1374-8.

14. Bywater RJ, Casewell MW. An assessment of the impact of antibiotic resistance in different bacterial species and of the contribution of animal sources to resistance in human infections. *J Antimicrob Chemother* 2000;46:643-645.

15. Cavallo JD, Fabre R, Leblanc F, et al. Antibiotic susceptibility and mechanisms of beta-lactam resistance in 1310 strains of pseudomonas aeruginosa: a french multicentre study (1996). *J Antimicrob Chemother* 2000;46:133-6.

16. Conover WJ. *Practical nonparametric statistics, second edition.* ed. New York: John Wiley and Sons, 1980;

17. Dargatz DA, Hill GW. Analysis of survey data. *Prev Vet Med* 1996;28:225-237.

18. Dargatz DA, Fedorka-Cray PJ, Ladely SR, et al. Antimicrobial susceptibility patterns of *Salmonella* isolates from cattle in feedlots. *J Am Vet Med Assoc* 2002;221:268-272.

19. Davis MA, Hancock DD, Besser TE, et al. Changes in antimicrobial resistance among Salmonella enterica Serovar typhimurium isolates from humans and cattle in the Northwestern United States, 1982-1997. *Emerg Infect Dis* 1999;5:802-6.

20. Dawson KA, Boling JA. Monensin-resistant bacteria in the rumens of calves on monensin- containing and unmedicated diets. *Appl Environ Microbiol* 1983;46:160-4.

21. Dohoo IR, Ducrot C, Fourichon C, et al. An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Prev Vet Med* 1997;29:221-39.

22. Donner A. The comparison of proportions in the presence of litter effects. *Prev Vet Med* 1993;18:17-26.

23. Donskey CJ, Chowdhry TK, Hecker MT, et al. Effect of antibiotic therapy on the density of vancomycin-resistant enterococci in the stool of colonized patients. *N Engl J Med* 2000;343:1925-1932.

24. Dunlop RH, McEwen SA, Meek AH, et al. Sampling considerations for herd-level measurement of faecal *Escherichia coli* antimicrobial resistance in finisher pigs. *Epidemiol Infect* 1999;122:485-496.

25. Eady EA, Coates P, Ross JI, et al. Antibiotic resistance patterns of aerobic coryneforms and furazolidone-resistant Gram-positive cocci from the skin surface of the human axilla and fourth toe cleft. *J Antimicrob Chemother* 2000;46:205-213.

26. Ebner PD, Mathew AG. Effects of antibiotic regimens on the fecal shedding patterns of pigs infected with Salmonella typhimurium. *J Food Prot* 2000;63:709-714.

27. Engberg J, Aarestrup FM, Taylor DE, et al. Quinolone and macrolide resistance in Campylobacter jejuni and C. coli: resistance mechanisms and trends in human isolates. *Emerg Infect Dis* 2001;7:24-34.

28. Fey PD, Safranek TJ, Rupp ME, et al. Ceftriaxone-resistant salmonella infection acquired by a child from cattle. *N Engl J Med* 2000;342:1242-9.

29. Finch RG. Antibiotic resistance. *J Antimicrob Chemother* 1998;42:125-8.

30. Fleiss, JL. Statistical methods for rates and proportions. 1981. New York, John Wiley & Sons, Inc.

31. Harwood VJ, Whitlock J, Withington V. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: Use in predicting the source of fecal contamination in subtropical waters. *Appl Environ Microbiol* 2000;66:3698-3704.

32. Herrero IA, Teshager T, Garde J, et al. Prevalence of vancomycin-resistant Enterococcus faecium (VREF) in pig faeces from slaughterhouses in Spain. *Prev Vet Med* 2000;47:255-62.

33. Hinton M, Hedges AJ, Linton AH. The ecology of Escherichia coli in market calves fed a milk-substitute diet. *J Appl Bacteriol* 1985;58:27-35.

34. Holmberg SD, Osterhom MT, Senger KA, et al. Drug-resistant Salmonella from animals fed antimicrobial. *N. E. Jour. Med.* 1984;311:617-622.

35. Hosmer DW, Lemeshow S . *Applied logistic regression*. New York: Wiley, 1989

36. Hox JJ, I.G.G. Kreft. Multilevel analysis methods. *Sociological Methods & Research* 1994;22:283-299.

37. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* 1988;26:2465-6.

38. Hurlbert SH. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology* 1971;52:577-856.

39. Johnson RA, D.W. Wichern. *Applied multivariate statistical techniques, third edition*. New Jersey: Prentice-Hall, Inc., 1992

40. Jones RN. Detection of emerging resistance patterns within longitudinal surveillance systems: data sensitivity and microbial susceptibility. MYSTIC Advisory Board. Meropenem Yearly Susceptibility Test Information Collection. *J Antimicrob Chemother* 2000;46 Suppl T2:1-8.

41. Kapperud G, Skjerve E, Vik L, et al. Epidemiological investigation of risk factors for campylobacter colonization in Norwegian broiler flocks. *Epidemiol Infect* 1993;111:245-255.

42. Kaspar CW, Burgess JL, Knight IT, et al. Antibiotic resistance indexing of Escherichia coli to identify sources of fecal contamination in water. *Can J Microbiol* 1990;36:891-4.

43. Kricher JC. Bird Species Diversity: The Effect of Species Richness and Equitability on the Diversity Index. *Ecology* 1972;53:278-282.

44. Krumperman PH. Multiple antibiotic resistance indexing of Escherichia coli to identify high-risk sources of fecal contamination of foods. *Appl Environ Microbiol* 1983;46:165-70.

45. London N, Nijsten R, Bogaard Avd, et al. Carriage of antibiotic-resistant *Escherichia coli* by health volunteers during a 15-week period. *Infection* 1994;22:187-192.

46. Low DE. Appropriate antibiotic use: reducing the emergence and dissemination of resistance. *Int J Clin Pract* 2000;106-110.

47. Malorny B, Schroeter A, Helmuth R. Incidence of quinolone resistance over the period 1986 to 1998 in veterinary Salmonella isolates from Germany. *Antimicrob Agents Chemother* 1999;43:2278-82.

48. Mathew AG, Saxton AM, Upchurch WG, et al. Multiple antibiotic resistance patterns of Escherichia coli isolates from swine farms. *Appl Environ Microbiol* 1999;65:2770-2.

49. McDermott JJ , Y.H. Schukken, M.M. Shoukri. Study design and analytical methods for data collected from clusters of animals. *Prev Vet Med* 1994;18:175-191.

50. Millar MR, Walsh TR, Linton CJ, et al. Carriage of antibiotic-resistant bacteria by healthy children. *J Antimicrob Chemother* 2001;47:605-10.

51. Morrison DF. *Multivariate statistical methods.* New York, New York: McGraw-Hill, Inc., 1976

52. National Antimicrobial Resistance Monitoring Program - Enteric Bacteria. Veterinary Isolates Final Report. *Athens, GA: USDA:ARS* 1997;20-26.

53. Normand EH, Gibson NR, Reid S, et al. Antimicrobial-resistance trends in bacterial isolates from companion-animal community practice in the UK. *Prev Vet Med* 2000;46:267-278.

54. Rajashekara G, Haverly E, Halvorson DA, et al. Multidrug-resistant Salmonella Typhimurium DT104 in poultry. *J Food Prot* 2000;63:155-161.

55. Randall LP, Cooles SW, Sayers AR, et al. Association between cyclohexane resistance in Salmonella of different serovars and increased, resistance to multiple antibiotics, disinfectants and dyes. *J Med Microbiol* 2001;50:919-924.

56. Rice LB. Emergence of vancomycin-resistant enterococci. *Emerg Infect Dis* 2001;7:183-7.

57. Saenz Y, Zarazaga M, Lantero M, et al. Antibiotic resistance in Campylobacter strains isolated from animals: Foods, and humans in Spain in 1997-1998. *Antimicrob. Agents Chemother.* 2000;44:267-271.

58. Sahm DF, Jones ME, Hickey ML, et al. Resistance surveillance of Streptococcus pneumoniae, Haemophilus influenzae and Moraxella catarrhalis isolated in Asia and Europe, 1997- 1998. *J Antimicrob Chemother* 2000;45:457-66.

59. Sahm DF, Karlowsky JA, Kelly LJ, et al. Need for annual surveillance of antimicrobial resistance in Streptococcus pneumoniae in the United States: 2-year longitudinal analysis. *Antimicrob. Agents Chemother.* 2001;45:1037-1042.

60. Sanches IS, Mato R, De Lencastre H, et al. Patterns of multidrug resistance among methicillin-resistant hospital isolates of coagulase-positive and coagulase-negative staphylococci collected in the International Multicenter Study RESIST in 1997 and 1998. *Microb Drug Resist.-Mechan Epidemiol Dis.* 2000;6:199-211.

61. Seyfarth AM, Wegener HC, Frimodt-Moller N. Antimicrobial resistance in Salmonella enterica subsp. enterica serovar typhimurium from humans and production animals. *J Antimicrob Chemother* 1997;40:67-75.

62. Sherley M., D.M. Gordan, P.J. Collignon. Variations in antibiotic resistance profile in Enterobateriaceae isolated from wild Australian mammals. *Envir Micro* 2000;2:620-631.

63. Simpson EH. Measurement of diversity. *Nature* 1949;163:688

64. Smith KE, Besser JM, Hedberg CW, et al. Quinolone-resistant Campylobacter jejuni infections in Minnesota, 1992- 1998. Investigation Team. *N Engl J Med* 1999;340:1525-32.

65. Spika JS, Waterman SH, Hoo GW, et al. Chloramphenicol-resistant Salmonella newport traced through hamburger to dairy farms. A major persisting source of human salmonellosis in California. *N Engl J Med* 1987;316:565-70.

66. Stephan R, Schumacher S. Resistance patterns of non-O157 Shiga toxin-producing Escherichia coli (STEC) strains isolated from animals, food and asymptomatic human carriers in Switzerland. *Lett. Appl. Microbiol.* 2001;32:114-117.

67. Stobberingh EE, van den Bogaard AE. Spread of antibiotic resistance from food animals to man. *Acta Vet Scand* 2000;47-50,51-52.

68. Tackett CO, Dominguez LB, Fisher HJ, et al. An outbreak of multiple-drug-resistant *Salmonella* Enteritis from raw milk. *J Am Med Assoc* 1985;253:2058-2060.

69. Thornsberry C, Jones ME, Hickey ML, et al. Resistance surveillance of Streptococcus pneumoniae, Haemophilus influenzae and Moraxella catarrhalis isolated in the United States, 1997-1998. *J Antimicrob Chemother* 1999;44:749-59.

70. Threlfall EJ , Ward LR, Frost JA, et al. The emergence and spread of antibiotic resistance in food-borne bacteria. *Int. J. Food Microbiol.* 2000;62:1-5.

71. Tsolia M, Kouppari G, Zaphiropoulou A, et al. Prevalence and patterns of resistance of Streptococcus pneumoniae strains isolated from carriers attending day care centers in the area of Athens. *Microb Drug Resist-Mechan Epidemiol Dis* 1999;5:271-278.

72. Wagner B.A., D.A. Dargatz, M.D. Salman, et al. Sampling Techniques for Antimicrobial Susceptibility of Enteric *Escherichia coli* Recovered From Feedlot Cattle. *Am J Vet Res* 2002;63:1662-1670.

73. Wagner BA, M.D. Salman, D.A. Dargatz, et al. In press. Factor analysis of minimum inhibitory concentrations for *Escherichia coli* isolated from feedlot cattle to model relationships among antimicrobial outcomes. *Prev Vet Med*

74. Wagner BA, Wells SE, Kott PS. Variance estimation for population attributable risk in a complex cross-sectional animal health survey. Prev Vet Med 2001; 48(1):1-13

75. White DG, McDermott PF. Emergence and transfer of antibacterial resistance. *J Dairy Sci.* 2001;84:E151-E155

76. Whitney CG, Farley MM, Hadler J, et al. Increasing prevalence of multidrug-resistant Streptococcus pneumoniae in the United States. *N Engl J Med* 2000;343:1917-24.

77. Wiggins BA. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. *Appl Environ Microbiol* 1996;62:3997-4002.

78. Wiggins BA, Andrews RW, Conway RA, et al. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *Appl Environ Microbiol* 1999;65:3483-6.

79. Wiuff C, Madsen M, Baggesen DL, et al. Quinolone resistance among Salmonella enterica from cattle, broilers, and swine in Denmark. *Microb Drug Resist* 2000;6:11-7.

80. Zhanel GG, Karlowsky JA, Palatnick L, et al. Prevalence of antimicrobial resistance in respiratory tract isolates of Streptococcus pneumoniae: Results of a Canadian national surveillance study (vol 43, pg 2504, 1999). *Antimicrob. Agents Chemother.* 2000;44:3247

CHAPTER 3

A COMPARISON OF SAMPLING TECHNIQUES FOR MEASURING THE ANTIMICROBIAL SUSCEPTIBILITY OF ENTERIC *ESCHERICHIA COLI* RECOVERED FROM FEEDLOT CATTLE

Summary

The objective of this study was to evaluate the effectiveness of various sampling methodologies for determining antimicrobial resistance patterns in *Escherichia coli* isolated from feces of feedlot cattle. Fecal samples obtained from 328 beef steers and 6 feedlot pens in which they resided. Single fecal samples were collected from the rectum of each steer and from floors of pens in which the cattle resided. Fecal material from each single sample was combined into pools containing 5 and 10 samples. Five isolates of *Escherichia coli* from each single sample and each pooled sample were tested for susceptibility to 17 antimicrobial drugs.

Patterns of antimicrobial resistance for fecal samples obtained from the rectum of cattle did not differ from fecal samples obtained from pen floors. Resistance patterns from pooled samples differed from patterns observed for single fecal samples. Little pen-to-pen variation in resistance prevalence was observed. Clustering of resistance phenotypes within samples was detected.

Studies of antimicrobial resistance in feedlot cattle can rely on fecal samples obtained from pen floors, thus avoiding the cost and effort of obtaining rectal samples. Pooling of fecal samples yielded resistance patterns that were consistent with those of single fecal samples when the prevalence of resistance to an antimicrobial was > 2%. Pooling may be a practical alternative when investigating patterns of resistance that are not rare. Apparent clustering of resistance phenotypes within samples argues for examining fewer isolates per fecal sample and more fecal samples per pen.

Introduction

Food producers and allied industries in the United States have come under increasing scrutiny because of concerns over the role antimicrobial use in various production systems may play in adverse health outcomes in humans as a result of antimicrobial-resistant bacteria. This scrutiny is motivating researchers to determine resistance patterns at various points along the animal production chain, in humans, and in the environment.[1-6] Additionally, focus is increasing on the utility of antimicrobial resistance monitoring and surveillance to evaluate trends and make information available to decision makers.[5;7-11]

Determination of the prevalence of antimicrobial resistance in bacteria from production animals, whether in a research or monitoring-surveillance context, is dependent on the ability to accurately and efficiently measure resistance. Measuring characteristics of antimicrobial resistance in bacterial populations, whether related to minimum inhibitory

concentration (MIC) or its categorized counterpart, susceptible-intermediate-resistant (SIR), can be strongly influenced by the sampling strategy.[12]

A number of variables can be considered when designing a sampling strategy for determining antimicrobial resistance in feedlot cattle. Perhaps the most obvious variable is the number of fecal samples to obtain from a study population to meet the precision requirements of the design. Furthermore, depending on the prevalence of the bacteria being studied, more than a single bacterial isolate can be obtained from a single fecal sample. Limited resources lead to a tradeoff between number of fecal samples collected and number of isolates tested per sample. Although several studies have involved examination of 2 to 4 isolates/fecal sample, Vosti et al[13] investigated the impact of isolating larger numbers of *Escherichia coli* isolates for each human fecal sample. They found that even though 15 to 25 isolates were examined, nearly 80% of serologic groups detected were identified after evaluation of the first 5 isolates. Another practical consideration is the method of sample collection. Feces can be collected directly from the rectum (swab specimen or a volume of feces) or from the pen floor. In most cattle feedlots, collection of fecal samples from the rectum of each animal would require restraining animals, whereas collection of samples from pen floors does not. Hoar et al[14]discussed the advantages of collecting fecal samples off the ground, compared with from the rectum, and the importance of comparing microbial measurements obtained for the 2 collection methods.

Other studies of antimicrobial resistance of *E. coli* have used various sampling approaches. Hinton et al.[15] used a swab to collect fecal samples from the rectum of each calf. Ten colonies from each sample were selected and tested against a panel of antimicrobial agents. Mercer et al.[16] also used a swab specimen obtained from the rectum when collecting samples from cattle and swine, but they examined 3 isolates/sample. Bourgue et al.[17] used methods similar to those of another study[15] to examine antimicrobial resistance in *E. coli* recovered from swine. In another study,[18] investigators tested 10 to 20 *E. coli* isolates from human fecal samples. It has been suggested[19] that it is necessary to examine at least 10 colonies from individual animals in a population to represent the diversity of serotypes of *E. coli* in that population.

The objective of the study reported here was to use non-type-specific *E. coli* to determine the effectiveness of various sampling methods and techniques for determining antimicrobial resistance patterns in feces recovered from feedlot cattle. Specifically, our intention was to evaluate the appropriateness of collecting a single fecal sample from the pen floor versus collecting a single sample from the rectum of each animal. Furthermore, antimicrobial resistance patterns for isolates cultured from individual fecal samples with those obtained from pooled fecal samples were compared to evaluate variation in test results on the basis of the various sample collection techniques.

Materials and methods

Study population:

Fecal samples were collected from 328 beef steers housed in 6 pens at a commercial feedlot in Colorado. Each pen housed between 54 and 56 steers and had an area of approximately 930 square meters. Steers in 3 pens were all from a single herd and had been at the feedlot for 75 days, whereas steers in the other 3 pens were from another single herd and had been at the feedlot for 81 days. All steers were fed tylosin in their rations at a rate of 10 g/907.2 kg of feed (approx 60 to 90 mg/steer/d). None of the steers received an antimicrobial injection during the month prior to collection of samples, except 1 steer that was injected with tilmicosin during the week prior to collection of fecal samples.

Collection and processing of samples:

Fecal samples (approx 50 g) were collected from the interior of 180 fresh fecal pats found on the floor of the pens (30 pats/pen). Disposable plastic gloves were used to pick up fecal material; gloves were changed between samples. Fecal samples from the pen floor could not be associated with a specific steer; however, efforts were made to collect fecal samples from sites throughout each pen to minimize the possibility of collecting multiple fecal samples from the same steer. While fecal samples were being collected from the floor of a pen, fecal samples were concurrently obtained from the rectum of 30 randomly selected steers from that pen. Steers were restrained in a chute for collection of fecal samples; plastic disposable gloves were used to obtain the samples. All samples were cooled and transported to the laboratory for further processing, which was accomplished within 4 hours after collection.

In the laboratory, tongue depressors were used to transfer fecal material to 50-mL conical tubes. Approximately 1 g of fecal material from each individual sample (collected from the rectum or from the pen floor) was placed in each tube. Additionally, pooled fecal samples were created in the laboratory; pooled samples contained 5 or 10 individual samples. Using a formal randomization scheme, 5 selected fecal samples each contributed approximately 0.2 g of fecal material to a single pool and 10 selected fecal samples each contributed approximately 0.1 g of fecal material to a single pool. As such, each pooled sample consisted of approximately 1 g of fecal material. Therefore, each pen was represented by 78 samples (30 individual samples collected from the rectum of 30 steers, 30 individual samples collected from the pen floor, 6 pooled samples consisting of samples obtained rectally from 5 steers, 6 pooled samples consisting of 5 samples obtained from the pen floor, 3 pooled samples consisting of samples collected rectally from 10 steers, and 3 pooled samples consisting of 10 samples obtained from the pen floor). Ten milliliters of saline (0.9% NaCl) solution was added to each tube to facilitate mixing of sample material. Tubes were agitated by use of a vortex to thoroughly mix the saline solution and fecal material. All individual and pooled samples were randomly assigned unique identification numbers to mask sample identity at the laboratory. Tubes were then stored overnight in a refrigerator at 3°C. On the subsequent day, samples were shipped via overnight delivery to a laboratory where culturing and antimicrobial susceptibility testing were conducted.

Culture and antimicrobial susceptibility testing:

Dilute fecal material was inoculated onto MacConkey-4-Methylumbelliferyl-β-D-Glucuronide agar plates,[a] and plates were incubated to provide colonies of *E. coli*. *Escherichia coli* was chosen as the species for antimicrobial susceptibility testing because of its ubiquitous nature. After incubation, individual *E. coli* colonies were identified by examination under UV light as lactose-positive (bright pink) and glucuronidase positive (colony periphery had a bluish appearance) colonies. Five colonies were selected from each agar plate, and each colony was transferred to a separate nutrient agar slant and incubated for another 18 to 24 hours. In preparation for antimicrobial susceptibility testing, MacConkey agar plates were inoculated with bacteria from the slants. Colonies were selected from these plates and placed into separate tubes with 5 mL of sterile water; turbidity was adjusted to a 0.5 McFarland standard. After mixing, 10 μL of bacterial suspension was used to inoculate 10 mL of cation-adjusted Mueller-Hinton broth, which was then used to inoculate 96-well plates containing various quantities of antimicrobial drugs for susceptibility testing. Plates were incubated at 35°C for 18 to 24 hours and then evaluated by use of a semiautomated broth dilution system.[b] On the basis of growth in individual wells, the MIC was determined for each of 17 antimicrobial drugs (amikacin, amoxicillin-clavulanic acid, ampicillin, apramycin, cefoxitin, ceftiofur, ceftriaxone, cephalothin, chloramphenicol, ciprofloxacin, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfamethoxazole, tetracycline, trimethoprim-sulfamethoxazole). These 17 antimicrobials were selected to parallel the panel used by the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS-EB).[20] The MIC for each antimicrobial was used to classify the isolate as susceptible, intermediate, or resistant in accordance with standards used to classify human isolates as determined by the National

Committee on Clinical Laboratory Standards and as used by the NARMS-EB surveillance program. Currently, veterinary breakpoints for many of the antimicrobials are not available.[21] The MIC values were truncated at the limits of the concentrations in the wells.

Statistical analysis:

Analyses were performed by use of 3 statistical programs.[c-e] Statistical analyses were performed on categorized data (SIR results) and MIC outcomes for antimicrobial susceptibility testing. For some analyses, intermediate results were grouped with susceptible results. The proportion of isolates resistant to each of the 17 antimicrobials and the associated SE was determined by use of statistical software.[d] The survey procedures in that program allowed for calculation of SE by use of the delta method (first-order Taylor series approximation) to account for the lack of independence of multiple isolates obtained from a sample and for the nesting of samples within pen.[22] Unless specified, all analyses performed accounted for the sampling design of the study.

A multiple-level model was used to examine sources of variation in antimicrobial resistance. Binomial response models were constructed for each of the 10 antimicrobials for which sufficient resistance occurred to allow for modeling. A restricted maximum-likelihood estimate was obtained for a simple random intercept model containing only a constant, using the restricted generalized iterative least-squares algorithm.[e] In this binary model, variance for level 1 (isolate level) was constrained such that it was equal to 1 (ie, extra-binomial variation was not permitted). As described elsewhere,[23] the proportion of

variance was calculated at each level, assuming that the level-1 variance on the logit scale was 3.29 (i.e., $\pi^2/3$).

Logistic regression models, which were essentially an extension of the intercept-only model used for multiple-level modeling, were constructed for each of the 8 antimicrobials to which *E. coli* were most commonly resistant (ampicillin, cephalothin, chloramphenicol, kanamycin, nalidixic acid, streptomycin, sulfamethoxazole, and tetracycline) by use of 1 of the statistical programs.[22,d] The outcome was susceptibility or resistance to each antimicrobial. The analysis was performed to evaluate the effects of pooling fecal samples and to compare individual samples from steers with individual samples from pen floors. Base levels (reference levels) for the logistic regression model were individual (nonpooled) samples and samples obtained per rectum. Isolates that were resistant to other antimicrobials were identified at a low frequency that did not allow modeling because of model instability.

Number of antimicrobials to which each isolate was resistant was determined. A χ^2 test was performed[22,d] to compare the distribution of multiple-resistant isolates between the factors of interest (individual samples from each steer versus individual samples from each pen floor, individual samples obtained rectally versus pooled samples of fecal samples obtained from 5 and 10 steers, and individual samples obtained from the pen floor versus pooled samples of 5 and 10 fecal samples from the pen floor).

An antibiogram was constructed to describe the phenotypic resistance patterns for each treatment combination. Intermediate and susceptible test results were combined for this analysis.

Clustering of antimicrobial phenotypes within individual samples (steer and pen floor samples) and within pools consisting of 5 samples was explored separately by use of simulated resampling of isolates. Pools consisting of 10 samples were not evaluated, because the number of pools per pen (n = 3) was considered too small. First, the actual number of antimicrobial phenotypes expressed within the 5 isolates of each individual or pooled sample was determined (value ranged from 1 to 5). For example, if all 5 isolates from a sample were susceptible to all antimicrobials tested, then the isolates represented a single antimicrobial phenotype, and the sample was assigned a value of 1. If all 5 isolates from a single sample had differing resistance phenotypes, then the sample was assigned a value of 5. The frequency distribution for the number of phenotypes identified in samples was constructed for the observed results for individual samples and the observed results for pooled samples consisting of 5 fecal samples. Next, simulated distributions were constructed to represent the distribution of the number of phenotypes that would have been expected if all the isolates had been randomly distributed. All isolates were sorted according to these randomly assigned numbers. The sorted list was divided sequentially into groups of 5 isolates such that each set of 5 isolates was considered to be equivalent to a sample. The number of unique phenotypes (phenotype count) was determined for each simulated group of 5 isolates. Frequency distribution for the phenotype count was then determined for the simulated samples. The simulation was repeated 1,000 times for

the individual samples and 1,000 times for the pooled samples consisting of 5 fecal samples. Mean percentage of samples in each category of phenotype count from the 1,000 simulations was compared with the observed frequency distribution by use of a χ^2 goodness-of-fit test that was not adjusted for the sample collection design. In the actual summarization and the simulation analysis, intermediate results were not grouped with susceptible results.

Analysis of MIC data included calculation of descriptive measures such as the 50th percentile MIC value (MIC_{50}) and geometric mean. Additionally, cumulative distribution functions were compared between factors of interest by use of the Kolmogorov-Smirnov test.[24] Significance was set at $\alpha = 0.05$ for all statistical tests in the study.

Results

A total of 2,316 isolates obtained from individual samples collected from steers, individual samples collected from pen floors, pooled samples from steers, and pooled samples from pen floors were tested (Table 3.1). The number of isolates tested reflects a loss of 24 isolates during processing. The proportion of *E. coli* isolates resistant to the antimicrobial drugs tested varied among samples collected from the rectum of individual steers, individual samples collected from the pen floors, and pooled samples. All isolates were susceptible to 4 antimicrobials (amikacin, apramycin, ceftriaxone, and gentamicin). Only 1 isolate from a pooled sample of feces collected from a pen floor was resistant to ciprofloxacin. Isolates from samples collected per rectum from individual steers and individual samples collected from the pen floor were resistant to 12 and 11

Table 3.1. Comparison of the proportion of *Escherichia coli* isolates from individual and pooled fecal samples obtained from the rectum of steers or from the pen floor with resistance to each of 17 antimicrobials

Antimicrobial	Individual samples				Pool of 5 samples				Pool of 10 samples			
	Rectum (n = 885)		Pen floor (n = 894)		Rectum (n =179)		Pen floor (n = 179)		Rectum (n = 89)		Pen floor (n = 90)	
Amikacin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Amoxicillin-clavulanic acid	0.9	(0.4)	0.6	(0.3)	1.1	(0.7)	0.6	(0.6)	0.0	(0.0)	0.0	(0.0)
Ampicillin	1.9	(0.7)	1.3	(0.7)	1.1	(0.8)	2.8	(2.2)	0.0	(0.0)	0.0	(0.0)
Apramycin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Cefoxitin	0.7	(0.4)	0.2	(0.2)	0.0	(0.0)	2.2	(2.2)	0.0	(0.0)	0.0	(0.0)
Ceftiofur	0.1	(0.1)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Ceftriaxone	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Cephalothi	1.1	(0.5)	0.6	(0.3)	1.1	(0.8)	1.7	(1.7)	0.0	(0.0)	0.0	(0.0)
Chloramphenicol	1.1	(0.5)	2.2	(0.8)	1.7	(1.2)	2.8	(2.2)	0.0	(0.0)	3.3	(3.2)
Ciprofloxacin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.6	(0.6)	0.0	(0.0)	0.0	(0.0)
Gentamicin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Kanamycin	0.7	(0.4)	1.5	(0.6)	2.8	(1.8)	1.1	(0.8)	0.0	(0.0)	5.6	(5.4)
Nalidixic acid	1.2	(0.5)	0.7	(0.5)	1.7	(1.2)	0.6	(0.6)	0.0	(0.0)	1.1	(1.1)
Streptomycin	12.7	(1.8)	10.3	(1.6)	9.5	(3.0)	12.8	(3.9)	9.0	(2.8)	7.8	(4.8)
Sulfamethoxazole	22.4	(2.3)	23.7	(2.5)	25.7	(5.2)	31.3	(5.6)	16.9	(5.5)	20.0	(7.0)
Tetracycline	31.2	(2.5)	30.6	(2.7)	31.8	(5.8)	42.5	(6.1)	28.1	(6.1)	28.9	(7.7)
Trimethoprim	0.1	(0.1)	0.1	(0.1)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)

Numbers in parentheses represent SE.

antimicrobials, respectively. Isolates from pooled fecal samples from 5 steers and pooled feces from 5 samples obtained from pen floors were resistant to 9 and 11 antimicrobials, respectively. Isolates from pooled fecal samples from 10 steers and pooled feces from 10 samples obtained from pen floors were resistant to only 3 and 6 antimicrobials, respectively. Each of the sample collection and pooling methods were able to document resistance to the 3 antimicrobials with the highest prevalence of resistance (streptomycin, sulfamethoxazole, and tetracycline). Resistance phenotypes associated with lower prevalences (< 2.0%), as estimated with individual samples, were not detected in the pools of 10 samples collected per rectum and were only variably detected in the pools of 10 samples collected from pen floors.

Results of multiple-level modeling suggested the proportion of variance in the occurrence of the 10 antimicrobials explained at the pen level did not exceed 0.066 (Table 3.2). The proportion of variation in the resistance proportion at the fecal sample level essentially divided the 10 antimicrobials into 2 groups: proportion of variance for 1 group was approximately 0.8 to 0.9, and proportion of variance for the second group was approximately 0.3 to 0.5 (tetracycline, streptomycin, and sulfamethoxazole). This grouping of antimicrobials was also evident by examining the proportion of variation that was unexplained (within-sample variation). The group of antimicrobials with the higher proportion of variance explained at the fecal sample level had a lower prevalence (1 to 3%) than the other group (8 to 40%).

Table 3.2. Proportion of total variation in antimicrobial resistance to each of 10 antimicrobials at the level of pen, fecal sample, and within fecal sample.

Antimicrobial	Proportion of total variation		
	Pen	Fecal sample	Within fecal sample
Amoxicillin-clavulanic acid	0.016	0.818	0.166
Ampicillin	0.023	0.827	0.150
Cefoxitin	0.023	0.887	0.090
Cephalothin	0.010	0.860	0.130
Chloramphenicol	0.066	0.775	0.159
Kanamycin	0.019	0.868	0.113
Nalidixic acid	0.009	0.872	0.118
Tetracycline	0.022	0.340	0.638
Streptomycin	0.028	0.476	0.496
Sulfamethoxazole	0.005	0.394	0.601

Odds ratios derived from the logistic regression model were compared relative to predefined base values (Table 3.3). Thus, the odds ratios represent the odds of isolates being detected as resistant to each of the 8 antimicrobials examined relative to nonpooled samples (all individual samples obtained from steers and pen floors) and all samples obtained from the rectum of steers (individual and pooled samples). Each odds ratio was adjusted for the other variable (pooled feces from 5 and 10 samples and all samples obtained from the pen floors, respectively) in the model. None of the odds ratios from the 8 separate antimicrobial logistic models was significantly different from unity indicating that there was no difference in the odds of an isolate being resistant regardless of whether it came from the rectum of a steer, a pen floor, an individual sample, or pooled fecal samples. Although not significant, 7 of 8 odds ratio estimates for pooled feces from 5 samples were less than unity suggesting a pattern toward reduced likelihood of resistance in the pooled samples relative to individual samples. In contrast, 5 of 6 odds ratios for pooled feces from 10 samples were greater than unity. Odds ratios for samples obtained from the pen floors were equally distributed around unity, inferring that the likelihood of resistance for isolates obtained from samples collected from pen floors was similar to that for isolates obtained from samples collected directly from the rectum of steers.

A majority (65.1%) of all isolates were susceptible to all 17 antimicrobials. Resistance to as many as 8 antimicrobials (ie, multiple resistance count = 8) was observed in single isolates (0.3%). Pooled feces from 5 steers and pooled feces from 10 pen floor samples did not have any isolates with multiple resistance to more than 4 drugs, whereas the

Table 3.3. Odds ratios for resistance to individual antimicrobials derived from logistic regression models.

Antimicrobial	Pools of 5 samples*		Pools of 10 samples*		Pen floor†	
	Odds ratio	*P*	Odds ratio	P	Odds ratio	*P*
Ampicillin	0.83	0.791	NE	NE	1.12	0.822
Cephalothin	0.60	0.504	NE	NE	1.52	0.513
Chloramphenicol	0.75	0.662	1.01	0.994	0.46	0.103
Kanamycin	0.54	0.327	0.38	0.363	0.55	0.299
Nalidixic acid	0.86	0.826	1.72	0.608	1.78	0.382
Streptomycin	1.03	0.915	1.42	0.366	1.15	0.497
Sulfamethoxazole	0.75	0.175	1.32	0.390	0.89	0.459
Tetracycline	0.76	0.176	1.14	0.612	0.95	0.726

*Reference level was individual samples. †Reference level was samples collected from the rectum of steers.
NE = Not estimable because all isolates were susceptible.

maximum multiple resistant count for pooled feces from 10 steers was 3 (Table 3.4). Distribution of multiple resistant counts was not significantly ($P = 0.486$) different between individual samples obtained from the rectum of steers and from pen floors.

Within samples collected per rectum, there was not a significant ($P = 0.819$) difference between the distribution of multiple resistant counts for individual samples and for pooled samples from 5 and 10 steers. Similarly for pen floor samples, there was not a significant ($P = 0.470$) difference between the distribution of multiple resistance counts for individual samples, pooled feces from 5 samples, and pooled feces from 10 samples. Forty resistance phenotypes were identified in the antibiograms from all isolates (Table 3.5). More phenotypes were identified among isolates from the individual samples obtained from steers and pen floors (28 and 25, respectively) than among the pooled feces of 5 or 10 samples (maximum of 15 phenotypes from pooled samples obtained from 5 steers). It should be mentioned that the number of isolates tested for the pooled feces of 5 and 10 samples were approximately one-fifth and one-tenth, respectively, the number of isolates tested for the individual samples. The phenotypes with higher prevalence (> 2% of isolates) that were found in individual samples collected from steers and pen floors were also detected in the antibiograms for the pooled samples. When the prevalence of a resistance phenotype in the antibiogram was ≤ 1.1% in individual samples from steers and the pen floor, then we inconsistently detected that phenotype in those pooled samples.

Number of phenotypes among the 5 isolates from each sample was tabulated separately

Table 3.4. Percentage of isolates from individual and pooled fecal samples obtained from the rectum of steers or from pen floors with the specified multiple resistance count (number of antimicrobials to which an isolate was resistant)

Multiple Resistance Count	Rectum			Pen floor		
	Individual	Pool of 5 samples	Pool of 10 samples	Individual	Pool of 5 samples	Pool of 10 samples
0	66.4 (2.5)	60.3 (5.9)	70.8 (6.1)	65.7 (2.7)	55.3 (6.1)	70.0 (7.7)
1	10.1 (1.5)	15.1 (3.2)	13.5 (5.0)	11.9 (1.7)	13.4 (3.7)	8.9 (3.9)
2	10.6 (1.5)	14.0 (3.8)	6.7 (3.9)	12.2 (1.8)	18.4 (4.8)	11.1 (5.2)
3	11.2 (1.7)	8.9 (2.8)	9.0 (2.8)	7.7 (1.4)	9.5 (3.4)	4.4 (2.0)
4	0.7 (0.4)	1.7 (1.2)	0.0 (0.0)	1.3 (0.6)	1.1 (0.8)	5.6 (3.8)
5	0.5 (0.3)	0.0 (0.0)	0.0 (0.0)	0.9 (0.6)	0.0 (0.0)	0.0 (0.0)
6	0.1 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.6 (0.6)	0.0 (0.0)
7	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.1 (0.1)	1.1 (0.1)	0.0 (0.0)
8	0.5 (0.4)	0.0 (0.0)	0.0 (0.0)	0.2 (0.2)	0.6 (0.6)	0.0 (0.0)

Numbers in parentheses represent SE.

Table 3.5. Percentage of isolates from individual and pooled fecal samples obtained from the rectum of steers or from the pen floors with the specified resistance phenotypes

Antimicrobial resistant Phenotypes	Rectum			Pen floor		
	Individual (n = 885)	Pool of 5 samples (n = 179)	Pool of 10 Samples (n = 89)	Individual (n = 894)	Pool of 5 samples (n = 179)	Pool of 10 samples (n = 90)
St, S, T	9.2	6.2	9.0	6.6	8.9	2.2
S, T	8.9	12.9	6.7	11.0	16.8	8.9
T	8.6	7.8	12.4	8.4	11.7	7.8
St, T	1.1	0.6	0	0.6	0	0
N, S, T	0.8	1.1	0	0.4	0.6	1.1
S	0.7	3.4	1.1	2.2	1.7	1.1
Ac, A, Cef, Cep, C, St, S, T	0.5	0	0	0.2	0.6	0
A, C, St, S, T	0.3	0	0	0.8	0	0
Ac, A, Cep	0.3	1.1	0	0	0	0
K, St, S, T	0.3	0.6	0	0.6	0	2.2
A, Cep, T	0.2	0	0	0	0	0
N	0.2	0.6	0	0.1	0	0
St	0.2	0.6	0	0.2	0	0
Ac, A, Cep, St, S, T	0.1	0	0	0	0	0
A, C, St, S	0.1	0	0	0.1	0	0
A, St, S	0.1	0	0	0	0	0
A, S, T	0.1	0	0	0	0	0
A, S	0.1	0	0	0	0	0
Cef, S, T	0.1	0	0	0	0	0
Cef	0.1	0	0	0	0	0
C, St, S, T, Tri	0.1	0	0	0.1	0	0
C, St, S, T	0.1	1.1	0	0.4	0.6	3.3
K, S, T	0.1	0.6	0	0.2	0	1.1
K, S	0.1	0	0	0.1	0	0
K	0.1	0.6	0	0.1	0	0
N, St, S, T	0.1	0	0	0.1	0	0
N, St, T	0.1	0	0	0	0	0
St, S,	0.1	0	0	0	0	0
Ac, A, Cep, T	0	0	0	0.1	0	0
Ac, Cep	0	0	0	0.1	0	0

Table 3.5 (cont.)

Antimicrobial resistant phenotypes	Rectum			Pen Floor		
	Individual (n = 885)	Pool of 5 samples (n = 179)	Pool of 10 Samples (n = 89)	Individual (n = 894)	Pool of 5 samples (n = 179)	Pool of 10 samples (n = 90)
A, Cef, C, St, S, T	0	0	0	0	0.6	0
A, St, S, T	0	0	0	0	0.6	0
C, S, T	0	0	0	0.1	0	0
C, S	0	0	0	0.1	0	0
C	0	0.6	0	0.2	0	0
Cip, T	0	0	0	0	0.6	0
K, St, T	0	0	0	0.2	0	0
K, T	0	0.6	0	0	0.6	0
Ac, A, Cep, C, St, S, T	0	0	0	0.1	0	0

*Isolates with missing susceptible-intermediate-resistant data for any antimicrobial are not included.
A = Ampicillin. Ac = Amoxicillin-clavulanic acid. C = Chloramphenicol. Cef = Ceftriaxone. Cep = Cephalothin. K = Kanamycin. Cip = Ciprofloxacin. N = Nalidixic acid. S = Sulfamethoxazole. St = Streptomycin. T = Tetracycline. Tri = Trimethoprim-sulfamethoxazole.

for the individual samples (steer and pen floor samples) and for the pooled feces from 5 samples (pooled steer and pen floor samples; Table 3.6). A majority of samples (75.6% of individual and 62.5% of pooled samples), had ≤ 2 resistance phenotypes. A single phenotype (e.g., all 5 isolates from a sample had a common phenotype) was found in 35.4% of individual samples. Most often, the single phenotype was susceptible to all antimicrobials (no resistance detected), but some samples had all isolates resistant to tetracycline alone (susceptible to all other antimicrobials tested). A single phenotype was detected in 25% of the pooled samples. As with individual samples, the single phenotype in pooled samples was predominantly susceptible to all antimicrobials, but some of the samples had a single phenotype that was resistant to sulfamethoxazole and tetracycline. Detection of 4 or 5 resistance phenotypes in a sample was relatively rare (5.2% of all samples).

Under the assumption of a random distribution of isolates, distribution of the resampling phenotype resulted in a bell-shaped distribution with the largest number of samples having 3 phenotypes. Results of the goodness-of-fit tests that were used to compare actual individual samples with simulated individual samples ($\chi^2 = 12.4$; $P = 0.014$) and actual pools of 5 samples with simulated pools of 5 samples (($\chi^2 = 19.54$; $P < 0.001$) indicated that the phenotypes detected among isolates from within samples were not behaving as though they were randomly distributed. Based on the apparent differences in the χ^2 values for the table and the cells, there was an overrepresentation of samples with a single phenotype.

Table 3.6. Percentage of actual and simulated random samples (5 isolates/sample) with the number of unique resistance phenotypes for individual samples and pooled feces of 5 samples

Number of unique phenotypes	Individual samples (n = 360)		Pool of 5 samples (n = 64)	
	Actual	Simulated random	Actual	Simulated random
1	35.44	8.27	25.00	4.48
2	40.19	32.56	37.50	26.48
3	19.62	38.66	29.69	41.13
4	4.11	17.50	6.25	23.30
5	0.63	3.00	1.56	4.61

Descriptive statistics for MIC values depicted a consistent pattern among the various sample collection strategies (Table 3.7). Chloramphenicol, however, had consistently higher geometric mean and MIC_{50} values in samples collected from the pen floor, compared with values for samples collected from each individual steer. Distributions of MIC values for all 17 antimicrobials were not significantly different between individual and pooled samples obtained from 5 steers ($P \geq 0.4$) and between individual and pooled samples obtained from 10 steers ($P \geq 0.6$). Distributions between individual and pooled feces of 5 pen floor samples were not significantly ($P \geq 0.3$) different, except for tetracycline ($P = 0.03$). The biggest difference in the cumulative distributions was at an MIC of 8.0 for tetracycline (difference of 11.8%) and 4.0 for cephalothin (difference of 10.3%). None of the 17 antimicrobial MIC distributions were significantly ($P \geq 0.8$) different between individual samples obtained from the pen floors and pooled feces from 10 pen floor samples. For 15 of 17 antimicrobials, MIC distributions for individual samples obtained from the steers and pen floor were not significantly ($P \geq 0.36$) different. The 2 antimicrobials in which the pattern differed significantly were apramycin ($P = 0.02$) and chloramphenicol ($P = 0.002$). The biggest difference in the cumulative distributions was at an MIC of 2 for apramycin (difference of 7.2%) and 4 for chloramphenicol (difference of 8.7%).

Discussion

A single best method for comparing antimicrobial resistance patterns among groups of animals does not exist at this time. As such, the comparison of antimicrobial resistant

Table 3.7 Descriptive statistics of minimum inhibitory concentration (MIC) values for individual and pooled fecal samples obtained from the rectum of steers or from the pen floors (MIC_{50}=the 50th percentile MIC value).

Antimicrobial	Individual samples				Pool of 5 samples				Pool of 10 samples			
	Rectal		Pen floor		Rectal		Pen floor		Rectal		Pen floor	
	Mean	MIC_{50}	Mean	MIC_{50}	Mean	MIC_{50}	Mean	MIC_{50}	Mean	MIC_{50}	Mean	MIC_{50}
Amikacin	4.01	4	4.01	4	4.00	4	4.05	4	4.0	4	4.0	4
Amoxicillin-clavulanic acid	2.73	4	2.72	4	2.58	2	2.98	4	2.63	2	2.62	2
Ampicillin	2.65	2	2.56	2	2.60	2	2.69	2	2.73	2	2.42	2
Apramycin	3.03	4	3.19	4	2.93	4	3.28	4	2.93	4	3.13	4
Cefoxitin	4.15	4	4.11	4	4.08	4	4.27	4	4.16	4	4.06	4
Ceftiofur	0.51	0.5	0.51	0.5	0.50	0.5	0.53	0.5	0.50	0.5	0.50	0.5
Ceftriaxone	0.26	0.25	0.25	0.25	0.25	0.25	0.27	0.25	0.25	0.25	0.25	0.25
Cephalothin	5.17	4	5.08	4	5.08	8	5.80	4	5.17	4	4.96	4
Chloramphenicol	5.73	4	6.17	8	5.50	4	6.34	8	5.51	4	6.50	8
Ciprofloxacin	0.02	0.015	0.02	0.015	0.02	0.015	0.02	0.015	0.02	0.015	0.02	0.015
Gentamicin	0.37	0.50	0.38	0.50	0.36	0.50	0.40	0.50	0.36	0.25	0.38	0.50
Kanamycin	16.16	16	16.35	16	16.63	16	16.31	16	16.00	16	17.28	16
Nalidixic acid	4.18	4	4.12	4	4.21	4	4.10	4	4.06	4	4.16	4
Streptomycin	35.60	32	35.23	32	34.71	32	35.94	32	34.06	32	35.37	32
Sulfamethoxazole	175.28	128	179.96	128	186.81	128	199.03	128	162.95	128	171.31	128
Tetracycline	7.62	4	7.43	4	7.81	4	9.24	4	7.29	4	7.13	4
Trimethoprim	0.14	0.12	0.14	0.12	0.14	0.12	0.14	0.12	0.13	0.12	0.13	0.12

patterns among groups of interest should be performed by use of a variety of methods. In the study reported here, we compared the prevalence of resistance to individual the study reported here, we compared the prevalence of resistance to individual antimicrobials, multiple resistance counts, antibiograms, and descriptive and inferential analysis of MIC values. Despite the diversity of analytical approaches, there is an inherent subjective component to the comparison of patterns that should be addressed prior to any such study. Characterization of relatively prevalent resistance patterns likely will necessitate a different approach to sampling approach than would be needed if the objective was to detect emerging resistance. In this study, phenotypes with a lower prevalence were more likely to be detected when the sampling intensity (i.e., number of isolates) was greatest, as would be expected. Regardless of the sample collection protocol, detection and estimation of resistance patterns that had a higher prevalence was more consistent.

In other studies that examined on-farm antimicrobial resistance in production animal populations, investigators have variably relied on individual samples from animals, including swab specimens obtained from the rectum[25] or fecal samples collected from the pen floor.[12;16;26] In the study reported here, we compared antimicrobial resistance patterns derived for fecal samples collected per rectum from individual cattle with those for samples obtained from the pen floors of the feedlot. Overall, patterns of antimicrobial resistance were extremely similar. Small differences, such as detection or lack of a rare phenotype in the antibiogram, were more likely a result of chance selection of the isolate than an inherent difference in the patterns. The only significant difference in

susceptibility patterns between the 2 methods was in the MIC distribution for tetracycline.

Many advantages exist for collecting fecal samples from the pen floor rather than the rectum of cattle in feedlots. Cattle must be restrained in a chute to allow collection of fecal samples from the rectum. This activity creates a disruption for the cattle, and it may also result in injuries to the cattle. It also is a costly procedure for the producer, because it can affect feed intake and rates of gain of the cattle. In contrast, samples can be easily obtained from the pen floors without substantial disruption to the cattle. Collection of samples from the pen floors generally does not affect the amount of work or expense for producers. Analysis of results of this study indicated that the use of samples collected from the pen floors to determine resistance patterns was essentially equivalent to the use of individual samples collected from the rectum of each steer.

Pooling of fecal samples can reduce costs associated with assessing antimicrobial resistance patterns. Primarily, the savings would be in culture costs if the number of isolates being examined is held constant. In this study, we compared individual samples and pooled samples consisting of 5 and 10 fecal samples. Five isolates were obtained from each sample. Consequently, 150, 30, and 15 isolates were used to describe each pen by use of individual samples, pooled feces from 5 samples, and pooled feces from 10 samples, respectively. Potentially, the cost savings for pooling of samples would be much greater, because the total number of isolates tested was greatly reduced for the pools. However, there were apparent differences between the distribution of antimicrobial

resistance patterns identified from individual fecal samples and pooled fecal samples. The most notable difference was the lack of isolates with antimicrobial resistance phenotypes that had a lower prevalence in the pooled feces from 10 samples. This lack of detection was reflected in the antibiograms, univariate proportions, and multiple resistant count. However, logistic regression models for antimicrobials with higher prevalences did not reveal significant differences between individual and pooled samples. Descriptive statistics for MIC values also tended to be comparable between individual samples and pooled samples.

Observed differences in susceptibility patterns obtained from individual samples and pooled samples may largely be attributable to differences in the number of isolates tested in each protocol. For example, 885 and 894 isolates were tested for susceptibility in the individual samples collected from steers and pen floors, respectively. Based on these sample sizes and assuming a perfect test (100% sensitivity and 100% specificity) and a sample of independent observations, we would have expected to detect a prevalence of approximately 0.3% with 95% confidence.[27] In comparison, the sample sizes of approximately 180 and 90 for the pooled feces of 5 and 10 samples would have been expected to detect antimicrobial resistance of approximately 1.6 and 3.3%, respectively, with 95% confidence.

For pooling to be an effective alternative to the use of individual samples for detection of antimicrobial resistance with a low prevalence, the number of total isolates obtained from the pools must be increased above those used in the study reported here. The number of

isolates can be increased by testing more isolates from each pool or by increasing the number of pools that are contributing isolates. Increasing the number of isolates for each pooled fecal sample is problematic because of the apparent clustering of antimicrobial resistance phenotypes that was observed in this study in individual samples and pooled feces from 5 samples. Essentially, when clustering is a problem, each successive isolate tested provides less information than the previously tested isolate. Therefore, effective sample size is reduced below actual sample size. The second alternative is to increase the number of pools that are formed. The limiting factor with increasing the number of pools will likely be the number of fresh fecal pats that are available for sample collection at a specific time point.

Dunlop et al[12] used a bootstrap resampling approach to document that maximizing the number of pigs contributing to composite (pooled) samples was the most cost-efficient sample collection plan for measuring concentrations of *E. coli* resistance. This result was at least in part attributable to the importance of between-pig variation. Also, they found that precision of their resistance estimate was more sensitive to the number of fecal samples collected than to the number of colonies selected per specimen. The latter 2 results are consistent with the findings for our study and measuring antimicrobial resistance in feedlot cattle. Results of the multiple-level model revealed that between-sample (between-animal) variation in resistance prevalence was important for all antimicrobials but even more so for the antimicrobials that had low prevalence of resistance. Higher between-sample variation in resistance prevalence relative to within-animal variation would seem to indicate that sample collection efforts should not rely on

simply increasing the number of isolates tested per sample but that more fecal samples are necessary.

Pen-to-pen variation in resistance was relatively small in our study. This minimal variation may be attributable in part to the homogeneous nature of the cattle that were in the 6 pens with regard to origin, weight, antimicrobial use in feed, and time at the feedlot. This result indicates that, for this particular feedlot, the number of pens that needed to be sampled was not as critical as the number of samples collected. Extrapolating the pen-level results to a more typical cross section of pens on a commercial feedlot, whereby more pen-to-pen variation might be expected, should be done cautiously. We believe it is important to look at sources of variation in resistance in a larger scale project to evaluate pen-to-pen and, potentially, feedlot-to-feedlot variation to see whether similar results can be obtained.

In contrast to the study reported here, Dunlop et al.[12] found that composite sampling was the most efficient sample collection method. This contrast may be the result of differences in the outcomes measured in the 2 studies. In that previous study, investigators measured concentrations of *E. coli* resistance to 2 antimicrobials (tetracycline and gentamicin), whereas we focused on estimating SIR and MIC characteristics for 17 antimicrobials without specifically enumerating abundance of organisms with each resistance phenotype. Secondly, the evaluation of composite sampling in that previous study was based on Monte-Carlo bootstrap resampling of bacterial abundance data, whereas our composites were actually constructed and tested

for resistance. Their simulation allowed them to create equal numbers of composite samples and individual samples on which to base their comparisons of methods. In our application, it does not appear to be feasible to create an equal number of pooled and individual samples. Additionally, in our study, the pooling of fecal samples following a designed randomization scheme was a resource-intensive effort in the laboratory.

The level at which clustering was identified in our study has implications for the overall sample collection design, total sample size, and number of isolates that should be used from a specific sample. Clustering was evident in individual samples and, to a slightly lesser degree, pooled feces of 5 samples. Clustering of resistance in our protocol could have arisen from 2 sources. The first source was clustering of phenotypes within the original sample (within-animal effect). Factors affecting the ecology of *E. coli* resistance within each steer, such as exposure to antimicrobials, management factors, or health-related circumstances, could have led to clustering. The second potential source of clustering was in the laboratory (bacterial growth and colony selection) phase of the study. Cannon and Nicholls[28] discussed the impact of clumping of bacteria (clones) on the detection of *Salmonella enterica* in fecal samples. They documented that, theoretically, detection of bacteria will increase when the sample is homogenized before an aliquot is obtained for testing. Their results pertained to detecting rare bacterial species, which is analogous to testing for rare resistance phenotypes. In the study reported here, vortexing of the fecal material, agitation during shipping, and further mixing in the laboratory represented a substantial effort to homogenize samples. Regardless of the effort to homogenize samples, clustering still occurred. It is possible

that some of the observed clustering could have arisen during other phases of sample handling and testing. Additionally, if all clustering of resistance phenotypes were attributable to within-animal clustering and none to between-animal clustering, then we would have expected reduced overall clustering with the pooled samples. Clustering of resistance phenotypes was only slightly reduced in the pooled samples, compared with the individual samples, which means that there likely was between-animal as well as within-animal clustering. With these data, it is not possible to separate whether clustering was the result of animal or laboratory-related factors.

The need for a large sample of isolates to identify rarely occurring resistance patterns must be balanced with the problem of clustering within samples. Within-sample clustering can be avoided by use of a single isolate from each sample. This approach would necessitate a substantial increase in the number of fecal samples that are collected to meet the sample size of isolates deemed necessary for detection or estimation of phenotype prevalences. A compromise is to select > 1 isolate/sample to increase sample size while considering the reduced benefit of the use of a large number of isolates per sample. Vosti et al.[13] raised similar issues when they identified serotypes of *E. coli* from human fecal samples. Approximately 50% of the specimens they examined had only 1 serologic group, whereas 10% contained 4 or 5 groups. Most of the serologic groups (almost 80%) were detected in the first 5 of the 15 to 25 isolates used for each sample. Dunlop et al.[12] examined the effect of various sample sizes on the precision of estimates of the proportion of resistance to 2 antimicrobials. They found that precision of the estimates did not noticeably improve when > 5 colonies were chosen for each fecal

sample. Consequently, when trying to obtain a desired sample size to estimate resistance proportions, it may be prudent to limit the number of isolates to ≤5 for each fecal sample.

Footnotes

[a]Fluorocult MacConkey Agar, EM Science, Gibbstown, NJ

[b]Sensititre, Trek diagnostics, Westlake, OH.

[c]SAS, version 8.1, SAS Institute Inc, Cary, NC.

[d]Stata, version 7.0, Stata Corp, College Station, Tex.

[e]MlwiN, version 1.1, Institute of Education, University of London, London, UK.

References

1. Baumgartner A, Kueffer M, Rohner P. Occurrence and antibiotic resistance of enterococci in various ready-to-eat foods. *Arch. Lebensm hyg* 2001;52:16–19.

2. Brophy PO, Caffrey PJ, Collins JD. Sensitivity patterns of *Escherichia coli* isolated from calves during and following prophylactic chlortetracycline therapy. *Br Vet J* 1977;133:340–345.

3. Dargatz DA, Fedorka-Cray PJ, Ladely SR, et al. Survey of *Salmonella* serotypes shed in feces of beef cows and their antimicrobial susceptibility patterns. *J Food Prot* 2000;63:1648–1653.

4. Seyfarth AM, Wegener HC, Frimodt-Moller N. Antimicrobial resistance in *Salmonella enterica* subsp. *enterica* serovar typhimurium from humans and production animals. *J Antimicrob Chemother* 1997;40:67–75.

5. White DG, Zhao S, Sudler R, et al. The isolation of antibiotic-resistant salmonella from retail ground meats. *N Engl J Med* 2001;345:1147–1154.

6. Wiggins BA, Andrews RW, Conway RA, et al. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *Appl Environ Microbiol* 1999;65:3483–3486.

7. Bywater RJ. Sense and nonsense in surveillance programs. *Acta Vet Scand – Suppl.* 2000;93:119–127.

8. Gould IM. Stewardship of antibiotic use and resistance surveillance: the international scene. *J Hosp Infect* 1999;43:S253–S260

9. Jones RN. Detection of emerging resistance patterns within longitudinal surveillance systems: data sensitivity and microbial susceptibility. MYSTIC Advisory Board. Meropenem Yearly Susceptibility Test Information Collection. *J Antimicrob Chemother* 2000;46(Suppl T2):1–8.

10. Marano NN, Rossiter S, Stamey K, et al. The National Antimicrobial Resistance Monitoring System (NARMS) for enteric bacteria, 1996–1999: surveillance for action. *J Am Vet Med Assoc* 2000;217:1829–1830.

11. Sahm DF, Karlowsky JA, Kelly LJ, et al. Need for annual surveillance of antimicrobial resistance in *Streptococcus pneumoniae* in the United States: 2-year longitudinal analysis. *Antimicrob Agents Chemother* 2001;45:1037–1042

12. Dunlop RH, McEwen SA, Meek AH, et al. Sampling considerations for herd-level measurement of faecal *Escherichia coli* antimicrobial resistance in finisher pigs. *Epidemiol Infect* 1999;122:485–496.

13. Vosti KL, Monto AS, Rantz LA. The importance of sample size in studies based upon the serologic classification of *Escherichia coli*. *Proc Exper Biol Med* 1962;iii:201–204.

14. Hoar BR, Atwill ER, Elmi C, et al. Comparison of fecal samples collected per rectum and off the ground for estimation of environmental contamination attributable to beef cattle. *Am J Vet Res* 1999;60:1352–1356.

15. Hinton M, Rixson PD, Allen V, et al. The persistence of drug resistant *Escherichia coli* in strains in the majority faecal flora of calves. *J Hyg Camb* 1984;93:547–557.

16. Mercer HD, Pocurull D, Gaines S, et al. Characteristics of antimicrobial resistance of *Escherichia coli* from animals: relationship to veterinary and management uses of antimicrobial agents. *Appl Microbiol* 1971;22:700–705.

17. Bourque R., Lallier R, Lariviere S. Influence of oral antibiotics on resistance and enterotoxigenicity of *Escherichia coli*. *Can J Comp Med* 1980;44:101–108.

18. Chang GW, Brill J, Lum R. Proportion of beta-D-glucuronidase-negative *Escherichia coli* in human fecal samples. *Appl Environ Microbiol* 1989;55:335–339.

19. Wray C. Some aspects of the occurrence of resistant bacteria in the normal animal flora. *J Antimicrob Chemother* 1986;18:141–147.

20. National Antimicrobial Resistance Monitoring Program - Enteric Bacteria. *Veterinary isolates final report*. Athens, GA: USDA:Agricultural Research Service, 1997;11.

21. National Committee for Clinical Laboratory Standards (NCCLS). Performance standards for antimicrobial susceptibility testing: Ninth informational supplement. NCCLS document M100-S12. Wayne, PA: National Committee for Clinical Laboratory Standards,2002; 90–93

22. StataCorp. *Stata reference manual, release 7.0, volume 4 SU-Z*. College Station, Tex: Stata Corporation, 2001.

23. Dohoo, IR, Tillard E, Stryhn H, and Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island. *Prev Vet Med* 2001;50:127–144

24. Conover WJ. Chapter 6. *Practical nonparametric statistics*. 2nd ed. New York: John Wiley and Sons, 1980;346–349

25. Langlois BE, Dawson KA, Leak I, et al. Effect of age and housing location on antibiotic resistance of fecal coliforms from pigs in a non antibiotic exposed herd. *Appl Environ Microbiol* 1988;54(6):1341–1344.

26. Dargatz DA, Fedorka-Cray PJ, Ladely SR, et al. Antimicrobial susceptibility patterns of *Salmonella* isolates from cattle in feedlots. *J Am Vet Med Assoc* 2002;221(2):268–272

27. Cannon RM. Sense and sensitivity - designing surveys based on an imperfect test. *Prev Vet Med* 2001;49:141–163.

28. Cannon RM, Nicholls, TJ. Relationship between sample weight, homogeneity, and sensitivity of fecal culture of *Salmonella enterica*. *J Vet Diagn Invest* 2002;14:60–62.

CHAPTER 4

SHORT-TERM REPEATABILITY OF MEASUREMENTS OF ANTIMICROBIAL SUSCEPTIBILITY OF *ESCHERICHIA COLI* ISOLATED FROM FECES OF FEEDLOT CATTLE

Summary

Short-term stability of measurements of antimicrobial susceptibility of *Escherichia coli* organisms isolated from feces of feedlot cattle is important in developing monitoring and surveillance programs. Frequent evaluations (i.e. daily) are resource intensive and in some situations may be impractical for long term sampling protocols. Consequently, a point-in-time measurement will need to be used to represent conditions in the peri-sampling period. In this study, 30 fecal samples were collected from each of six cattle pens on a commercial cattle feedlot on two occasions separated by 48 hours. *Escherichia coli* was isolated from single and pooled samples. The isolates were tested for antimicrobial susceptibility against a panel of 17 antimicrobials. Resistance to five antimicrobials (ampicillin, nalidixic acid, streptomycin, sulfamethoxazole, and tetracycline) was detected in single and pooled samples from both sampling periods (Day 1 and Day 3). The prevalence of isolates resistant to these five antimicrobials was two percent or higher in all treatment combinations except for pools obtained from Day 3 samples. Lower levels of resistance to six more antimicrobials were detected inconsistently across the single and pooled samples. Logistic models constructed for the antimicrobials for which the *E. coli* isolates were most commonly resistant demonstrated

that there were no significant differences between periods (P>0.10) and between single and pooled samples (P>0.20). The distribution of the number of antimicrobials to which isolates were resistant was consistent for the single samples across periods but there appeared to be a lower prevalence of any resistance in Day 1 pooled samples. A larger number of resistance phenotypes were detected in the single samples than in the pooled samples and resistance phenotypes with prevalence of less than two percent were detected inconsistently across periods and single/pooled samples. Resistance to individual antimicrobials was consistent by all measures when the prevalence was at least two percent. Inconsistent results were obtained for antimicrobials to which resistance rarely occurred. The apparent inconsistencies do not appear to be related to external factors but rather to sampling intensity. Short-term stability is a plausible assumption under sampling strategies that are designed to detect specific levels of prevalence. However, when resistance levels fall below these levels there will likely be fluctuations in the presence/absence of rare resistance phenotypes and in their prevalence and central tendency measures.

Introduction

Monitoring systems for antimicrobial resistance associated with food animals are being developed and implemented throughout the world in response to public and animal health concerns. In the U.S., the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS-EB) is a collaborative effort between the Centers for Disease Control and Prevention (CDC), the Food and Drug Administration (FDA) and the United States Department of Agriculture (USDA) formed to monitor antimicrobial resistance in

bacteria isolated from humans, animals, and animal products.[4] One of the specific goals of NARMS-EB is to provide descriptive data regarding the extent and temporal trends in antimicrobial resistance in enteric bacteria from human and animal populations. The NARMS-EB isolates selected for resistance testing are from both clinical and nonclinical sources.[8] A second monitoring system in the U.S. is part of the FDA's Center for Veterinary Medicine proposed framework for evaluating the use of antimicrobial drugs in food animals.[3] The framework includes both pre-approval and post-approval monitoring for antimicrobial resistance development. These monitoring efforts, as well as other studies of antimicrobial resistance in food animal populations, require the use of appropriate sampling methods to ensure that the resulting data can be used effectively to make sound decisions.

Monitoring efforts that are not solely reliant on clinical isolates will need to involve active sample collection strategies from the animal species of interest. The breadth of the animal populations that need to be monitored on a national basis in the U.S. and the expense of antimicrobial susceptibility testing will likely make it impractical to implement longitudinal sampling on a frequent basis. Similarly, specific studies which will be designed to focus on longer term changes in the antimicrobial resistance pattern in a group of animals, such as feedlot cattle, likely will depend on weekly, monthly, or quarterly sampling efforts to describe resistance in the bacterial population. The use of a point-in-time description of antimicrobial resistance, such as a specific day in a week, is based on the assumption that the measured outcome is representative of the time period. If the antimicrobial resistance outcomes were representative of the time period, one

would expect that short-term variability in resistance profiles should be minimal barring an intervention, such as diet, weather change or new exposure to antimicrobial selection pressure. The issue of resistance variability extends beyond that which may be occurring biologically to include variability due to sampling, handling, processing, and antimicrobial susceptibility testing. In most situations, it would be difficult to separate the variation in resistance due to animal/environmental factors and factors related to the process of susceptibility testing. If the total variation in measured antimicrobial resistance over short time periods is large, then it may be difficult to interpret findings when differences in resistance patterns are detected over time.

The objective of this study was to determine the short-term repeatability of measurements of antimicrobial resistance in *Escherichia coli* isolated from the feces of feedlot cattle. Specifically, our intention was to compare resistance patterns for *E. coli* organisms isolated from fecal samples collected from pen floors two days apart using both single and pooled fecal samples.

Methods

Study population

Fecal samples were collected on two separate occasions (Day1 and Day3) from each of 6 randomly selected pens of cattle on a commercial feedlot in central Colorado. The two sampling occasions were 48 hours apart. Four of the pens were 41,800 square feet in size and the remaining two pens were 83,600 square feet. The smaller pens contained between 191 and 261 beef steers while the larger pens had 519 and 565 beef steers,

respectively. All cattle had been on the feedlot for at least 129 days and were on a finishing diet which contained tylosin and an ionophore. On each visit, fecal samples were collected from the interior of 30 fresh fecal pats found on the floor of each pen. Gloves were used to obtain approximately 50 grams of fecal material from each fecal pat. Gloves were changed between samples. Efforts were made to collect the fecal samples from sites throughout the pen to minimize the likelihood of collecting multiple fecal samples from the same animal. The samples were cooled on ice and transported to the laboratory.

In the laboratory, approximately 1 gram of fecal material from each sample was placed in a 50 ml conical tube. Additionally, pools of five fecal samples were created in the laboratory. Each single fecal sample was assigned randomly to a pool of five single samples. Each single fecal sample contributed approximately 0.2 grams to only one pool of five samples. A wooden applicator was used to transfer the fecal material to 50 ml conical tubes. Ten ml of 0.9 percent (normal) saline was added to each tube. The tubes were vortexed to mix the saline and the fecal material. The tubes were refrigerated overnight at 3°C. On the next day the diluted samples were shipped for overnight delivery to a second laboratory that conducted the culturing and antimicrobial susceptibility testing.

Culture and antimicrobial susceptibility testing

Culture and testing methodologies were previously described in detail.[10] Briefly, diluted fecal material was inoculated on MacConkey-MUG agar plates[a] using swabs then struck

for isolation of *E. coli*. After incubation, individual *E. coli* colonies were identified as lactose-positive (bright pink) and, under ultraviolet light, glucouronidase positive (colony periphery had a bluish appearance). One colony from each single sample agar plate and three colonies from each pool of five agar plate were selected and transferred to individual nutrient agar slants and incubated for another 18-24 hours. In preparation for antimicrobial susceptibility testing, MacConkey agar plates were inoculated with bacteria from the slants. Colonies were selected from these plates and placed into separate tubes with 5 ml of sterile water and turbidity was adjusted to a 0.5 McFarland standard. After mixing, 10 µl of the bacterial suspension were used to inoculate 10 ml of cation-adjusted Mueller-Hinton broth, which was then used to inoculate 96-well plates containing different quantities of antimicrobial drugs for susceptibility testing. The plates were incubated at 35° C for 18-24 hours and then read using a semi-automated broth dilution system.[b] Based on growth in individual wells, the minimum inhibitory concentrations (MIC) of 17 antimicrobial drugs were determined using an automated plate reading apparatus. The 17 antimicrobials and dilution ranges were selected to parallel the panel used by the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS-EB).[51] The MIC for each antimicrobial was used to classify the isolate as susceptible, intermediate, or resistant according to standards used to classify bacteria isolated from human sources as determined by the National Committee on Clinical Laboratory Standards (NCCLS) and as used by the NARMS-EB surveillance program.

---

[1] Amikacin (4-32 µg/ml), Amoxicillin-Clavulanic acid (0.5/0.25-32/16 µg/ml), Ampicillin (2-64 µg/ml), Apramycin (2-32 µg/ml), Cefoxitin (4-32 µg/ml), Ceftiofur (0.5-16 µg/ml), Ceftriaxone (0.25-64 µg/ml), Cephalothin (4-32 µg/ml), Chloramphenicol (4-32 µg/ml), Ciprofloxacin (0.015-2 µg/ml), Gentamicin (0.25-16 µg/ml), Kanamycin (16-64 µg/ml), Naladixic acid (4-256 µg/ml), Streptomycin (32-256 µg/ml), Sulfamethoxazole (128-512 µg/ml), Tetracycline (4-32 µg/ml), Trimethoprim-Sulfamethoxazole (0.12/2.38-4/76 µg/ml).

Currently, veterinary breakpoints for the antimicrobials used in this study have not been verified.[6] Recorded MIC values were truncated at the limits of the concentrations in the wells.

Statistical analysis

Statistical analyses were performed both on the categorical SIR and the MIC outcomes of the antimicrobial resistance testing. The proportion of isolates resistant to each of the 17 antimicrobials and the associated standard error were determined using Stata software.[c] The survey procedures in Stata allow for calculation of standard errors using the delta method (first order Taylor series approximation) to account for the lack of independence of multiple isolates taken from a sample and for the nesting of samples within pen.[7] Unless specified, all statistical analyses accounted for the study's sampling design. Logistic regression models were constructed to evaluate the effects of pooling (single versus pooled samples) and period (Day1 versus Day3) on resistance to the antimicrobials for which the *E. coli* were most commonly resistant (ampicillin, cephalothin, chloramphenicol, kanamycin, nalidixic acid, streptomycin, sulfamethoxazole, and tetracycline). For other antimicrobials the prevalence of resistance was low resulting in unstable logistic models.

The number of antimicrobials to which each isolate was resistant, the multiple resistance count, was determined. A chi-square test was performed to compare the distribution of the multiple resistance count between the treatments of interest (single sample Day1

period versus single sample Day3, pools of five samples Day1 versus pools of five samples Day3 and single animal samples versus the pools of five samples).

An antibiogram was constructed to describe phenotypic resistance patterns for single and pooled samples from each period. Intermediate test results were grouped with susceptible results for this purpose.

The analysis of the MIC data included calculation of descriptive measures such as the MIC_{50} and the geometric mean. Additionally, a design-based chi-square test was used to evaluate differences in the MIC distributions between the two periods and the differences between single and pooled samples. The MIC values were treated as categorical for this analysis because of the relatively few MIC values that were observed for each antimicrobial drug.

For the chi-square test and others used in this study, the alpha value was set at 0.05. All analyses not performed in Stata were performed using version 8.1 of SAS.[d]

Results

A total of 570 isolates from single (n=357) and pooled (n=213) samples was evaluated to characterize antimicrobial susceptibility. The total number of isolates tested reflects the loss to analysis of 6 isolates during processing. Resistance to five antimicrobials (ampicillin, nalidixic acid, streptomycin, sulfamethoxazole, and tetracycline) was present in isolates of *Escherichia coli* (not serotyped) in both time periods and in the single and

pooled samples (Table 4.1). No resistance was found to six antimicrobials (amikacin, amoxicillin/clavulanic acid, apramycin, cefoxitin, ceftiofur, and gentamicin) in any isolates from single or pooled samples. Resistance to trimethoprim/sulfamethoxazole was detected in isolates from single samples from both periods (prevalence equal 1.7 and 0.6 percent, respectively) but only in isolates from the Day1 pooled samples (prevalence=0.9 percent). Similarly, cephalothin resistance was detected in isolates from single samples from both periods (prevalence=0.6 percent in both periods) but was only detected in isolates from Day3 pooled samples (prevalence=1.9 percent). Kanamycin resistance was only detected in isolates from the Day3 samples (prevalence $\geq$0.6 percent). Ceftriaxone resistance was detected in the Day1 single sample isolates (prevalence= 0.6 percent) and the Day3 pooled sample isolates (prevalence= 1.0 percent). Resistance to the last two antimicrobials, chloramphenicol and ciprofloxacin, only occurred in the Day3 single sample isolates (prevalence= 1.7 percent) and the Day1 pooled sample isolates (prevalence= 0.9 percent), respectively.

The base levels (reference levels) used for the five logistic regression models were single samples and Day1 samples. Odds ratios represent the likelihood of an isolate being resistant compared to these base levels. Based on the fitted logistic regression models, the odds of an isolate being resistant on Day3 were similar to that of an isolate obtained on Day1 for these antimicrobials (P>0.10, Table 4.2). Similarly, the odds of isolates from a pooled sample being resistant to any of the five antimicrobials did not differ from the isolates from single samples (P>0.20).

Table 4.1 Percent[1] of *Escherichia coli* isolates from single and pooled samples taken during two samplings periods (48 hours apart) with resistance to each of 17 antimicrobials.

	Day 1				Day 3			
Antimicrobial	Single		Pools of 5		Single		Pools of 5	
Amikacin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Amoxicillin-clavulanic acid	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Ampicillin	3.9	(1.5)	2.8	(1.5)	3.4	(1.3)	2.9	(1.6)
Apramycin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Cefoxitin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Ceftiofur	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Ceftriaxone	0.6	(0.6)	0.0	(0.0)	0.0	(0.0)	1.0	(0.9)
Cephalothin	0.6	(0.6)	0.0	(0.0)	0.6	(0.6)	1.9	(1.9)
Chloramphenicol	0.0	(0.0)	0.0	(0.0)	1.7	(1.0)	0.0	(0.0)
Ciprofloxacin	0.0	(0.0)	0.9	(0.9)	0.0	(0.0)	0.0	(0.0)
Gentamicin	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Kanamycin	0.0	(0.0)	0.0	(0.0)	0.6	(0.6)	1.0	(0.9)
Nalidixic acid	2.8	(1.2)	3.7	(2.2)	2.8	(1.2)	1.0	(0.9)
Streptomycin	7.9	(2.0)	6.5	(2.2)	9.5	(2.2)	8.6	(3.6)
Sulfamethoxazole	16.3	(2.8)	6.5	(2.9)	13.4	(2.5)	24.8	(5.0)
Tetracycline	40.4	(3.7)	27.8	(5.7)	39.1	(3.7)	40.0	(5.6)
Trimethoprim	1.7	(0.6)	0.9	(0.9)	0.6	(0.6)	0.0	(0.0)

[1]standard errors are in parentheses

Table 4.2. Odds ratios derived from logistic regression models on five antimicrobials. The reference levels for the models were Day 1 samples and single samples (non-pooled).

	Model	Day 3*		Pools of Five**	
Antimicrobial	P-value	Odds ratio	P-value	Odds ratio	P-value
Ampicillin	0.855	1.12	0.811	1.29	0.604
Nalidixic acid	0.628	1.51	0.459	1.21	0.757
Streptomycin	0.628	0.79	0.452	1.17	0.667
Sulfamethoxazole	0.294	0.67	0.123	0.95	0.861
Tetracycline	0.374	0.86	0.430	1.28	0.235

*Reference level is Day1

**Reference level is single samples

The distributions of multiple resistance counts from the single-sample isolates were similar between the two periods (P=0.94, Table 4.3). In both periods, penta-resistance occurred rarely in the single samples and was not detected in any of the pooled samples. Single isolates of the resistance phenotypes Amp/Str/Sul/Tet/TMS and Chl/Nal/Str/Sul/Tet were found in the Day1 and Day3 single samples, respectively.[2] The distribution of the multiple resistance counts in the single samples compared to the pooled samples was not significantly different in either the Day1 (P=0.160) or the Day3 (P=0.721). The percent of isolates with no resistance was substantially higher in the Day1 pooled samples. Concomitantly, fewer of the isolates in the Day1 pooled samples had resistance to only one antimicrobial.

In combination, the antibiograms for both periods and the single/pooled samples resulted in the identification of 28 unique resistance phenotypes (Table 4.4). Resistance to tetracycline only was the most common resistance phenotype in both periods and in both single and pooled samples. Two other resistance phenotypes, Sul/Tet and Sul/Tet/Str, were found in isolates from both periods and from single and pooled samples.

The number of resistance phenotypes found in isolates from the Day1 and Day3 period single samples, 16 and 19 respectively, while being to similar to each other were high compared to the number identified from pooled samples from both periods, 11 and 10, respectively. Of the 28 resistance phenotypes, nine were found only in the single samples and two were found only in the pooled samples. Single samples from the two periods,

[2] Amp=ampicillin, Chl=chloramphenicol, Nal=nalidixic acid, Str=streptomycin, Sul=sulfamethoxazole, Tet=tetracycline, TMS=trimethoprim-sulfamethoxazole

Table 4.3. Percent* of *Escherichia coli* isolates from single and pooled fecal samples from both periods with the specified multiple resistance count (number of antimicrobials to which an isolate was resistant).

Number of resistant antimicrobials	Day 1		Day 3	
	Single	Pool of 5	Single	Pool of 5
0	54.5(3.7)	71.3 (5.6)	54.2 (3.8)	47.6 (4.9)
1	27.0 (3.3)	15.7 (4.4)	30.2 (3.4)	35.2 (4.3)
2	11.8 (2.4)	7.4 (2.6)	9.5 (2.2)	8.6 (3.1)
3	3.9 (1.5)	3.7 (1.7)	2.8 (1.2)	5.7 (2.5)
4	2.2 (1.1)	1.9 (1.3)	2.8 (1.2)	2.9 (2.1)
5	0.6 (0.6)	0 (0.0)	0.6 (0.6)	0 (0.0)

*number in parentheses are standard errors

Table 4.4. Frequency distribution* of resistance phenotypes (antibiograms) for single and pooled samples for each period (numbers in parentheses are percents).

	Day 1		Day 3	
Antimicrobials	Single	Pool of 5	Single	Pool of 5
Tet	39 (21.91)	15 (13.89)	42 (23.46)	24 (22.86)
Sul, Tet	10 (5.62)	1 (0.93)	6 (3.35)	7 (6.67)
Str, Sul, Tet	6 (3.37)	1 (0.93)	3 (1.68)	4 (3.81)
Sul	4 (2.25)		6 (3.35)	9 (8.57)
Str, Tet	3 (1.69)	3 (2.78)	2 (1.12)	
Nal, Tet	2 (1.12)	2 (1.85)	2 (1.12)	
Amp, Str, Sul, Tet, Tri	1 (0.56)			
Amp, Str, Sul, Tet	1 (0.56)		1 (0.56)	
Amp, Sul	1 (0.56)		1 (0.56)	
Amp	1 (0.56)			
Ceft, Tet	1 (0.56)			1 (0.95)
Cep, Str, Sul, Tet	1 (0.56)		1 (0.56)	2(1.90)
Nal, Str, Sul, Tet	1 (0.56)	1 (0.93)	1 (0.56)	
Nal, Sul, Tet, Tri	1 (0.56)	1 (0.93)		
Nal, Sul, Tet,	1 (0.56)			
Sul, Tri	1 (0.56)			
Amp, Kan, Str, Tet			1 (0.56)	1 (0.95)

Table 4.4 (cont.)

Amp, Str, Tet		1 (0.93)		
Amp, Sul, Tet		1 (0.93)		2 (1.90)
Amp, Tet		1 (0.93)	1 (0.56)	
Chl, Nal, Str, Sul, Tet			1 (0.56)	
Chl, Sul, Tet			1 (0.56)	
Chl, Tet			1 (0.56)	
Cip		1 (0.93)		
Nal			1 (0.56)	1 (0.95)
Str, Sul, Tet, Tri			1 (0.56)	
Str, Sul,			1 (0.56)	
Str			3 (1.68)	2 (1.90)
Total isolates** (n)	178	108	179	105

Amo= Amoxicillin-Clavulanic acid, Amp=Ampicillin, Ceft=Ceftriaxone, Cep=Cephalothin, Chl=Chloramphenicol, Kan=Kanamycin, Nal=Nalidixic acid, Str=Streptomycin, Sul=Sulfamethoxazole, Tet=Tetracycline, Tri= trimethoprim-sulfamethoxazole

*Number of isolates; percent of the total number of isolates is shown in parantheses

**Isolates with missing SIR data for any antimicrobial are not included in this table.

had ten resistance phenotypes in common while pooled samples had only four resistance phenotypes in common. Eleven resistance phenotypes occurred in only one of the period/pooled sampling combinations. All eleven of these resistance phenotypes were represented by a single isolate.

The descriptive statistics for MICs depicted a consistent pattern across the four different sampling combinations (combinations of Day1, Day3 and single/pooled samples) for the majority of antimicrobials (Table 4.5). For nine of the antimicrobials the MIC values were predominated (>90%) by a single dilution value, the lower end of the detection limit for the plates. The design-adjusted chi-square test revealed significant differences between the sampling combinations in the MIC values for three antimicrobials, chloramphenicol ($P<0.001$), ciprofloxacin ($P=0.006$), sulfamethoxazole ($P=0.013$). The MIC values for chloramphenicol for isolates from Day3 were higher than those for the samples taken from Day1 although the differences were not large. The MIC distribution for ciprofloxacin was skewed to the right because the lowest value represented 90 percent or more of the isolates from both periods and in both single and pooled samples. The skewness resulted in sparse data at higher MIC values. A single isolate with an MIC value of 4.0 μg/ml for ciprofloxacin was found in a Day1 pooled sample and the next highest MIC value for any isolate was 0.25 μg/ml. Isolates from Day3 pooled samples appeared to have lower MIC values for streptomycin than those of the single samples or Day1 pooled samples although all the distributions appeared to be bimodal. The distribution for tetracycline MIC values also tended to be bimodal although a slightly

Table 4.5. Descriptive statistics of the MIC values for 17 antimicrobials for single and pooled samples from each period. *90 percent of more of the isolates in each period and method had the same MIC value, **mean is geometric mean

Antimicrobial	Day 1				Day 3			
	Single		Pool of 5		Single		Pool of 5	
	Mean**	MIC_{50}	Mean	MIC_{50}	Mean	MIC_{50}	Mean	MIC_{50}
Amikacin*	4.03	4	4.03	4	4.00	4	4.08	4
Amoxocillin- Clavulanic acid	2.65	2	2.64	2	2.44	2	2.57	2
Ampicillin	2.69	2	2.79	2	2.68	2	2.89	2
Apramycin	2.93	4	2.79	2	2.66	2	2.64	2
Cefoxitin*	4.13	4	4.03	4	4.14	4	4.24	4
Ceftiofur*	0.50	0.5	0.50	0.5	0.50	0.5	0.50	0.5
Ceftriaxone*	0.26	0.25	0.25	0.25	0.25	0.25	0.27	0.25
Cephalothin	5.21	4	5.14	4	4.73	4	5.28	4
Chloramphenicol	5.07	4	5.22	4	6.20	8	6.35	8
Ciprofloxacin*	0.02	0.015	0.02	0.015	0.02	0.015	0.02	0.015
Gentamicin	0.39	0.50	0.39	0.50	0.39	0.50	0.41	0.50
Kanamycin*	16.00	16	16.10	16	16.19	16	16.32	16
Nalidixic acid*	4.57	4	4.70	4	4.52	4	4.19	4
Streptomycin*	34.47	32	33.47	32	34.44	32	34.18	32
Sulfamethoxazole	161.90	128	140.15	128	155.68	128	180.42	128
Tetracycline	9.07	8	7.64	4	8.45	4	8.89	4
Trimethoprim*	0.13	0.12	0.12	0.12	0.13	0.12	0.12	0.12

higher percentage (31.4%) of the isolates from the Day3 pooled samples occurred at the highest MIC value (32) as compared to the other sampling combinations.

Discussion

The assumption of short-term repeatability in measuring antimicrobial resistance in isolates obtained from feedlot cattle is critical for allowing researchers to interpret susceptibility results and, ultimately, in making longitudinal and monitoring studies practical in terms of both human and economic resources. In this study, several criteria, based on both SIR and MIC data, were used to evaluate short-term repeatability. Prevalence of resistance to individual antimicrobials was consistent between the two periods and between single and pooled samples when the resistance prevalence for the tested isolates was at least two percent. However, for the antimicrobials to which resistance occurred rarely, the detection of resistance using either the single or pooled samples was inconsistent between the two periods. The multiple resistance count was similar for the two periods and for the single versus pooled samples but evaluation of the antibiograms revealed that there were substantial differences in resistance phenotypes detected with the different sampling methods. The ranges and measures of central tendency for MIC values were fairly consistent for both the two periods and the single/pooled samples.

Pooling can result in a cost savings due to reduced culture costs but there are disadvantages when sampling for antimicrobial resistance. A recent study demonstrated that pooling may not be practical because of the limitations placed on the number of

isolates that can be tested from a single pool.[10] Bias in prevalence estimates could potentially be introduced if unequal portions of fecal material from different cattle are mixed or if there is some competitive advantage for specific resistance phenotypes in culture. Also, when pooled samples are mixed, it may not be possible to completely disperse clusters of cells even from a single animal.[1,10] Consequently, when subsamples are taken for culture there is a predisposition for sampling from one or two animals rather than a truly equal probability of sampling organisms from all animals. Results from this study, however, do not indicate the presence of a consistent bias in the ability of pooled samples to estimate antimicrobial resistance relative to the individual samples. Neither pooled nor individual samples consistently detected the rarer resistance phenotypes in this study.

Repeatability of antimicrobial resistance outcomes can be affected by factors related to the sampling design, such as sample size and susceptibility testing processes, or by actual changes in the antimicrobial susceptibility pattern of the bacterial populations. Although a change in the baseline antimicrobial susceptibility in the populations cannot be ruled out, there was no evidence that external factors had been altering the population directly over the two-day study period. In the two weeks prior to sampling, only one animal in the six pens was treated parenterally with an antimicrobial (ceftiofur). All of the cattle had been on feed in their respective pens for at least 129 days (average body weight>1143 pounds). The cattle had been on the same finishing ration for approximately 100 days.

If the inconsistencies in antimicrobial resistance measurements observed in this study, especially in the detection of rare resistance phenotypes, are not readily attributable to shifts in the microbial population then sampling issues must be considered. A recent study demonstrated that differences in the presence/absence of resistant phenotypes in samples taken from feedlot cattle were attributable to sample sizes that were probably insufficient for detecting rare resistance phenotypes.[10] Another study found that the number of serologic types found in human fecal samples increased with the larger numbers of isolates tested.[9] In a third study, fecal swabs were obtained from 20 pigs daily for 32 days.[11] Twenty-five *Escherichia coli* isolates were tested for antimicrobial susceptibility to two antimicrobial drugs. Substantial variation was observed in the proportion of isolates that were resistant throughout the period. The conclusion was that feces must be sampled daily for not less than 25 days to account for the "random appearance" of resistant phenotypes that persist for only a day or two. Sampling at intervals of five to seven days was not recommended because of the observed variability. The laboratory burden of their sampling protocol (20 pigs * 25 isolates * 32 day), notably, was very time and resource intensive.

The sampling design utilized in this study was based on a proposed design for a longer term monitoring program that was intended to have repeated samples taken from a set of feedlot cattle pens over the duration of the feeding period. Our study demonstrated that short-term variation in bacterial resistance measures appears to be relatively small. Reduction of variation in the short-term potentially should allow for these sampling methods to detect biologically and statistically significant changes in resistance measures affected by management or environmental factors over longer time periods. However, if

the objective of the monitoring program is to detect changes in the occurrence of rarer resistance phenotypes, in the absence of management or environmental interventions which would substantially alter the prevalence, then the short-term variation likely will make it difficult to detect long-term changes.

Simply increasing the number of isolates taken from each agar plate appears to be one solution for increasing the sensitivity of the monitoring program to detect changes in the prevalence of rare resistance phenotypes. However, the potential for clustering of antimicrobial resistance phenotypes within a sample suggests a non-linear response in the amount of information gained from increasing the number of isolates within a fecal sample. One study of antimicrobial resistance in swine found that the precision of resistance estimates did not improve noticeably when more than five colonies were selected per swine fecal sample.[2] Another study demonstrated clustering of antimicrobial resistance phenotypes among isolates from feedlot cattle fecal samples and suggested that fewer than five isolates per sample is a reasonable guideline.[10] They suggested increasing the number of fecal samples, if possible, to increase the total number of isolates being tested for susceptibility.

Resource constraints, given the current culture and susceptibility testing methodologies, make it impractical to sample large numbers of isolates and evaluate susceptibility on a daily basis when the objective is to determine long-term trends. Therefore, the strategy of taking samples taken at a given point in time is one way to represent the susceptibility patterns during the peri-sampling period. The results of our study indicate that short-term

stability makes this a plausible assumption. This study also indicates, however, that researchers should be cautious about the interpretation of the presence/absence of rare resistance phenotypes or small fluctuations in prevalence and central tendency measures. Intermittent occurrence of rare isolates may or may not be reflecting actual trends in the antimicrobial resistance pattern of the bacterial population. Small differences in the occurrence of rarer resistance phenotypes over long time periods, thus, may not be reflecting actual differences in prevalence.

Sources and manufacturers

a. Fluorocult MacConkey agar, EM Science, Gibbstown, NJ.

b. Sensititre, Trek Diagnostics, Westlake, Ohio.

c. Stata, version 7.0, Stata Corp., College Station, Texas.

d. SAS, version 8.1, SAS Institute Inc., Cary, NC.

References

1. Cannon, R.M. and T.J. Nicholls. 2002. Relationship between sample weight, homogeneity, and sensitivity of fecal culture of *Salmonella enterica*. *J. Vet. Diag. Invest.* 14:60-62.

2. Dunlop, R.H., McEwen, S.A., Meek, A.H., Friendship, R.M., Black, W.D., and Clarke, R.C. 1999. Sampling considerations for herd-level measurement of faecal *Escherichia coli* antimicrobial resistance in finisher pigs. *Epidemiol. Infect.* 122:485-496.

3. FDA-CVM. A proposed framework for evaluating and assuring the human safety of the microbial effects of antimicrobial new animal drugs intended for use in food-producing animals. 2002.

4. Marano, N.N., Rossiter, S., Stamey, K., Joyce, K., Barrett, T.J., Tollefson, L.K., and Angulo, F.J. 2000. The National Antimicrobial Resistance Monitoring System (NARMS) for enteric bacteria, 1996-1999: surveillance for action. *J. Am. Vet. Med. Assoc.* 217:1829-1830.

5. National Antimicrobial Resistance Monitoring Program - Enteric Bacteria. 1997. Veterinary Isolates Final Report. *Athens, GA: USDA:ARS.* 20-26.

6. National Committee for Clinical Laboratory Standards (NCCLS). 1999. Performance standards for antimicrobial susceptibility testing: Ninth informational supplement. NCCLS document M100-S9. Wayne, PA: National Committee for Clinical Laboratory Standards. 90-93.

7. StataCorp. 2001. Stata statistical software: Release 7.0. Stata Corporation, College Station, TX.

8. Tollefson, L., F.J. Angulo, and P.J. Fedorka-Cray. 1998. National surveillance for antibiotic resistance in zoonotic pathogens. *Vet. Clin. North Am.: Food Anim. Pract.* 14:141-150.

9. Vosti, K.L., A.S. Monto, and L.A. Rantz. 1962. The importance of sample size in studies based upon the serologic classification of *Escherichia coli*. *Proc. Exper. Biol. Med.* iii:201-204.

10. Wagner B.A., D.A. Dargatz, M.D. Salman, P.S. Morley, T.E. Wittum, and T.J. Keefe. 2002. Sampling Techniques for Antimicrobial Susceptibility of Enteric *Escherichia coli* Recovered From Feedlot Cattle. *A. J. Vet.Res.* 63:1662-1670.

11. Walton, J.R. 1972. A dynamic study of drug resistance in populations of faecal *Escherichia coli* from pigs fed nitrovin and furazolidone continuously. *Zbl. Vet. Med. B.* 19:646-654.

CHAPTER 5

FACTOR ANALYSIS OF MINIMUM INHIBITORY CONCENTRATIONS FOR *ESCHERICHIA COLI* ISOLATED FROM FEEDLOT CATTLE TO MODEL RELATIONSHIPS AMONG ANTIMICROBIAL RESISTANCE OUTCOMES

Summary

Factor analysis was used to assess relationships in the minimum inhibitory concentration among 17 antimicrobials tested on isolates of *Escherichia coli* isolated from 360 faecal samples obtained from feedlot cattle. Six factors were extracted using maximum-likelihood factor analysis. The factors were interpretable antimicrobial groupings based on class of antimicrobial and previously described associations. New generation cephalosporins, older generation beta-lactams, fluoroquinolones and aminoglycosides grouped separately as classes of antimicrobials on four of the six factors. One of the remaining factors was a grouping of antimicrobials that had been identified as being related in previous feedlot studies. The last factor was a grouping of three of the five antimicrobials that comprise the antimicrobials found in penta-resistant strains of *Salmonella* Typhimurium. The factor analysis described patterns in the MIC data that would not have been apparent if only susceptible-resistance categorized antimicrobial-resistance data had been analyzed.

Introduction

Antimicrobial resistance, especially in pathogenic bacteria, is a global problem that has emerged in the past two decades (Blondeau and Vaughan 2000; Aarestrup 2000). Resistance in bacterial isolates from agricultural animals has raised concern both among public-health professionals and veterinarians. The public-health concern is that antimicrobial-resistance genes can be incorporated into the genome of potential human pathogens (e.g. *Salmonella, Campylobacter, Escherichia coli*) and transmitted through food-borne routes to humans. In addition, there is concern that resistance genes in non-pathogenic bacteria could be transmitted to humans where the genes would be available to non-resistant pathogenic bacteria in the human gastrointestinal tract. Veterinarians and other animal-health authorities are concerned about their ability to control bacterial infections in domestic animals and about the zoonotic potential. These concerns have motivated considerable effort in both human medical and veterinary areas to identify emerging resistance, estimate prevalence, and monitor trends (Sahm et al. 2001; Jones 2000; Cavallo et al. 2000; Smith et al. 1999).

The importance of resistance surveillance mandates the development of effective and efficient monitoring systems (Sahm et al. 2001). Within the framework of surveillance, there is a need to refine statistical methods for identifying patterns of resistance and to be able to communicate the results to the scientific community (Jones 2000). Surveillance systems can generate large amounts of data that must be distilled into meaningful summaries. An important attribute of resistance surveillance data is the occurrence of multiple outcomes per individual bacterial isolate. Multiple outcomes typically arise

when a panel of antimicrobials is tested on each isolate. Methods that have been used to address the multiple outcomes include antibiograms, indices (Krumperman 1983; Kaspar et al. 1990), and discriminant analysis (Wiggins 1996).

Antibiograms are essentially a tabulation of the occurrence of antimicrobial resistance. Resistance to a specific antimicrobial can occur singly or jointly with other antimicrobials. Interpretation of antibiograms depends on visual examination of the tabulation to identify patterns. Indices typically are a summary value of the richness, the number of antimicrobials to which an isolate is resistant, or diversity of the pattern being examined. The summary value yields little information concerning the patterns of resistance in the data. Discriminant analysis is the only method that addresses the multiple outcomes with multivariate statistical methods. The objective of the discriminant analysis was to construct a linear function of resistant/non-resistant responses to multiple antimicrobials to separate isolates into categories for identification of sources such as in the case of non-point source pollution (Wiggins 1996).

Our objective was to describe the use of factor analysis in identification of patterns in antimicrobial minimum-inhibitory concentrations. Specifically, factor analysis was applied to identify patterns within a panel of 17 antimicrobials used to test bacterial isolates of *E. coli* from feces of feedlot cattle. The usefulness of the method was assessed by the ability to identify patterns of both susceptibility and resistance that had interpretations based on the biology of resistance development and antimicrobial action.

Materials and methods

Study population

Faecal samples were collected from cattle housed in six pens at a commercial feedlot in Colorado on May 2, 2000. Each pen had approximately 10,000 square feet of area and contained between 54 and 56 beef steers. Cattle in three of the pens had been on feed for 75 days while the other three pens had been on feed for 81 days. All cattle were receiving tylosin in their rations at a rate of 10 gram per ton of feed (approximately 60-90 mg per head per day). One animal had received an injection of an antimicrobial (micotil) in the week prior to sampling. No other cattle had antimicrobial injections in the month prior to sampling. Within each pen, 30 fresh faecal samples (approximately 50g each) were collected from individually identifiable faecal pats on the ground. Disposable plastic gloves were used to pick up faecal material and were changed between samples. Efforts were made to collect the faecal samples from sites throughout the pen to minimize the opportunity for collecting multiple faecal samples from the same animal.

Additionally, while the pen floor samples were being collected, 30 faecal samples were obtained per rectum from randomly selected individual animals from each of the six pens while the cattle were being restrained in a chute. The randomization was performed by assigning random numbers to the count of cattle in the pen (numbers ranging from 1 to 54 or 56 depending on pen size). The random numbers were sorted and the top 30 were chosen for sampling. The count value was translated to the order going through the chute. All samples were cooled and transported to the laboratory for further processing within 4 hours of collection.

In the laboratory, approximately 1 gram of faecal material from each sample was placed in a 50-ml conical tube. Ten ml of 0.5% normal saline was added to each tube. The tubes were shaken using a vortex to mix the saline and the faecal material. The tubes were then kept overnight in a refrigerator at 3°C. On the next day, the samples were shipped for overnight delivery to a second lab where the culturing and antimicrobial susceptibility testing was conducted.

Culture and antimicrobial susceptibility testing

Dilute faecal material was streaked on MacConkey-4-Methylumbelliferyl-β-D-Glucuronide agar plates to isolate colonies of *E. coli*. *Escherichia coli* was chosen as the species for antimicrobial resistance testing because of its ubiquitous nature. Plates were incubated at 37°C for 18-24 hours. Individual *E. coli* colonies were identified under ultraviolet light as lactose-positive (bright pink) and glucouronidase-positive (colony periphery had a bluish appearance). Five colonies were selected from each agar plate and transferred to individual nutrient agar slants and incubated for another 18-24 hours. The slants then were checked for growth and stored at 2-8°C until the individual isolates were tested for susceptibility. In preparation for antimicrobial susceptibility testing, MacConkey agar plates were inoculated with bacteria from the slants. Colonies were selected from these plates and placed into separate tubes with 5 ml of sterile water and turbidity was adjusted to a 0.5 McFarland standard. After mixing, 10 µl of the bacterial suspension was used to inoculate 10 ml of cation-adjusted Mueller-Hinton broth, which then was used to inoculate antimicrobial-sensitivity plates. The plates were incubated at

35°C for 18-24 hours and then read using a semi-automated system (Sensititre™). Based on growth in individual wells, the minimum inhibitory concentrations (MICs) against a panel of 17 antimicrobial drugs were determined. The 17 antimicrobials (Table 1) were selected to parallel the panel used by the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS-EB) in the U.S. The ATTC *E. coli* 25922 was used as a quality control organism. The MIC for each antimicrobial was used to classify the isolate as susceptible, intermediate, or resistant according to standards used to classify human isolates as determined by the National Committee on Clinical Laboratory Standards. (Breakpoints for veterinary MICs values are not currently available.)

Statistical analysis

All analyses were performed using version 8.1 of SAS[3]. All observations with missing MIC results for any of the 17 antimicrobials were deleted for all analyses. Descriptive statistics were computed to examine the univariate MIC distributions for the 17 antimicrobials. Some of the techniques we used assume a multivariate Normal distribution. To accommodate this assumption, data for individual variables should be approximately normally distributed. Skewness of the MIC distributions indicated the need for a transformation to more-closely approximate normal distributions. The data were log transformed to correct for the lack of normality. Shapiro-Wilk tests for normality were used to assess the normality assumption of the transformed data (Conover, 1980).

[3] SAS version 8.1: SAS Software, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513

Three factor-analytic methods were implemented to allow for comparison of fit criteria: principal-component (PC), unweighted least-squares (ULS), and maximum-likelihood (ML). A varimax rotation was used in each method. The fit criteria examined were root mean square (RMS) of the off-diagonal residual correlation and RMS of the off-diagonal partial correlations. The number of factors to keep in the factor analysis was determined by evaluating three criteria based on results from the PC and ML methods: a scree plot, a minimum eigenvalue of 1.00, and a large-sample test for the number of common factors (Johnson and Wichern, 1992). The scree plot, a plot of the factor number and its associated eigenvalue, was assessed by looking for a leveling of the slope of the curve. The large-sample test was used by sequentially fitting models with increasing numbers of factors kept. A chi-square test tested significance of adding factors using output from the maximum-likelihood factor analysis. Johnson and Wichern (1992) warn that large sample sizes and a small number of factors being examined compared to the number of variables entering the factor analysis can lead to retention of more common factors than are necessary to explore the data. Consequently, they suggest exercising judgment in selecting the number of factors since additional factors can be significant without providing much additional insight into data. The scree plot and the minimum eigenvalue cutoff of 1.00 came from the PC method while the chi-square tests were obtained from the ML method. The final model determined from transformed data was compared to a model developed from non-transformed data to assess the effect of the transformation on the factors. A P-value ≤ 0.05 was used in determining statistical significance.

Multiple isolates were obtained from a single faecal sample and, potentially, faecal samples could have originated from the same individual making the assumption of independence needed for inferential analysis problematic. We ran the factor model with randomly selected single isolates from each sample to assess the impact that five isolates from a single sample might have had on the model. Johnson and Wichern (1992) discuss an example where the lack of independence among repeated observations was evaluated in this manner.

An antibiogram table was developed for comparison with the factor-analysis results. Intermediate results were combined with susceptible results to create a dichotomous susceptible/resistant outcome for constructing the antibiogram.

Results

Minimum inhibitory concentration was measured on a total of 1737 isolates (one isolate lost to testing and 42 isolates with missing MIC data). The descriptive analysis of the untransformed data indicated that the MIC values were not normally distributed (Table 5.1). MIC distributions for almost all of the antimicrobials were highly skewed and all 17 of the Shapiro-Wilks tests for normality were highly significant ($P<0.001$). The log transformation improved the skewness problem for all 17 antimicrobials although the Shapiro-Wilks tests for normality still indicated significant deviations from normality ($P<0.001$) for the transformed MIC values for all antimicrobials. The deviations from normality are likely due to the discrete characteristics of the MIC data and the relatively large sample size which allows for detection of small deviations from normality.

Table 5.1. Descriptive statistics for MIC data (n=1737). MIC units are μg/ml.

Antimicrobial	5th percentile	Median	95th percentile	Maximum	Non-transformed skewness	Transformed skewness
Amikacin	4.0	4.0	4.0	16.0	23.6	18.4
Amoxicillin-Clavulanic acid	1.0	4.0	4.0	32.0	7.8	-0.1
Ampicillin	2.0	2.0	4.0	32.0	6.7	2.8
Apramycin	2.0	4.0	4.0	16.0	2.3	0.0
Cefoxitin	4.0	4.0	4.0	32.0	11.9	7.7
Ceftiofur	0.5	0.5	0.5	4.0	14.8	13.3
Ceftriaxone	0.25	0.25	0.25	8.0	18.2	14.4
Cephalothin	2.0	4.0	16.0	32.0	3.1	0.0
Chloramphenicol	4.0	8.0	8.0	32.0	4.5	1.0
Ciprofloxacin	0.015	0.015	0.015	0.025	14.4	8.3
Gentamicin	0.25	0.50	0.50	4.0	5.7	0.5
Kanamycin	16.0	16.0	16.0	64.0	9.7	9.5
Nalidixic acid	4.0	4.0	4.0	128.0	12.8	9.5
Streptomycin	32.0	32.0	64.0	256.0	6.6	3.9
Sulfamethoxazole	128.0	128.0	512.0	512.0	1.3	1.3
Tetracycline	4.0	4.0	32.0	32.0	1.0	0.9
Trimethoprim	0.12	0.12	0.25	4.0	16.2	3.2

The antibiogram revealed 36 different combinations of resistance to antimicrobials (Table 5.2). Most of the isolates (1148/1737; 66.1%) were susceptible to all antimicrobials; resistance to a single antimicrobial was indicated in 189 (10.9%) of the isolates. No resistance was ever observed for 6 antimicrobials (amikacin, apramycin, ceftiofur, ceftriaxone, ciprofloxacin, and gentamicin). Most of the antimicrobial-resistance combinations were relatively rare (<1.0% of isolates). Varying combinations of resistance to streptomycin, sulfamethoxazole, tetracycline, and trimethoprim-sulfamethoxazole predominated.

Based on the scree plot, five to six factors would be recommended although most of the variation in the data was represented in three to four factors (Figure 5.1). Based on the criteria of keeping eigenvalues ≥1.0, six factors should be kept. The large-sample chi-square tests used to check for sequential addition of factors indicated that at least 7 factors could be kept in the analysis (Table 5.3). We chose to keep 6 factors.

The best fit between the six factor models was judged by examining root mean square of the off-diagonal residual correlation and root mean square of the off-diagonal partial correlations (Table 5.4). These two criteria are used to compare the relative ability of the models to fit the observed correlation matrix (1992). Larger values for both the criteria imply a poorer fit because more of the original correlation matrix is left unexplained by the factor analysis. The PC method had the poorest fit based on these two diagnostics. The other two methods had comparable values for the two diagnostics.

Table 5.2. Frequency distribution of resistance groupings (n=1737).

Antimicrobials	Frequency	Percent
Amo, Amp, Cef, Cep, Chl, Str, Sul, Tet	6	0.35
Amo, Amp, Cep, Chl, Str, Sul, Tet	1	0.06
Amo, Amp, Cep, Str, Sul, Tet	1	0.06
Amo, Amp, Cep, Tet	1	0.06
Amo, Amp, Cep	3	0.17
Amo, Cep	1	0.06
Amp, Cep, Tet	2	0.12
Amp, Chl, Str, Sul, Tet	10	0.58
Amp, Chl, Str, Sul	2	0.12
Amp, Str, Sul	1	0.06
Amp, Sul, Tet	1	0.06
Amp, Sul	1	0.06
Cef, Sul, Tet	1	0.06
Cef	1	0.06
Chl, Str, Sul, Tet, Tri	2	0.12
Chl, Str, Sul, Tet	5	0.29
Chl, Sul, Tet	1	0.06
Chl, Sul	1	0.06
Chl	2	0.12
Kan, Str, Sul, Tet	8	0.46
Kan, Str, Tet	2	0.12
Kan, Sul, Tet	3	0.17
Kan, Sul	2	0.12
Kan	2	0.12
Nal, Str, Sul, Tet	2	0.12
Nal, Str, Tet	1	0.06
Nal, Sul, Tet	11	0.63
Nal	3	0.17
Str, Sul, Tet	140	8.05
Str, Sul	1	0.06
Str, Tet	15	0.86
Str	4	0.23
Sul, Tet	175	10.07
Sul	26	1.50
Tet	151	8.69

Amo= Amoxicillin-Clavulanic acid, Amp=Ampicillin, Cef=Cefoxitin, Cep=Cephalothin, Chl=Chloramphenicol, Kan=Kanamycin, Nal=Naladixic acid, Str=Streptomycin, Sul=Sulfamethoxazole, Tet=Tetracycline, Tri= trimethoprim-sulfamethoxazole

Figure 5.1. Scree plot depicting the eigenvalues for eight factors. Factors with eigenvalues of at least one were kept in the final factor model.

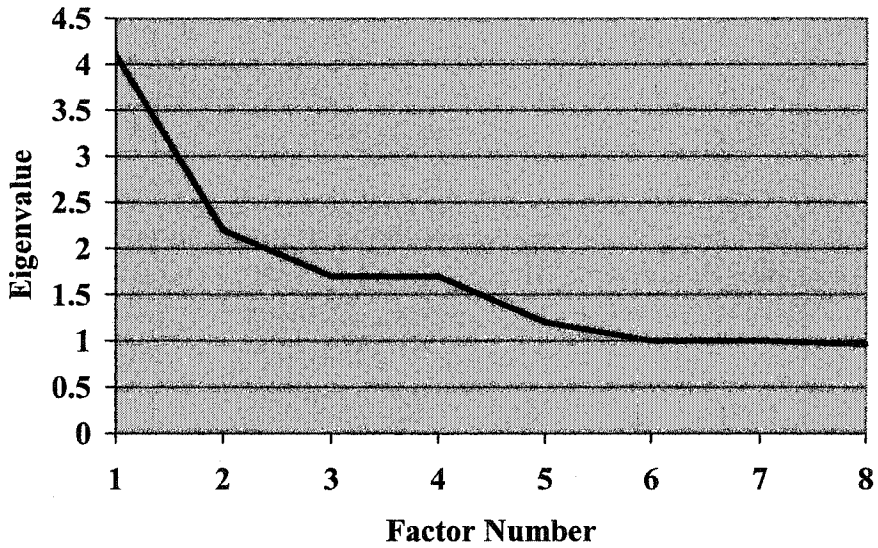


Table 5.3. Large-sample chi-square tests for the number of factors to keep in maximum-likelihood factor analysis model.

Factor to be added	Chi-square	df	P-value
2	3348.4	16	<0.0001
3	1647.0	15	<0.0001
4	1212.8	14	<0.0001
5	717.5	13	<0.0001
6	396.1	12	<0.0001
7	77.5	11	<0.0001

Table 5.4. Root mean square (RMS) for the off-diagonal residual and partial correlations for the six factor models.

Factor analysis method	Residual RMS	Partial correlation RMS
Maximum likelihood	0.01	0.03
Principal components	0.06	0.21
Unweighted least squares	0.01	0.03

The loadings for the six factors nevertheless were examined for all three models. The factor loadings were very similar between the ULS and the ML methods and, thus, only the ML and PC method loadings are presented here (Table 5.5). The PC loadings differed from the results for the other two methods primarily by a switching of factor 1 and factor 2. Also, kanamycin loaded on the sixth factor extracted using the PC method but did not load on any factor using the other methods. Because of fit criteria, the consistencies between the ULS and ML factor loading patterns, and the general agreement across all three methods, the remaining discussion of results will focus on the factor analysis results based on the ML method.

Two antimicrobials, ceftiofur and ceftriaxone, loaded very high (loading>0.90) on the first factor while a third antimicrobial, cefoxitin, loaded moderately high (loading>0.50). All three of these antimicrobials are cephalosporins (a type of beta-lactam), both second and third generation. Thus, factor one can be described as "newer generation cephalosporins." Sulfamethoxazole and tetracycline loaded heavily on factor two while trimethoprim-sulfamethoxazole and, to a lesser extent, streptomycin loaded moderately high. These antimicrobials do not come from a single class of antimicrobials -- but resistances to these two drugs occur together frequently in bacterial isolates. Factor two also can be considered to be a group of antimicrobials with higher prevalences of resistance than other antimicrobials on the panel. Factor three was dominated by amoxicillin-clavulanic acid and cephalothin with a lesser influence by ampicillin. All three of these antimicrobials are older generation beta-lactams. Only ciprofloxacin and nalidixic acid loaded high on factor four. These are the only fluoroquinolones in the

Table 5.5. Loadings* for the six factors from the maximum-likelihood and principal component factor analysis of minimum inhibitory concentrations.

Antimicrobial	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Maximum-likelihood loadings						
Amikacin	-0.010	0.039	0.033	0.007	-0.041	0.011
Amoxicillin-Clavulanic acid	0.166	0.088	**0.769**	0.023	0.120	0.192
Ampicillin	0.196	0.092	**0.470**	0.025	0.064	**0.557**
Apramycin	0.017	0.070	0.138	0.015	**0.648**	0.098
Cefoxitin	**0.552**	0.015	0.252	-0.025	-0.007	0.192
Ceftiofur	**0.955**	0.056	0.082	0.016	0.031	0.133
Ceftriaxone	**0.949**	0.049	0.053	0.030	0.028	0.137
Cephalothin	0.125	0.152	**0.807**	0.036	0.124	0.029
Chloramphenicol	0.148	0.053	0.074	-0.028	0.039	**0.486**
Ciprofloxacin	0.024	0.034	0.035	**0.893**	0.042	-0.026
Gentamicin	-0.007	0.047	0.115	0.014	**0.891**	0.049
Kanamycin	-0.016	0.125	0.000	-0.006	-0.019	0.009
Nalidixic acid	-0.008	0.037	0.024	**0.776**	-0.026	-0.006
Streptomycin	0.266	**0.537**	0.021	-0.012	0.146	**0.495**
Sulfamethoxazole	0.068	**0.891**	0.081	0.111	0.066	0.031
Tetracycline	0.055	**0.809**	0.170	0.110	0.074	0.011
Trimethoprim-Sulfamethoxazole	0.084	**0.695**	0.083	-0.061	0.058	0.097
Principal-component method						
Amikacin	0.034	-0.061	0.107	0.002	-0.135	0.135
Amoxicillin-Clavulanic acid	0.100	0.156	**0.863**	0.020	0.114	0.029
Ampicillin	0.087	0.294	**0.637**	0.007	0.093	**0.341**
Apramycin	0.084	0.014	0.135	0.011	**0.869**	0.007
Cefoxitin	0.025	**0.715**	0.296	-0.036	-0.024	0.006
Ceftiofur	0.099	**0.938**	0.081	0.022	0.032	0.015
Ceftriaxone	0.093	**0.940**	0.053	0.034	0.028	0.020
Cephalothin	0.170	0.082	**0.857**	0.038	0.085	-0.065
Chloramphenicol	-0.024	0.244	0.139	-0.055	0.116	**0.752**
Ciprofloxacin	0.041	0.019	0.029	**0.918**	0.043	0.005
Gentamicin	0.083	-0.023	0.130	0.011	**0.873**	-0.006
Kanamycin	0.103	-0.120	-0.049	0.028	-0.099	**0.548**
Nalidixic acid	0.039	-0.002	0.022	**0.914**	-0.023	-0.019
Streptomycin	**0.608**	0.346	0.064	-0.036	0.183	**0.402**
Sulfamethoxazole	**0.892**	0.021	0.078	0.114	0.034	0.089
Tetracycline	**0.852**	0.014	0.165	0.116	0.031	0.025
Trimethoprim-Sulfamethoxazole	**0.825**	0.072	0.083	-0.103	0.026	-0.023

*Moderate and high loadings are in bold.

panel. Consequently, factor four is a grouping of fluoroquinolones. Gentamicin loaded high while apramycin loaded moderately high on factor 5. Both antimicrobials are in the aminoglycoside class of antimicrobials. Other aminoglycosides, specifically amikacin, kanamycin, and streptomycin did not load high on factor 5 (amikacin and kanamycin actually had negative loadings on factor 5). Ampicillin, chloramphenicol, and streptomycin represent the highest loadings on factor 6. Chloramphenicol had its highest loading on factor 6 while the loadings in factor 6 for ampicillin and streptomycin were similar to those in factor 3 and factor 2, respectively.

Neither amikacin nor kanamycin (both aminoglycosides) loaded strongly on any of the factors created using either the ML or the ULS methods. Amikacin dominated the sixth factor created by the PC method.

The ML method was run to extract six factors from the untransformed data. The resulting loading patterns were the same as obtained from the transformed data although the order of the factors was mixed and streptomycin and trimethoprim-sulfamethoxazole did not load as highly with sulfamethoxazole and tetracycline.

Maximum-likelihood models using a randomly selected single isolate consistently mirrored the results that we obtained when all five isolates were available to the model. However, because of the reduced number of isolates in the single isolate model, occasionally the model would not solve because of singularities in the model (all MIC values the same for at least one antimicrobial).

Discussion

Patterns in minimum inhibitory concentrations identified by factor analysis essentially fell into two types of relationships: a link amongst antimicrobials within the same class (e.g. cephalosporins, fluoroquinolones, or aminoglycosides) and groupings of antimicrobials (not necessarily from the same class) that were consistent with susceptibility/resistance associations in other studies of antimicrobial resistance in beef cattle.

The ability to attach specific interpretations to the factors demonstrates the value of this type of analysis. Factor one is a grouping of newer generation cephalosporins. Under selection pressure for the development of resistance for at least one new generation cephalosporin, it is logical that resistance might emerge for all the cephalosporins. Similarly, in the absence of selection pressure for this class of antimicrobials it is also logical to find them loading together on the same factor. Thus, factor one represents a susceptibility pattern in the data that might not be apparent when reviewing the antibiogram (which is based solely on resistance).

Factor two differs from factor one in that the former represents a mixture of antimicrobial classes. Three of the four antimicrobials in this group had higher univariate MIC values (Table 5.1) and the highest frequency of resistance (Table 5.2) among the antimicrobials tested. However, high MIC values and frequency of resistance does not appear to be the reason that trimethoprim/sulfamethoxazole was correlated with the other antimicrobials.

Other studies of resistance patterns of enteric bacteria from feedlot cattle (Dargatz et al., 2002) and beef cows (Dargatz et al., 2000) have demonstrated the co-occurrence of streptomycin and trimethoprim/sulfamethoxazole resistance as well as sulfamethoxazole and tetracycline resistance.

The third factor was comprised of antimicrobials from the older group of beta-lactams. As with factor one, this grouping appears to be a grouping of antimicrobials from the same class. One difference between the antimicrobials that loaded high in factor 3 and factor 1, respectively, is that the MIC values tend to be higher in the antimicrobials with higher loadings on the factor 3. In factor 3, the antimicrobials with high loadings have been in use for a longer period of time and, perhaps, are more frequently used in feedlot operations. Approximately 50% of large feedlots (≥8,000 head) use Cephalosporins (ceftiofur) to treat respiratory disease (USDA, 2000) -- one of the most common diseases in feedlot cattle. However, only about 9 percent of feedlots use it as their primary initial treatment (USDA, 2000) -- so this class of antimicrobials is being used only in special circumstances as a primary treatment. Treatment records from the study feedlot indicated beta-lactams were not used as a primary initial treatment for respiratory disease.

Factor 4 was dominated by the fluoroquinolones, a new class of antimicrobials of special concern because resistance has not yet developed widely in the human or animal populations (Sahm et al., 2001; Smith et al., 1999; Thornsberry et al., 1999). Nalidixic acid, the older of the two fluoroquinolones, had relatively higher MIC values which were reflected in the presence of resistance in some isolates but ciprofloxacin resistance was

absent in these 1737 isolates. Consequently, the factor analysis was able to identify a correlation between the two fluoroquinolones using MIC data that would not have been apparent using SIR data. Ruiz et al. (Ruiz et al., 2002) demonstrated a relationship between naladixic acid resistance prevalence and ciprofloxacin MIC values in human *E. coli* isolates. In their study, increasing prevalence of nalidixic acid resistance was associated with increasing ciprofloxacin MIC values in ciprofloxacin-susceptible isolates.

Factor 5 is problematic because only two of the five aminoglycosides, apramycin and gentamicin, loaded high on the factor (amikacin, kanamycin, and streptomycin did not). Streptomycin is an older antimicrobial and its common link with other antimicrobials was strongly represented in Factor 2. Amikacin is a synthetic derivative of kanamycin so its activity against bacteria is similar (Prescott et al., 2000). Consequently, it is not surprising to see the two behave similarly in a factor analysis. However, the relationship, or lack thereof, between these two antimicrobials and the two that loaded heavily on this factor is not well understood.

The antimicrobials in factor 6, in the same manner as factor 2, come from different classes of antimicrobials. The grouping of antimicrobials in factor 6 parallels an important combination of antimicrobial resistance found in *Salmonella* Typhimurium. An epidemic strain of *S.* Typhimurium definitive type 104 has a penta-resistant combination to ampicillin, chloramphenicol, streptomycin, sulphonomides and tetracyclines and a worldwide distribution (Davis et al., 1999). The latter two antimicrobials did not load substantially on factor 6 but loaded heavily on factor 2.

Six antimicrobials (amikacin, apramycin, ceftiofur, ceftriaxone, ciprofloxacin, and gentamicin) did not contribute information to the antibiogram because all isolates were susceptible. When the MIC values were used in the factor analysis, only amikacin was not related substantially to any of the 6 factors. The categorization of MIC values into susceptibility/resistance categories for creation of the antibiogram resulted in a loss of information.

Many reports present only the dichotomy of susceptibility/resistance for each antimicrobial on the panel that was tested in the study along with the breakpoint for designation (Walker and Thornsberry. 1998). Solely reporting descriptive information univariably on resistance/susceptibility has some disadvantages. First, the MIC breakpoints could change -- thereby limiting the value over time of studies, which dichotomized the data (Walker and Thornsberry, 1998). Secondly, dichotomization reduces the amount of information that is available for analysis and interpretation. Shifts in MIC values might become apparent or be of interest before "cutpoints" are crossed. Jones (2000) presents a case for showing the MICs as cumulative distributions. Lastly, the biological processes underlying resistance development and transmission lead to resistance/susceptibility linkages among antimicrobials that are not easily identifiable when univariate summaries are presented.

The potential for violation of assumptions when using MIC data in factor analysis should be assessed. The MIC data are not truly continuous data and might even be left or right

truncated. These characteristics -- along with the potential for skewness in the distributions -- make it difficult to assume that the underlying distribution is normal. The log transformation we used in this study removed only some of the skewness. These deficiencies make the assumption of multivariate normality somewhat tenuous unless, with the large sample size in the current study, the multivariate version of the Central-Limit Theorem is invoked (Morrison, 1976). The similarity of the factors obtained using the PC and the ULS methods -- neither of which depend on multivariate normality for factor extraction (Johnson and Wichern, 1992; Harman and Jones, 1966) -- with the factors obtained using the ML method suggests that the ML method provided valid results even if the multivariate normal assumption was violated. Similarly, the potential lack of independence due to multiple isolates per sample did not adversely impact the factor analytic results but rather the increased number of isolates added stability to the factor extraction process.

Factor loadings using the ML method were similar for the transformed and non-transformed data implying the transformation did not substantially alter inherent patterns in the data. Also, in situations where inferential analyses are not being conducted, it may not be necessary to transform MIC data for use in any of the factor analysis methods.

Conclusions

Factor analysis provides insight into correlations, if they exist, among the MIC patterns for the antimicrobials used in a panel. In this sense, the method can be applied to any panel being used to explore patterns of resistance associated with any bacterial isolates.

In addition, when designing future antimicrobial testing studies, a cost effective alternative might be to consider accounting for the correlations in the data. Antimicrobial resistance testing is fairly expensive, especially when the number of isolates being tested is large. If certain antimicrobials are strongly correlated, then it might be possible to omit some of the antimicrobials from the panel without substantially reducing the amount of information that is available. Also, as was the case in this study, if there is little variation in the MIC values for a specific antimicrobial, it may not load substantially on any factor. If identification of patterns is the study objective rather than detecting emerging resistance, then it may be useful to omit antimicrobials with little variation in the MIC values from the panel.

References

Aarestrup, F.M. 2000. Occurrence, selection and spread of resistance to antimicrobial agents used for growth promotion for food animals in Denmark. APMIS Suppl 101: 1-48.

Blondeau, J. M. and Vaughan, D. 2000. A review of antimicrobial resistance in Canada. Can. J. Microbiol. 46(10): 867-877.

Cavallo, J.D., Fabre, R., Leblanc, F., Nicolas-Chanoine, M.H., and Thabaut, A. 2000. Antibiotic susceptibility and mechanisms of beta-lactam resistance in 1310 strains of pseudomonas aeruginosa: a french multicentre study (1996). J. Antimicrob. Chemother. 46: 133-6.

Conover, W.J. 1980. Practical nonparametric statistics, second edition. John Wiley and Sons, New York.

Dargatz, D.A., Fedorka-Cray, P.J., Ladely, S.R., and Ferris, K.E. 2000. Survey of Salmonella serotypes shed in feces of beef cows and their antimicrobial susceptibility patterns. J. Food Prot. 63: 1648-1653.

Dargatz, D.A., Fedorka-Cray, P.J., Ladely, S.R., Ferris, K.E., Green, A.L., and Headrick, M.L. 2002. Antimicrobial susceptibility patterns of *Salmonella* isolates from cattle in feedlots. J. Am. Vet. Med. Assoc. 221(2):268-272.

Davis, M.A., Hancock, D.D., Besser, T.E., Rice, D.H., Gay, J.M., Gay, C., Gearhart, L., and DiGiacomo, R. 1999. Changes in antimicrobial resistance among Salmonella enterica Serovar typhimurium isolates from humans and cattle in the Northwestern United States, 1982-1997. Emerg. Infect. Dis. 5: 802-6.

Harman, H.H. and Jones, W.H. 1966. Factor analysis by minimizing residuals (MINRES). Psychometrika 31: 351-368.

Johnson, R.A. and Wichern, D.W. 1992. Applied multivariate statistical techniques, third edition. Prentice-Hall, Inc., New Jersey.

Jones, R.N. 2000. Detection of emerging resistance patterns within longitudinal surveillance systems: data sensitivity and microbial susceptibility. MYSTIC Advisory Board. Meropenem Yearly Susceptibility Test Information Collection. J Antimicrob Chemother 46 Suppl T2: 1-8.

Kaspar, C.W., Burgess, J.L., Knight, I.T., and Colwell, R.R. 1990. Antibiotic resistance indexing of Escherichia coli to identify sources of fecal contamination in water. Can. J. Microbiol. 36: 891-4.

Krumperman, P.H. 1983. Multiple antibiotic resistance indexing of Escherichia coli to identify high-risk sources of fecal contamination of foods. Appl. Environ. Microbiol. 46: 165-170.

Morrison, D.F. 1976. Multivariate statistical methods. McGraw-Hill, Inc., New York, New York.

Prescott, J.F., Baggot, J.D. and Walker, R.D. 2000. Antimicrobial therapy in veterinary medicine. Iowa State University Press, Ames, Iowa.

Ruiz, J., Gomez, J., Navia, M.M., Ribera, A., Sierra, J.M., Marco, F., Mensa, J., and Vila, J. 2002. High prevalence of nalidixic acid resistant, ciprofloxacin susceptible phenotype among clinical isolates of *Escherichia coli* and other *Enterobacteriaceae*. Diagn. Microbiol. Infect. Dis. 42: 257-261

Sahm, D.F., Karlowsky, J.A., Kelly, L.J., Critchley, I.A., Jones, M.E., Thornsberry, C., Mauriz, Y., and Kahn, J. 2001. Need for annual surveillance of antimicrobial resistance in Streptococcus pneumoniae in the United States: 2-year longitudinal analysis. Antimicrob. Agents Chemother. 45: 1037-1042.

Smith, K.E., Besser, J.M., Hedberg, C.W., Leano, F.T., Bender, J.B., Wicklund, J.H., Johnson, B.P., Moore, K.A., and Osterholm, M.T. 1999. Quinolone-resistant Campylobacter jejuni infections in Minnesota, 1992- 1998. Investigation Team. N. Engl. J. Med. 340: 1525-32.

Thornsberry, C., Jones, M.E., Hickey, M.L., Mauriz, Y., Kahn, J., and Sahm, D.F. 1999. Resistance surveillance of Streptococcus pneumoniae, Haemophilus influenzae and Moraxella catarrhalis isolated in the United States, 1997-1998. J. Antimicrob. Chemother. 44: 749-59.

USDA. 2000. Part III: Health management and biosecurity in U.S. Feedlots. Part III: Health management and biosecurity in U.S. Feedlots. USDA:APHIS:VS, CEAH, National Animal Health Monitoring System. Fort Collins, CO #N336.1200.

Walker, R.D. and Thornsberry, C. 1998. Decrease in antibiotic susceptibility or increase in resistance? J. Antimicrob. Chemother. 41: 1-4.

Wiggins, B.A. 1996. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl. Environ. Microbiol. 62: 3997-4002.

CHAPTER 6

REVIEW OF HERD-LEVEL TESTING IN VETERINARY MEDICINE

Summary

Production animals often are raised in clusters such as a herd or flock. Determination of the health status of the cluster requires a diagnostic test. The use of imperfect diagnostic tests to determine health status of a cluster or a group of clusters has received much attention in the veterinary literature. Important considerations in evaluating herd-level tests include animal-level test sensitivity and specificity, sample size, cutpoint, within-herd prevalence, and herd size. The notion of herd-level sensitivity and specificity arises from characterizing the effect of the combination of these considerations on the ability to correctly classify the status of a cluster.

A number of models have been constructed to examine the relationship between the animal-level and herd-level testing. The models range from fixed parameter inputs to allowing use of probability distributions to model variability. The distribution of within-herd (cluster) prevalence has been modeled using the intracluster correlation coefficient and the beta binomial probability distribution. Some of the models can be used to explore two-stage sampling designs.

Introduction

The notion of imperfect test sensitivity and specificity of tests as they apply to individuals is well recognized in the medical and veterinary communities. However, in animal production most individuals are part of a herd or a flock. Epidemiologists are often interested in testing outcomes for a herd with the goal of being able to estimate herd-level disease prevalence or, in some instances, herd-level incidence. Herd-level prevalence is an important measure for determining geographic distribution of diseases and has implications for animal health decisions in terms of regionalization, certification, and control programs. Additionally, herd-level disease status determination is critical in examining herd-level disease risk factors. The purpose of this chapter is to review the development and theory of herd-level (aggregate) testing.

Literature review

The theory surrounding the application of diagnostic testing in herds has developed over the last 10 years. Martin et al. (1992) compiled and explored many of the important issues that surround the development of herd-level testing. They initially addressed the impact of imperfect animal-level test sensitivity and specificity. Earlier research had focused primarily on the importance of imperfect animal-level sensitivity (Marchevsky et al. 1989; Adler and Wiggens 1973; MacDiarmid 1988). After reviewing previously published adjustments for sample size if the test is less than 100 percent sensitive, Martin et al. (1992) modified Cannon and Roe's (1982) sample size formula for detecting disease by essentially multiplying the number of diseased animals in a population by the sensitivity.

Use and interpretation of herd-level tests becomes problematic when individual test specificity is less than 100%. Martin et al. (1992) raised the important issue of false positive errors. If an individual-level test lacks perfect specificity, then testing at the herd-level can result in either herd misclassification of a non-diseased herd (false positive) or the correct classification of a diseased herd, albeit for the wrong reason. They define the probability that an individual will give a positive result (become a reactor) which is directly related to the true prevalence within a herd and the sensitivity of a test. This probability is inversely related to test specificity.

Martin et al. (1992) described the implications of test positivity, both true and false positive test results, on herd-level status. As sample size increases herd-level sensitivity increases (i.e. the probability of a test positive increases) but herd specificity decreases because false positives will lead to incorrect classification of disease-free herds. In order to compensate for decreased specificity, they introduced the concept of the cutpoint number of animals used to declare a herd diseased which is based on the probability distribution of the number of reactors in a herd. Their probability distribution for the number test positive is based on test sensitivity, specificity, and number tested. If the number of test positives exceeds a predetermined number of animals (the cutpoint), the herd would be considered diseased. Thus, they are defining herd sensitivity as the proportion of diseased herds in which the number of reactors meets or exceeds a specified cutpoint number of reactors. Herd-level specificity can be increased under this definition but there will be a trade off with herd-level sensitivity. Herd-level sensitivity decreases

with the use of a cutpoint and the implications for herd-level test diagnostics should be considered carefully especially in terms of herd-level predictive values.

Donald (1993) extended the theoretical framework for aggregate testing by incorporating the clustering, or similarity, of disease states within a farm and even within an animal while still accounting for test sensitivity and specificity. He considered both the cluster-level prevalence (the proportion of clusters in a population that have at least one diseased element) and the element-level prevalence (the proportion of elements, in the population, that are diseased). Donald suggested modeling cluster-correlated binary responses with the beta-binomial distribution which can be related directly to the intracluster correlation coefficient, ρ (see Donald et al. 1994). The intracluster correlation coefficient measures disease outcome dependencies within a cluster. If $\rho = 0$ then disease is randomly distributed throughout the population (no disease clustering – disease is randomly distributed). If $\rho = 1$ then all clusters are concordant (all animals within a cluster share the same diseased or non-diseased status.

Donald (1993) constructed a model for the apparent distribution of disease assuming there are two subpopulations of clusters, diseased and non-diseased, and that within the diseased subpopulation the disease is distributed according to the beta-binomial distribution (incorporates the intracluster correlation). The purpose of hypothesizing two populations is to allow for zero prevalence herds which would not occur when using the beta-binomial distribution. The model uses probability distributions similar to those described by Martin et al. (1992) for element-level test positivity for diseased and non-

diseased subpopulations. The model is used to construct a frequency distribution for the number of test-positive results based on the model parameters (cluster size, element-level prevalence, sensitivity, specificity, and the prevalence pattern parameter (percent of herds diseased and non-diseased)). He then used this model to examine three potential cluster-level prevalence estimators, a crude estimator, an empirical estimator, and a corrected estimator (conventional method developed by (Rogan and Gladen 1978)). He examined the properties associated with the estimators (bias and variance). The crude and conventional estimators are biased relative to values of the intracluster correlation coefficient. The empirical estimator appears to overcome the bias. He noted that an advantage of the empirical estimator is that its resistance to bias is not dependent on having a prior estimate of the intracluster correlation.

Although Donald appears to support the use of the empirical estimator, he acknowledged limitations including the estimation of the disease prevalence parameter and no allowance for variable cluster size, cutpoints (other than a single test positive) or subsampling within herds. The empirical estimator has not been widely accepted, perhaps because of these limitations and perhaps because of the complexity of the method (Jordan and McEwen 1998).

Donald et al. (1994) revisited the aggregate testing topic to address some of the limitations from the first study. They incorporated variable cutpoints, correlated test results and "disease-free" status. Test result correlation can occur following vaccination or other seroconversion, including natural exposure. In the case of test correlation, if one

animal is test positive then another animal within the same herd is more likely to test positive, even if both animals are free of disease. They modify the beta-binomial based model constructed by Donald (1993) by including six parameters. The six parameters were disease prevalence, test sensitivity, test specificity, and the intraclass correlations for each of the three parameters. The last two parameters are additions to the previous model. Cutpoints are evaluated by comparing the aggregate sensitivity against 1 minus the aggregate specificity on a receiver operating characteristic curve. The optimal cutpoint is selected as the minimum geometric distance from the perfect aggregate test. A final modification of their model is to allow some tolerance for accepting low probability of disease herds as being disease negative.

The discussion of the implications of the theoretical findings from their model including the effects of sample size, cutpoints, and disease distribution, follows, in many regards, the work of Martin et al. (1992) and Donald (1993). However, they acknowledge that the effect of within-herd test correlations is complex but, in general, increased within-herd sensitivity (or specificity) correlation will decrease aggregate sensitivity (or specificity). Sensitivity correlation will not affect aggregate specificity but the converse is not true. In concluding remarks, Donald et al. (1994) note that much of what they have developed is theoretical and should be used primarily to help clarify the issues surrounding aggregate testing. They suggest investigators use these ideas to help focus on correct sample size and cutpoints for aggregate testing. Investigators should be aware that there is likely to be variation in herd-level sensitivity and specificity. In the absence of adequate information on animal- and herd-level disease prevalence and the distribution of disease

among herds they recommend computing aggregate sensitivities, specificities and predictive values over a range of parameters and select the most conservative estimate. They provide a FORTRAN program to accomplish the task. Other researchers have found the complexity of this presentation to be a practical drawback to implementation (Jordan and McEwen 1998).

Carpenter and Gardner (1996) used a simulation model of serologic testing for porcine parvovirus to demonstrate the inherent problems in extrapolating animal test results to the herd level. Their outcome of interest was a three-level categorization of seroprevalence (low, medium, and high). Herd-level predictive values were their measure of overall test efficacy. They compared their results to those of Martin et al. (1992) and found some similarities and some differences. Increasing sample size under imperfect sensitivity increased herd-level sensitivity and predictive value negative in Martin's development of the problem. With three prevalence categories, Carpenter and Gardner (1996) noted that under similar conditions their model showed increased negative predictive value and herd-level sensitivity with the exception of a decrease in herd-level sensitivity in the high prevalence herds. Martin et al. (1992) also showed that for fixed sample size and specificity, increasing test specificity decreased herd sensitivity. Carpenter and Gardner (1996) observed a different result in their medium seroprevalence level but had results in the high prevalence group that agreed with Martin et al. (1992). The strength of the approach by Carpenter and Gardner was to demonstrate that, in addition to the analytic approaches used by previous researchers, there is value in taking a simulation model approach.

Christensen and Gardner (2000) reviewed the theoretical aspects of herd-level interpretation of individual animal tests and presented some new considerations for herd-level sensitivity and specificity estimation including potential problems with biased estimates of animal-level test characteristics. Precision associated with individual-animal sensitivity and specificity will be reflected in potentially severely reduced precision in herd-level tests. They showed that test-evaluation studies with small sample sizes resulted in wide confidence interval estimates for animal-level test specificity and sensitivity, which translated to wide confidence intervals for herd level sensitivity and specificity (as wide as 0 to 1). Similarly, bias in individual-level sensitivity and specificity can result in biased herd-level test characteristics. They suggested that direct estimation of herd-level test characteristics from field studies may be more appropriate in some situations than a theoretical approach. They also discussed use of testing pooled samples, targeted (prevalence-directed) sampling and use of a combination of tests. The practicality of pooled tests depends on many factors but is likely to be most effective when true prevalence is low. Targeted sampling can increase herd sensitivity essentially by increasing the within-herd prevalence of the herds being tested. Multiple tests can either increase herd-level sensitivity or specificity depending on the test interpretation (serial or parallel)

In the later half of the 1990's, interest in aggregate testing shifted to applying the theory to substantiating freedom from disease and complete survey design and building applications (Cameron and Baldock 1998b; Cameron and Baldock 1998a; Jordan and

McEwen 1998; Audige and Beckett 1999; Jordan and McEwen 1998; Audige and Beckett 1999). Cameron and Baldock (1998a) developed probability estimates for perfect and imperfect tests under finite and infinite populations. Cameron and Baldock (1998a) then investigated hypothesis testing for disease freedom. Their specification of the null hypothesis is that disease is present at a level greater than or equal to a prevalence of interest. The alternative hypothesis is that the prevalence is below the prevalence of interest. They then construct probability distributions based on the null and alternative hypothesis with cutpoints. At small sample sizes these distributions can overlap (given a realistic hypothesis for freedom of disease). Increasing the sample size will lead to separation of these distributions. Sample size can be set to the level at which both power and the type I error rate are satisfactory. A program was provided by the authors to accomplish this computation. The probability estimates under finite sampling conditions is a basic extension of Martin's work except for a shifting of the adjustment for sensitivity into the probability of success for the binomial trial. Martin et al. (1992) alter the exponent of the binomial distribution by multiplying it by the number of diseased animals. The impact of this shift has not been fully explored. They conclude that their probabilistic approach is appropriate to use in two-stage sampling which they discuss in a companion paper (Cameron and Baldock, 1998b).

In their companion paper, Cameron and Baldock (1998b) extended their approach from within-herd to herd-level survey development. They take a two-stage approach for which the first stage consists of herds and the second stage consists of animals within herds. Within the second stage they use their approach developed in the first paper to define

within-herd type I error and power levels based on sensitivity, specificity, minimum prevalence (hypothesis based) and cutpoint criteria. The herd-level sensitivity is then defined as (1 - type I error) and herd-level specificity is (1-type II error). The imperfect herd-level test is treated similarly to the imperfect within-herd test and cutpoints are utilized.

Cameron and Baldock (1998a; 1998b) do not directly consider some of the parameters that others have considered important for aggregate testing such as intracluster correlations and variable test characteristics. Their two-step approach is intuitively appealing and appears to be applicable to real world situations, but it does not account for variability in sensitivity and specificity or for clustering of test results and disease. However, both papers by Cameron and Baldock (1998a; 1998b) demonstrate the point made by Carpenter and Gardner (1996) that non-analytical but computationally intense solutions are practical.

A generalized model for aggregate testing was developed by Jordan and McEwen (1998). The generalizations were incorporated into a Monte Carlo model which allows for input of stochastic distributions for animal-level sensitivity/specificity, animal- and herd-level disease, and herd size. Within-herd testing protocols (all animals/fixed number/fixed proportion/or some combination) were also flexible. Two modes were offered, free-living-contagion and obligate-parasite mode. They also introduced the use of likelihood ratios to assess testing protocols. The likelihood ratio for a positive herd test is the probability of a positive herd-test result in a diseased herd divided by the probability of a

positive herd-test in a non-diseased herd. Similarly, the likelihood ratio for a negative herd test is the probability of a negative herd-test result in a non-diseased herd divided by the probability of a negative herd-test in a diseased herd. These ratios, which are herd-level characteristics, are summarized by their computer program for the population being sampled. An example using *E. coli* O157:H7 was used.

Jordan and McEwen (1998) discuss the data limitations for justifying specific parametric distributions for their input variables and note the potential need for expending resources to accurately describe input variables. The use of continuous probability distributions to model within-herd prevalence levels, although not specifically addressed by Jordan and McEwen (1998), is problematic. Given a finite and potentially small population and a continuous probability distribution, it is possible that a simulated infected herd (non-zero but low disease prevalence) will, by chance, have no infected animals. This issue was addressed by Donald (1993) by creating a non-infected subpopulation. Jordan and McEwen (1998) offer the alternative of either accepting the herds as non-infected or forcing all infected herds to have at least one infected animal.

Independently, but in the same time frame as Jordan and McEwen, Audigé and Beckett (1999) developed a model which focused on assessing health surveys versus constructing herd-level testing protocols. Briefly, their model had two stages. In stage one, the input parameters (animal-level test characteristics, within-cluster prevalence, cluster size and within-cluster sample size) were used to model the number of positive test results in both infected and non-infected herds. Beta distributions were used to model uncertainty in

animal-level test characteristics. Prevalence was modeled using a BetaPERT model (a modification of the Beta distribution), but an assumption was made of no relationship between cluster size and within-cluster prevalence. The number of animals tested per cluster was fixed. A cutpoint was determined and herd-level test characteristics were determined using a receiver operating characteristic curve. In the second stage, these herd-level test characteristics were applied to an infected population and an uninfected population of clusters. The number of positive clusters expected from the uninfected and infected populations was simulated. Audigé and Beckett (1999) used likelihood ratios as did Jordan and McEwen (1999) but in a different sense. They calculated the ratio, for each number of possible positive clusters, as the probability of observing the number if the country, or population of herds, is infected divided by the probability of observing the number if the country is uninfected.

In 2000, Audigé et al. (2001) extended their model for use in substantiating freedom from disease. One of the major changes was to allow for a Bayesian approach to determining post-survey probability of disease freedom. Assessing the pre-survey (prior) probability was not within the scope of their paper, but the authors listed several criteria including history of disease, animal movement and veterinary infrastructure. They modified their model to allow for varying sample sizes per herd and thus were able to compare the efficiency of whole-herd ELISA testing for IBR as compared to sampling five animals per herd.

Cannon (2001) continued the development of aggregated testing applications by providing some mathematical tools and by exploring disease freedom with and without perfect specificity. He demonstrated that the inverse of the beta function (available in some software packages) can be used to calculate confidence intervals for proportions. Other tools include a "fractional" binomial approximation for the mixture of hypergeometric and binomial distributions that arise when sampling with a non-perfect test. He used the approximation to model the number of reactors in a sample from a herd. He also used the approximation along with the inverse beta function to calculate a confidence interval for the proportion of disease animals in a population. Because calculation of hypergeometric probabilities can overwhelm computer resources, he suggested the use of a ratio of log gamma functions to calculate probabilities.

Cannon (2001) compared four methods of calculating samples sizes when the objective is to demonstrate freedom from disease using a perfectly specific test. One of the methods he used was equivalent to an approach used by Garner et al. (1997) in a study of porcine reproductive and respiratory syndrome in Australia. He also considered two approaches to estimating the sample size for herds based on an average herd size and average number of animals sampled per herd. His discussion of determining freedom under the conditions of imperfect specificity paralleled the approach of Cameron and Baldock (1998b). The tools that Cannon has developed were implemented in an Excel spreadsheet.

The issues regarding aggregate testing are complex and the appropriate methods are still being debated in the literature. Paisley et al. (2000) used a simulation model to demonstrate that a national surveillance program for bovine paratuberculosis was impractical under the conditions of low herd-level prevalence (0.2%), poor individual animal test sensitivity (mean 45%), low within herd prevalence and small herd sizes. Stärk et al. (2000) compared two sampling designs, individual random animal and two-stage sampling. They concluded that the latter approach requires a larger sample size but that the increase in information about herd status makes the effort worthwhile. Ziller et al. (2002) compared three within-herd sampling strategies: cluster (all animals sampled), subsampling within the herd based on herd characteristics, and a limited fixed sample size for each herd. They conclude that cluster sampling provided the highest certainty of freedom but at the highest cost. The subsampling approach was more efficient than cluster sampling, and the limited sampling was least expensive and requires more herds to be sampled. They recommend the limited sampling approach for large survey efforts.

Conclusions

In summary, the theoretical framework for investigating the factors associated with aggregate testing was developed in the early 1990's. The impacts of test sensitivity/specificity, disease clustering and test correlation, cutpoints, and prevalence have all been examined. Researchers in the second half of the 1990's emphasized application of the theory to real world examples but found that in many situations data on which to base assumptions were not available. Optimization in the face of costs

associated with testing has been discussed by Cameron and Baldock (1998b) and Ziller et al. (2002) but has not been adequately developed.

References

Adler, H.A. and Wiggens, A.D. 1973. Interpretation of serologic tests for *Mycoplasma gallisepticum*. World Poultry Sci. J. 9: 345-353.

Audige, L. and Beckett, S. 1999. A quantitative assessment of the validity of animal-health surveys using stochastic modelling. Prev. Vet. Med. 38: 259-76.

Audige, L., Doherr, M.G., Hauser, R., and Salman, M.D. 2001. Stochastic modelling as a tool for planning animal-health surveys and interpreting screening-test results. Prev. Vet. Med. 49: 1-17.

Cameron, A.R. and Baldock, F.C. 1998a. A new probability formula for surveys to substantiate freedom from disease. Prev. Vet. Med. 34: 1-17.

Cameron, A.R. and Baldock, F.C. 1998b. Two-stage sampling in surveys to substantiate freedom from disease. Prev. Vet. Med. 34: 19-30.

Cannon, R.M. 2001. Sense and sensitivity - designing surveys based on an imperfect test. Prev. Vet. Med. 49: 141-163.

Cannon, R. M. and Roe, R. T. Livestock disease surveys: A field manual for veterinarians. 1982. Australian Bureau of Animal Health Canberra.

Carpenter, T.E. and Gardner, I.A. 1996. Simulation modeling to determine herd-level predictive values and sensitivity based on individual-animal test sensitivity and specificity and sample size. Prev. Vet. Med. 27: 57-66.

Christensen, J. and Gardner, I.A. 2000. Herd-level interpretation of test results for epidemiologic studies of animal diseases. Prev. Vet. Med. 45: 83-106.

Donald, A. 1993. Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm. Prev. Vet. Med 15: 125-145.

Donald, A.W., Gardner, I.A., and Wiggins, A.D. 1994. Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. Prev. Vet. Med. 19: 167-187.

Garner, M.G., Gleeson, L.J., Holyoake, P.K., Cannon, R.M., and Doughty, W.J. 1997. A national serological survey to verify Australia's freedom from porcine reproductive and respiratory syndrome. Aust. Vet. J 75: 596-600.

Jordan, D. and McEwen, S.A. 1998. Herd-level test performance based on uncertain estimates of individual test performance, individual true prevalence and herd true prevalence. Prev. Vet. Med. 36: 187-209.

MacDiarmid, S.C. 1988. Future options for brucellosis surveillance in New Zealand beef herds. New Zealand Vet. J. 36: 39-42.

Marchevsky, N., Held J.R., and Garcia-Carrillo C. 1989. Probability of introducing diseases because of false negative test results. Am. J. Epidemiol 130: 611-614.

Martin, S.W., Shoukri, M., and Thorburn, M.A. 1992. Evaluating the health status of herds based on tests applied to individuals. Prev. Vet. Med. 14: 33--43.

Paisley, L.G., Tharaldsen, J., and Jarp, J. 2000. A simulated surveillance program for bovine paratuberculosis in dairy herds in Norway. Prev Vet Med 44: 141-51.

Rogan, W.J. and Gladen, B. 1978. Estimating prevalence from the results of a screening test. Am. J. Epidemiol. 107: 71-76.

Stark, K.D., Mortensen, S., Olsen, A.M., Barfod, K., Botner, A., Lavritsen, D.T., and Strandbygard, B. 2000. Designing serological surveillance programmes to document freedom from disease with special reference to exotic viral diseases of pigs in Denmark. Rev Sci Tech 19: 715-24.

Ziller, M., Selhorst, T., Teuffert, J., Kramer, M., and Schluter, H. 2002. Analysis of sampling strategies to substantiate freedom from disease in large areas. Prev Vet Med 52: 333-43.

CHAPTER 7

ESTIMATION OF HERD-LEVEL PREVALENCE: STRATEGIES FOR SAMPLING DESIGN

Summary

Two-stage sampling has been used to study disease freedom in agricultural animal production. The first stage of sampling is the herd or flock and the second stage is sampling within the herd. At both stages sampling is directed at the detection of disease. Herd-level test characteristics are based on individual animal diagnostic test characteristics, within-herd prevalence, and cutpoint number of positive tests for declaring a herd positive.

Studies of the distribution of disease in a population may be concerned with estimating herd-level prevalence rather the disease detection. We propose the use of a herd-level sample size formula based on a common adjustment for prevalence estimates when diagnostic tests are imperfect. The formula depends on an estimate of herd-level sensitivity and specificity. A Monte Carlo simulation model was developed to calculate these herd-level test characteristics when disease may be clustered in herds.

Simulations explore the effect of varying levels of intracluster correlation on herd-level sensitivity and specificity. At low prevalence, herd-level sensitivity increased with

increasing intracluster correlation, but the sensitivity was less affected at higher prevalence. Also at low prevalence, many herds are being classified as positive based only on false positive test results. Positive predictive values drop sharply with increasing intracluster correlation.

A hypothetical example of two-stage sampling was created to examine the impact on herd and within-herd sample sizes of a single test positive versus a variable number of positives for declaring a herd positive. More herds but fewer animals within herd were needed for the single test positive design versus the variable cutpoint design. Real life two-stage sampling design examples were developed for estimating Ovine Progressive Pneumonia in sheep and *Salmonella* spp. in finishing swine herds

The two stage sampling approach for estimating herd-level prevalence is a flexible methodology that allows researchers to address sampling objectives. The use of a distribution for within-herd prevalence results in a conservative estimate of herd-level test characteristics. The model allows for researchers to achieve a balance between the number of herds and the total number of animals sampled by manipulating herd-level test characteristics such as the number of animals sampled within a herd. The model can be used to adjust designs to minimize the potential for differential misclassification of herd status.

1. Introduction

Surveys of disease prevalence are often intended to provide information to substantiate claims of disease freedom in a regional or national level. A two-stage sampling approach has been proposed as a basis for such a purpose (Garner et al. 1997; Cameron and Baldock 1998b; Audige and Beckett 1999; Stark et al. 2000). The first stage of sampling is the herd, and the second stage is within-herd sampling.

The use of two-stage sampling has evolved to meet surveillance objectives for two reasons. First, list frames of animals for randomized sample selection do not typically exist at a regional or national level while list frames of herds can be more readily constructed and maintained. Secondly, the theory and application of within-herd sampling with imperfect diagnostic tests has been well developed by a number of researchers (Martin et al. 1992; Donald 1993; Donald et al. 1994; Carpenter and Gardner 1996; Jordan and McEwen 1998; Cameron and Baldock 1998a). The within-herd sampling research has guided the approach to sampling to classify the herd's disease or infection status.

The herd-level sensitivity (HSE) and specificity (HSP) depend on a number of criteria including individual animal test characteristics, sample size, within-herd prevalence and the cutpoint number of reactors (Martin et al. 1992; Cameron and Baldock 1998a). The HSE and HSP become test characteristics which can be applied at the herd-level in a

manner equivalent to animal-level sensitivity (SE) and specificity (SP) at the within-herd level (Cameron and Baldock 1998b). The HSE and HSP are usually based on detecting disease if it occurs above a fixed level which is determined according to the epidemiology of the disease under study or specific national or international rules. For surveys intended to investigate disease freedom, the sampling of herds at the first stage is usually designed to detect disease if it is present. If the disease is present it is assumed to occur above a minimum level in infected herds.

However the determination of HSE and HSP becomes problematic if a minimum within-herd prevalence level is not assumed. The distribution of within-herd prevalence becomes an important consideration. Donald (1993) suggested modeling the distribution of within-herd prevalence as cluster-correlated binary responses using the beta-binomial to account for the non-independence of disease status within herds. He related the intracluster correlation coefficient to the beta-binomial distribution.

The National Animal Health Monitoring System (NAHMS) conducts national surveys of animal health and production (Wineland and Dargatz 1998). One of the goals of NAHMS surveys is to estimate herd-level prevalence for various diseases of interest in cattle, sheep, swine, equine, and poultry. The NAHMS studies have been implemented as two-stage sampling designs but have not typically considered the affect of imperfect tests or disease clustering on HSE and HSP. Unfortunately, two-stage sampling methods for substantiating disease freedom are not directly applicable to surveys for estimating herd-level prevalence. Because substantiating disease freedom relies on detection of

disease at both the within-herd and herd levels, an estimate of herd-level prevalence may not have the desired precision. Also, within-herd sampling for disease freedom often assumes a fixed prevalence to be detected. When conducting a national survey to estimate herd-level prevalence, such as in a NAHMS study, it may be more appropriate to assume a distribution for within-herd prevalence of disease, especially as affected by clustering (Donald 1993; Donald et al. 1994; Jordan and McEwen 1998).

The objective of this study is to develop two-stage sampling strategies for estimating herd-level prevalence given imperfect diagnostic tests, varying herd size, and clustering of disease. To meet this objective a sample size formula for first stage sampling is proposed and a Monte Carlo simulation model is used to investigate the effect of disease clustering and various sampling strategies. Application of the approach is demonstrated with two examples.

2. Material and Methods

2.1 Sample size formula for herd-level estimation

Rogan and Gladen (1978) proposed a prevalence estimator that was adjusted for imperfect test characteristics. Although their estimator was intended for individual test results, the estimator can readily be applied to herd-level estimation (Donald 1993). The herd-level estimate of disease is:

$$\hat{\theta}_{RG} = \frac{\hat{\theta} + HSP - 1}{HSE + HSP - 1}$$

where $\hat{\theta}_{RG}$ denotes the adjusted herd-level prevalence and $\hat{\theta}$ denotes the apparent or unadjusted herd-level prevalence.

Donald (1993) showed the variance of the estimator, assuming fixed HSE and HSP is:

$$Var(\hat{\theta}_{RG}) = \frac{Var(\hat{\theta})}{(HSE + HSP - 1)^2}$$

The variance formula can be used to derive a sample size equation (Appendix 1). A sample size estimate for the number of herds (n) needed to achieve an acceptable error bound (Error bound denotes one-half the maximum width for the confidence interval for θ_{RG}) with a specified confidence level (1-α) is:

$$n = \frac{\hat{p}(1-\hat{p})}{(HSE + HSP - 1)^2}\left(\frac{Z_{\alpha/2}}{Error}\right)^2$$

2.2 Simulation model characteristics

A model was constructed to simulate two-stage sampling using a commercially available software package (@risk, Palisade Corporation, Newfield, NY). Inputs into the model include herd size, animal-level sensitivity (SE) and specificity (SP), animal-level prevalence, within-herd sample size, percent of herds that are negative, and the cutpoint needed to designate a herd as positive. Herd size can be input as a fixed size or it can

follow a discrete distribution. Sensitivity and specificity can be input as point estimates or beta pert distributions (minimum, maximum and most likely) (Audige and Beckett, 1999). The sample size can be a fixed number, a fixed percent of the herd size, or based on the discrete distribution of herd size. The cutpoint is the number of test positive results that are needed to designate a herd as infected. The default cutpoint is one. These input parameters are similar to those used by Audige and Beckett (1999) and Jordan and McEwen (1998). When fixed parameters are input, results from the model follow those reported by Cameron and Baldock (1998b).

In each simulation iteration, the input parameters are used to create infected and non-infected herds. A beta-binomial distribution was used to determine the number of infected animals in each herd. The parameters for the beta-binomial distribution (α and β) were determined by the specified animal-level prevalence (p) and the intracluster correlation coefficient (ρ) using the following formulas (Bohning and Greiner 1998):

$$\alpha = \frac{p}{\rho} - p$$

and

$$\beta = \frac{(1-p)}{\rho} + p - 1$$

Infected herds (regardless of the herd size) could be forced to have at least a single infected animal, as was used by Donald et al. (1994), but this option was not used in this study. After the number of infected animals in a herd was determined the model used a hypergeometric distribution function to randomly choose the number of infected animals in the sample from the herd. The number of non-infected animals in the sample is the

difference between the sample size and the number of infected animals in the sample as determined by the hypergeometric distribution function. A binomial probability distribution was used to estimate the number of infected animals in cases when the number of infected animals was large and the hypergeometric function failed.

A binomial distribution function was used to simulate the diagnostic testing of the sample of animals from the herd. Sensitivity was input into the binomial probability distribution function along with the number of infected animals to determine the number of true test positives and false test negatives in each sample. Similarly, SP was used to determine the number of true test negatives and false test positives.

The simulated test results for each herd were cross-classified into the traditional 2x2 table based the true herd status and the results from the simulated testing. This classification allowed for calculation of herd-level sensitivity, specificity, and positive and negative predictive values. Each simulation was executed 10,000 times.

2.3 Survey model implementation

The model was used to investigate both theoretical and applied within-herd and herd-level sampling design. First, the effect of within-herd sample size, intracluster correlation coefficient and cutpoint on HSE, HSP, positive predictive value, and negative predictive value were assessed by simulation. For these simulations, herd-size (N) was 200, SE=90%, SP=90% and proportion of the herds that was negative (ψ=60%) was fixed. Six levels of intracluster correlation (0, 0.05, 0.1, 0.3, 0.5, and 0.8), two animal-

level prevalence levels (1% and 20%), three within-herd sample sizes (10, 20, and 30) and three cutpoints (1, 2, and 3) were used in the simulation. The impact of imperfect specificity was assessed by allowing SP to rise to 100%.

Two-stage sampling was then simulated using the model where SN=98%, SP=99%, and prevalence=20%. The intracluster correlation coefficient (ρ=0.1) and the proportion of herds that were negative (ψ=60%) were fixed. In the first simulated design scenario within-herd sample sizes were determined separately for each herd size category with the constraint that the cutpoint be fixed at one. The within-herd sample size was set to keep the HSE above 90% and the HSP above 80%. In the second simulated design scenario the within-herd sample sizes were simulated again but the cutpoint was allowed to vary. Under this design scenario, the sample size was determined to have approximately HSE= 95% and HSP=90%. Positive and negative predictive values were calculated for both design scenarios. Each of the within-herd design scenarios was then modeled simultaneously for all herd size categories to simulate two-stage sampling from a population of herds. The proportion of herds taken from each herd-size category was simulated in two ways for each of the within-herd sampling design scenarios. In the first herd-level sampling simulation, the proportion of herds from each herd size category followed the proportion of herds in the population. In the second herd-level simulation, the proportion of herds sampled from each herd size category was equivalent (25% from each category). Survey-level test characteristics were calculated for each of the within-herd sampling design scenarios under both of the herd-level sampling strategies.

The survey-level test characteristics from the four combinations of within-herd and herd-level design scenarios were used to determine the number of herds needed to estimate the herd-level prevalence with 95% confidence and an error bound of 10%. The confidence intervals were evaluated by simulating sampling the required number of herds under each of the four design scenarios 1,000 times to see what proportion of the adjusted herd-level prevalence estimates fell within the confidence interval. Thus, if the number of herds required to be 95% confident with an error bound of 10% was 150 then 1,000 simulations of 150 herds was implemented. The simulation results were exported to a statistical software package (SAS, SAS Institute, Cary, N.C.) from which adjusted prevalence for groups of herds was calculated and descriptive statistics were computed.

As the last part of the two-stage simulation, the total number of animals needed to be sampled in each design scenario was calculated by multiplying the number of herds needed in each herd size category by the within-herd sampling requirements for that size category. The number of herds sampled in each herd size category depended on whether the herd-level sampling was proportional to the population or set to equivalent proportions.

2.4 Survey model application

The model was then used to design potential surveys to estimate herd-level prevalence for Ovine Progressive Pneumonia (OPP) in sheep flocks and *Salmonella* spp. in finishing swine herds. These two examples were chosen because they have been the recent focus

of NAHMS studies, and data were available for estimating the input parameters of the model.

In the 2001 NAHMS sheep study a total of 21,525 sheep in 687 flocks were tested for OPP using a competitive-inhibition enzyme-linked immunosorbent assay (cELISA). The SN and SP of the cELISA were estimated to be 98.6% and 96.9%, respectively (K. Marshall, pers. comm.). The proportion of herds in each herd size category followed the herd allocation used by NAHMS. Only flocks with 20 or more animals were eligible. Weighted animal-level prevalence was estimated from the data and adjusted using the Rogan and Gladen formula to obtain an animal-level prevalence estimate for the model. The intracluster correlation coefficient was estimated using hierarchical modeling software (MlwiN, version 1.1, Institute of Education, University of London, London, UK)

For comparison, two sampling designs were constructed for OPP in sheep. The first design restricted the cutpoint to a single positive for determining herd-level status. The sample size was chosen to keep HSP about 70%. The second design allowed flexible cutpoints for each herd-size category and attempted to keep HSE sensitivity about 80%.

The 2000 NAHMS swine study tested a total of 5,470 fecal samples for *Salmonella* spp. using a culture test with SE assumed to be about 75%. The SP for the purpose of the sampling design was assumed to be 100%. The fecal samples were collected from 861 pens in 194 buildings on 123 farms. Hierarchical modeling was used to estimate the

proportion of variance in the presence of *Salmonella* spp. for the complete hierarchy (farm, building, pen, and sample) and for the farm and sample only. The latter model was used to estimate intracluster correlation for the simulation model. Unweighted animal-level prevalence was calculated and adjusted using the Rogan and Gladen formula. The distribution of finishing herd sizes was obtained from the NAHMS data. Only finishing herds with 100 or more finishing pigs were eligible for the NAHMS study.

As with the sheep OPP survey, two sampling designs for sampling for *Salmonella* spp. in finishing swine were developed using the model. Since the specificity of the culture test was assumed to be 100%, the two sampling designs compared were both based on a cutpoint of one for declaring a herd positive. The within-herd sample size for the first design was selected to achieve a HSE of 70% while the second design had a goal of a HSE of 80%.

The percent of herds that were negative was not estimated as an input parameter into either the sheep or the swine herd simulation. Instead, only ρ, calculated from the raw data, was used to determine the distribution of within-herd prevalence. All sampling designs were constructed to achieve an estimate of the herd-level prevalence with 95% confidence and an error bound of 10%.

3. Results

3.1 Herd-level modeling results

Herd-level sensitivity and specificity levels changed in relation to within-herd sample size, cutpoint, animal-level prevalence and ρ (Tables 7.1 and 7.2). At all sample sizes and values of ρ, when the animal-level test was imperfect, the effect of increasing cutpoint was to decrease HSE and increase HSP. Increasing sample size, regardless of cutpoint, within-herd prevalence and ρ, increased HSE and decreased HSP. When the within-herd prevalence was relatively low (π=1%), HSE increased with increasing ρ. When the within-herd prevalence was higher (π=20%) the HSE varied less but appeared to decrease in the middle of the range of ρ. Assuming a perfect animal-level test specificity substantially decreased HSE at both within-herd prevalences. Additionally, trends across the levels of ρ were consistent with the results observed for imperfect specificity at the two animal-level prevalences examined.

Positive and negative predictive values were influenced by ρ, sample size, and cutpoint. Negative predictive values decreased with increasing sample size and, like HSE, appeared to reach a minimum at intermediate values of ρ (Figure 7.1). Negative predictive values remained above 0.8 for all the simulated conditions. Conversely, positive predictive value increased with sample size but remained below about 0.5 and decreased monotonically with increasing ρ.

Table 7.1. Effect of within-herd disease correlation (ρ) with fixed herd size (N=200) and animal-level prevalence (π=.01), and variable animal-level specificity, sample size (n) and cut-off points (k) on herd-level sensitivity and specificity[1]

		Herd-level sensitivity						Herd-level specificity
n	k	HSE (ρ=0)	HSE (ρ=0.05)	HSE (ρ=0.1)	HSE (ρ=0.3)	HSE (ρ=0.5)	HSE (ρ=0.8)	
Animal-level specificity= 0.9								
10	1	68.9	72.3	73.0	81.2	85.7	92.8	35.0
	2	29.8	34.7	38.1	47.6	57.1	69.7	73.7
	3	9.3	11.4	13.4	24.9	36.2	64.2	93.0
20	1	89.9	91.8	92.1	93.3	97.3	96.4	12.2
	2	66.1	70.6	72.3	81.2	85.5	85.8	39.4
	3	37.0	43.3	48.2	56.0	63.8	75.0	67.6
30	1	96.2	98.1	99.0	97.7	98.6	100.0	4.2
	2	85.1	88.8	91.0	89.5	92.8	94.3	18.4
	3	65.1	69.8	71.9	77.0	82.6	90.2	41.4
Animal-level specificity = 1.0								
10	1	10.3	19.5	26.5	34.1	46.9	65.3	100
20	1	19.7	31.2	40.1	45.7	59.9	76.9	100
30	1	27.0	39.8	51.4	55.6	64.7	78.9	100

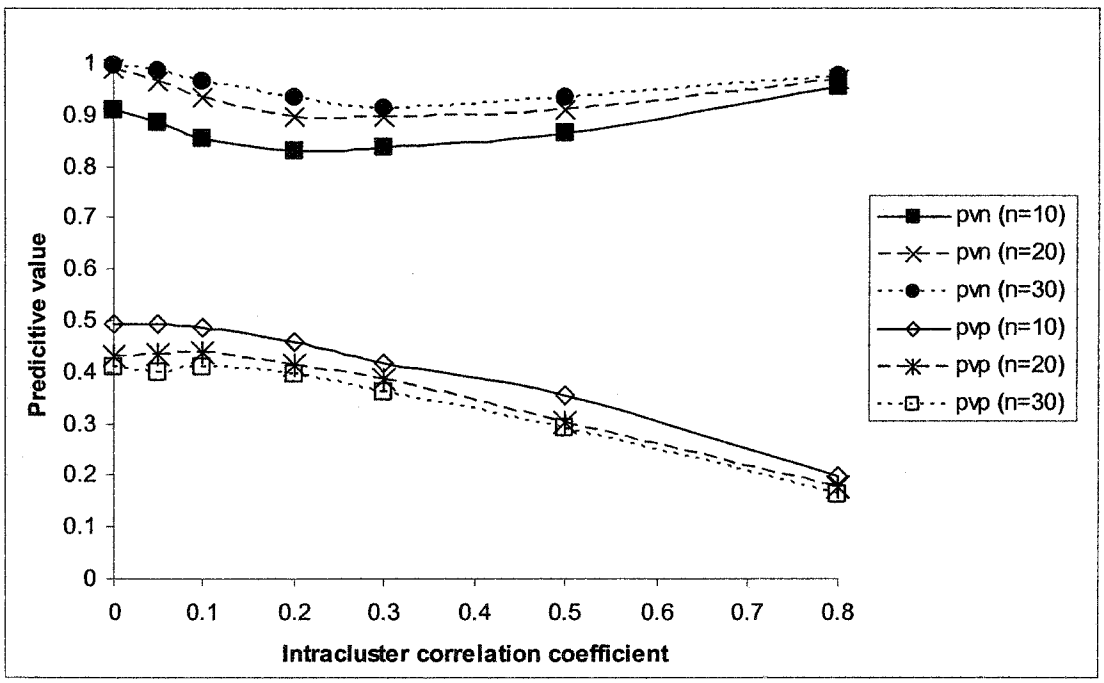
[1]Animal-level sensitivity=0.9 and the proportion of negative herds = 0.6

Table 7.2. Effect of within-herd disease correlation (ρ) with fixed herd size (N=200) and animal-level prevalence (π=.2), and variable animal-level specificity, sample size (n) and cut-off points (k) on herd-level sensitivity and specificity[1]

		Herd-level sensitivity						Herd-level specificity
n	k	HSE (ρ=0)	HSE (ρ=0.05)	HSE (ρ=0.1)	HSE (ρ=0.3)	HSE (ρ=0.5)	HSE (ρ=0.8)	
Animal-level specificity= 0.9								
10	1	94.6	93.2	91.0	87.3	86.6	90.7	34.8
	2	77.6	74.4	71.5	66.8	68.9	77.0	73.8
	3	51.0	49.6	47.4	46.3	48.8	67.6	92.8
20	1	99.8	99.4	98.6	97.7	97.0	97.8	12.3
	2	98.2	96.0	93.7	88.1	87.4	91.5	39.2
	3	91.8	87.7	83.1	77.7	74.8	80.9	67.8
30	1	100	99.9	99.8	99.3	99.4	99.4	4.3
	2	99.9	99.2	98.3	95.8	95.4	96.1	18.4
	3	99.0	97.7	94.3	89.0	88.6	90.7	40.7
Animal-level specificity = 1.0								
10	1	85.5	79.8	75.7	65.7	64.7	75.7	100
20	1	98.0	93.9	88.8	75.8	74.9	79.2	100
30	1	99.8	97.5	94.1	82.6	78.5	83.7	100

[1]Animal-level sensitivity=0.9 and the proportion of negative herds = 0.6

Figure 7.1. The relationship between intracluster correlation coefficient and the predictive value positive (pvp) and predictive value negative (pvn) of a herd-level level test at different within-herd sample sizes.



Negative predictive value decreased with increasing cutpoint while the opposite was true for positive predictive value (Figure 7.2). However, the impact of the change on negative predictive value was about 7% while the cutpoint value changed the positive predictive value by more than 20%. At a ρ of 0.8, the negative predictive values for the three cutpoint levels were similar, but the positive predictive values did not converge in a similar manner.

3.2 Modeling two-stage sampling with a hypothetical population

The hypothetical population of herds was constructed with four herd size categories (10-30, 31-50, 51-100, and 101-200) (Table 7.3). The proportions of herds in the population in each herd size category ranged from 60 percent in the small herd category to 5 percent in the large herd category.

In the sample design with a fixed cutpoint of one test positive, the within-herd sample size ranged from 12 to 20 animals per herd. With these sample sizes, herd-level sensitivity and specificity exceed 91% and 81%, respectively, in all herd size categories. Positive predictive values were between 76% and 81% while negative predictive values remained at 94% or higher.

When the sample design allowed a cutpoint of one or more, the sample size increased in all herd size categories except the small herd size where allowing a higher cutpoint made it impossible to attain an adequate HSE level. In herds with more than 30 animals, all animals were tested up to 55 then in herds with over 55 animals only 55 animals needed to be tested. Both positive and negative predictive values increased in the three herd size

Figure 7.2. The relationship between intracluster correlation coefficient and the predictive value positive (pvp) and predictive value negative (pvn) of a herd-level level test at different cutpoints for determining herd-level status (n=20).

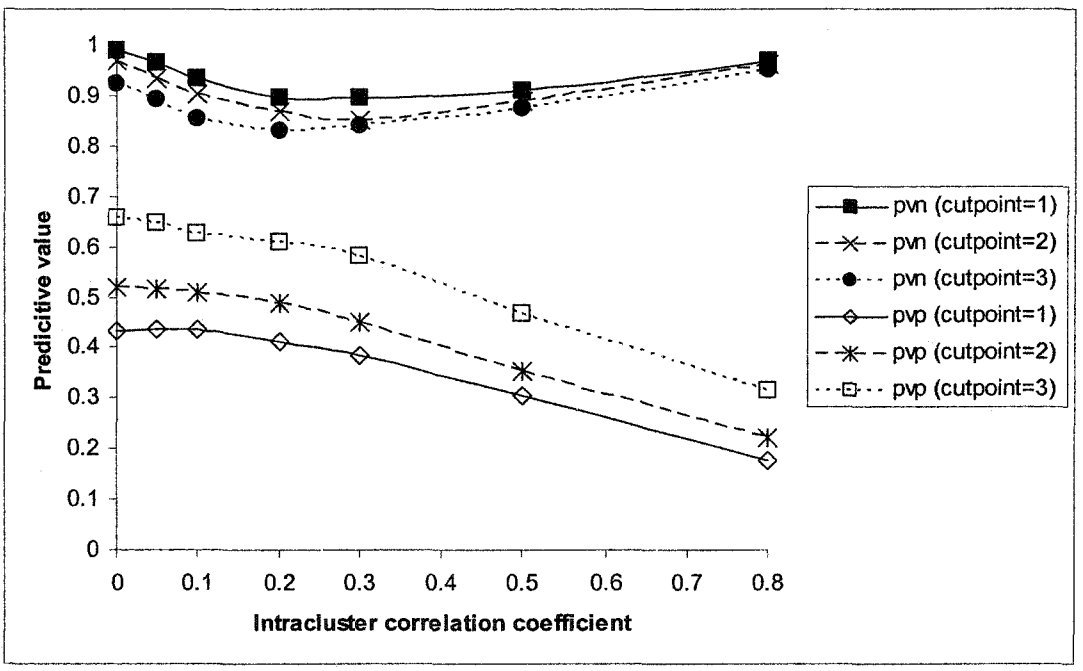


Table 7.3. Herd-level sensitivity, specificity, negative predictive value and negative predictive value for herd size categories under two herd-level design strategies, with and without a fixed cutpoint of one. (SE=98%, SP=99%, π=20%, Ψ=60%, ρ=.1)

Herd size	Percent of herds	Cutpoint	Within-herd sampling	HSE	HSP	Positive Predictive Value	Negative Predictive Value
Fixed cutpoint							
10-30	60	1	All up to 12 then 13	93.3	87.7	80.5	96.1
31-50	20	1	15 up to 40 then 18	92.3	85.1	79.6	94.6
51-100	15	1	19	91.4	82.1	76.5	93.8
101-200	5	1	20	92.7	81.1	76.4	94.4
Variable cutpoint[1]							
10-30	60	1	All up to 12 then 13	93.3	87.7	80.5	96.1
31-50	20	2	Select all	95.8	93.6	90.6	97.2
51-100	15	1	All up to 55 then 55	97.1	89.7	86.2	97.9
101-200	5	1	55	95.8	89.4	85.8	97.0

[1]Cutpoint equal to one up to herd size of 30 then equal to two for the rest of the herd sizes

categories for which the sample size increased above the level in the simulation using a cutpoint of one.

The survey-level sensitivity and specificity for the fixed cutpoint design was approximately 93% and 86%, respectively, regardless of whether the herds were sampled in proportion to the population or equal proportions were taken from each herd size category (Table 7.4). The survey-level sensitivity and specificity for the variable cutpoint design were close to 95% and 90%, respectively, again regardless of the herd sampling protocol.

The number of herds required to attain the desired confidence in estimation was between 127 and 152 (Table 7.5). The number of herds required for the fixed cutpoint sampling was approximately 20 greater than was required by the variable cutpoint sampling design. Confidence intervals created using the 5^{th} and 95^{th} percentile estimates of the adjusted herd-level prevalence fell within 10 percent of the true prevalence, suggesting that the sample size was sufficient to create the desired confidence interval.

The within-herd sampling designs resulted in substantially varying number of animals that needed to be tested (Table 7.6). Although more herds were required for the fixed cutpoint design, the total number of animals was less than required for variable cutpoint designs. The variable cutpoint design with equal proportions of samples taken from each herd size category had the largest total sample size requirement because of the greater sample size requirements in the three larger herd size categories.

Table 7.4. Survey-level sensitivity (SN_s), specificity (SP_s), positive predictive value (PPV) and negative predictive value (NPV) for herd size categories under two herd-level design strategies, with and without a fixed cutpoint of one, and two herd sampling strategies (proportion to the population and equivalent proportion in each herd size category). (SE=98%, SP=99%, π=20%, Ψ=60%, ρ=.1)

Herd sampling strategy	SN_s	SP_s	PPV	NPV
Fixed cutpoint =1				
In proportion to population	93.4	85.8	79.4	95.7
25% of herds from each herd size category	93.0	85.8	79.4	95.4
Variable cutpoint[1]				
In proportion to population	95.1	90.7	86.7	96.7
25% of herds from each herd size category	94.3	90.1	84.9	96.4

[1] Cutpoint equal to one up to herd size of 30 then equal to two for the rest of the herd sizes

Table 7.5. Sample size needed to estimate herd-level prevalence (95% confident with an error bound of 10 %), true prevalence and simulated adjusted herd-level prevalence estimates under two herd-level design strategies, with and without a fixed cutpoint of one, and two herd sampling strategies (proportion to the population and equivalent proportion in each herd size category).

Sample design and strategy	Number of herds	True herd-level prevalence	Adjusted prevalence		
			Mean prevalence	5th percentile	95th percentile
Cutpoint=1 proportion to pop.	150	38.1	36.6	28.6	44.8
Cutpoint=1 (25% from each herd size category)	152	37.1	39.2	30.7	47.5
Variable cutpoint, Proportion to pop.	127	38.8	37.3	29.1	45.3
Variable cutpoint, (25% from each herd size category)	131	37.1	38.7	30.4	42.5

[1]Distribution characteristics from 1,000 simulations of sampling the specified number of herds

Table 7.6. Estimated number of animals needed to estimate herd-level prevalence under two herd-level sampling strategies (fixed cutpoint of one and variable cutpoint). Within each strategy herds were sampled by two different ways in proportion to how they occurred in the population and a fixed 25% within each herd category.

Herd size	Mean number sampled per herd	Number of herds	Number of animals
Cutpoint=1, sample relative to proportion in population			
10-30	12.9	90	1161
31-50	16.5	30	495
51-100	19	22	399
101-200	20	8	160
Total		150	2215
Cutpoint=1, sample equal proportion in each herd size category			
10-30	12.9	38	490
31-50	16.5	38	627
51-100	19	38	722
101-200	20	38	760
Total		152	2599
Variable cutpoint, sample relative to proportion in population			
10-30	12.9	76	980
31-50	40	26	1040
51-100	54.75	19	1040
101-200	55	6	330
Total		127	3390
Variable cutpoint, sample equal proportion in each herd size category			
10-30	12.9	33	426
31-50	40	33	1320
51-100	54.75	33	1807
101-200	55	32	1760
Total		131	5313

3.3 Modeling two-stage sampling: OPP and Salmonella examples

Apparent within-flock prevalence of OPP in sheep flocks was highly variable (Figure 7.3) which resulted in an estimated ρ of 0.48. Six flock size categories were defined and almost 50 percent of the flocks had 100 or fewer adult ewes (Table 7.7). When a cutpoint was fixed, between 11 and 16 samples were required from each flock to keep HSP at approximately 70%. The HSP declined in the largest flock size to avoid HSE from declining further. When the cutpoint was variable and the HSE objective was 80%, the HSP was between 82.1% and 47.9% and sample sizes increased to a maximum of 90% in the largest flocks. Positive predictive values in the flock size categories remained relatively high in both designs but the negative predictive values dropped to low levels in large flocks in both designs albeit more in the fixed cutpoint design. Survey-level test characteristics showed an improvement in all measures for the variable cutpoint design compared to the fixed cutpoint design.

The improved test characteristics of the variable cutpoint design for OPP resulted in a smaller sample size of flocks than did the fixed cutpoint design (Table 7.8). However, the increased number of animals that needed to be tested in each flock size category resulted in more than double the number of animal tests than did the fixed cutpoint design.

The distribution of apparent within-herd prevalence for *Salmonella* spp. in finishing pigs is skewed to the right (Figure 7.4). The four-level hierarchical model for describing the proportion of the variance of *Salmonella* spp. at each level indicated that most of the

Figure 7.3. Distribution of apparent within-flock prevalence of ovine progressive pneumonia in sheep flocks sampled in the 2001 National Animal Health Monitoring System's sheep study.

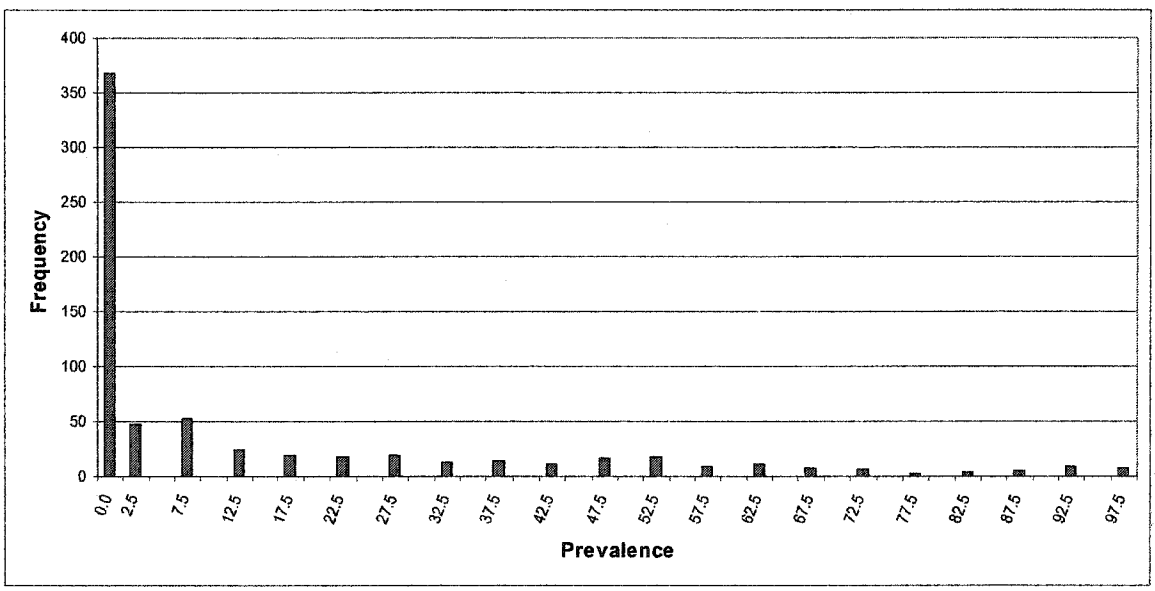


Table 7.7. Herd-level sensitivity, specificity, positive predictive value and negative predictive value for herd size categories under two herd-level design strategies (HSP=70% and HSE=80%), for testing for Ovine Progressive Pneumonia in sheep flocks.

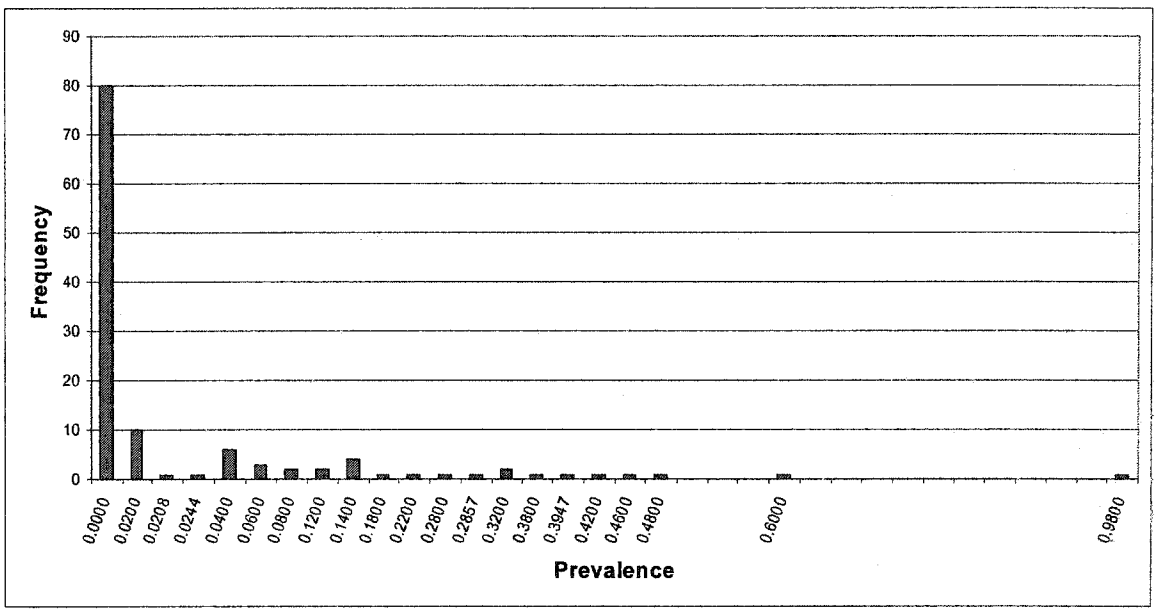
Herd size	Percent of herds	Within-herd sampling	HSE	HSP	Positive Predictive Value	Negative Predictive Value
Fixed cutpoint=1, keeping herd-level specificity approximately 70%						
20-50	24	11	85.7	70.8	71.5	85.3
51-100	24	11	79.7	69.9	73.4	76.8
101-200	17	11 up to 150, then 12	75.3	69.3	76.4	68.0
201-500	15	12	73.1	69.7	78.9	62.5
501-2000	15	12	68.3	68.4	82.1	50.4
2001-5000	5	16	71.7	60.7	82.5	45.4
Sample of herds from population		--	76.3	70.0	77.1	69.2
Variable cutpoint[1], keep herd-level sensitivity approximately 80%						
20-50	24	All up to 25, 25	86.3	82.1	79.9	87.9
51-100	24	25 up to 80 then 30	81.5	79.9	80.7	80.6
101-200	17	30 up to 150, then 50	78.6	77.6	81.4	74.4
201-500	15	50 up to 250, then 60	78.9	72.9	82.0	68.8
501-2000	15	75	81.5	58.7	80.7	60.0
2001-5000	5	90	84.2	47.9	80.6	54.0
Sample of herds from population		--	81.4	77.6	81.5	77.4

[1]Cutpoint =2 up to herd sizes of 150 then increases to 3.

Table 7.8. Sampling design for herd-level prevalence estimation of Ovine Progressive Pneumonia for two herd-level sampling strategies (fixed cutpoint of one and flexible cutpoint).

Herd size	Mean number sampled per herd	Number of herds	Number of animals
Fixed cutpoint=1, keeping herd-level specificity about 70%			
20-50	11	106	1166
51-100	11	104	1144
101-200	11.5	74	851
201-500	12	67	804
501-2000	12	67	804
2001-5000	16	21	336
Total		439	5105
Variable cutpoint, keep herd-level sensitivity about 80%			
20-50	24.5	66	1617
51-100	27	65	1755
101-200	40	46	1840
201-500	58.3	42	2450
501-2000	75	42	3150
2001-5000	90	13	1170
Total		274	11982

Figure 7.4. Distribution of apparent within-herd prevalence of *Salmonella* spp.in finish pigs sampled in the 2000 National Animal Health Monitoring System's swine study



variation in the response is at the farm level (51.7%) followed by the animal (30.0%), pen (10.4%) and building levels (7.9%). When the hierarchical model was reduced to just farm and animal level, the proportion of variance at each level was 0.67 and 0.33, respectively. This model provided the estimate of ρ (0.33) for designing the two-stage sampling.

Five herd size categories were constructed with the relatively even proportions of the herd population in all but the largest herd size category (Table 7.9). The sample sizes required to attain 70% HSE in the herd size categories were between 45 and 225 while the sample sizes required to attain 80% HSE were between 88 and 600. The 600 samples in the largest herd size category did not result in a HSE of 80%, but from an applied perspective it did not seem feasible to continue to inflate the sample size. A cutpoint of one was used because the animal-level test specificity was assumed to be one. The perfect test specificity also resulted in HSP and positive predictive values of one. Negative predictive values declined with increasing herd size categories in both designs.

The design intended to attain a HSE of 70% required sampling 47 more farms than did the design for HSE of 80% to meet the sample size requirements to construct a 95% confidence interval with a error bound of 10% (Table 7.10). The number of tests in the latter design was almost double the required number in the former design. The sample size in the design for a HSE of 80% would have been even higher if the sample size in the largest herd size category had been allowed to exceed 600.

Table 7.9. Herd-level sensitivity (HSE_f), and negative predictive value for herd size categories under two herd-level design strategies (HSE=70% and HSE=80%), for testing for *Salmonella* spp. in finishing pigs. Herd-level specificity and positive predictive value are equal to one because animal-level test specificity=100%.

Herd size	Percent of herds	Within-herd sampling	HSE	Negative Predictive Value
Fixed cutpoint=1, target of herd-level sensitivity of 70%				
100-600	26	45	69.9	70.2
601-1500	24	78	69.9	61.0
1501-3000	23	115	70.1	57.2
3001-8000	22	157	69.7	53.1
8000-20000	5	225	69.9	46.8
Sample from population		--	69.5	59.7
Fixed cutpoint=1, target of herd-level sensitivity of 80%				
100-600	26	88	80.0	76.3
601-1500	24	180	80.1	71.5
1501-3000	23	295	80.3	67.1
3001-8000	22	420	80.0	62.8
8000-20000	5	600	78.8	57.1
Sample from population		--	79.5	69.7

Table 7.10. Sampling design for herd-level prevalence estimation of *Salmonella* spp. for two herd-level sampling strategies (herd-level sensitivity approximately 70% and 80% respectively).

Herd size	Number sampled per herd	Number of herds	Number of animals
Fixed cutpoint=1, keeping herd-level sensitivity approximately 70%			
100-600	45	52	2340
601-1500	78	47	3666
1501-3000	115	45	5155
3001-8000	157	45	7065
8000-20000	225	10	2250
Total		199	20496
Fixed cutpoint=1, keep herd-level sensitivity approximately 80%			
100-600	88	40	3520
601-1500	180	36	6480
1501-3000	295	34	10030
3001-8000	420	34	14280
8000-20000	600	8	4800
Total		152	39110

4. Discussion

Combining the model to estimate herd-level sensitivity and specificity and survey-level test characteristics with the formula for herd sample size provides a practical approach to designing two-stage sampling plans for estimating herd-level prevalence. Often a practical limitation of a modeling approach is the lack of empirical information upon which to base assumptions (Jordan and McEwen, 1998) but the minimal requirements for this model are the distribution of herd sizes, reliable estimates of animal-level test sensitivity and specificity, and some information on the potential distribution of within-herd prevalence.

The intracluster correlation coefficient along with the animal level prevalence can be used to specify the distribution of within-herd prevalence in a population. Results of this study indicate that herd-level test characteristics can be influenced by the level of clustering in herds, especially at lower prevalence levels. McDermott and Schukken (1994) emphasized the need to consider cluster effects in epidemiological studies of animal populations and described the ANOVA method for estimating ρ when cluster sizes are equal; they used the approach to estimate a range for ρ when the herd sizes were unequal. They also listed correlation coefficients from a number of studies on various species and diseases. Detailed data were available to allow us to estimate ρ using a hierarchical model. Estimates of ρ from actual test data will most likely be affected by test sensitivity and specificity. However, the estimate provides some basis for designing studies. Further research efforts are needed to develop adjusted estimates of ρ.

When no estimate of ρ is available, a qualitative assessment of ρ may be appropriate. Ariwan and Frerichs (1996) suggested qualitative levels of low, medium, and high intracluster correlation which they associated with values of the design effect: 2, 4 and 7 respectively. The design effect is the variance of an estimated proportion obtained by a cluster sample divided by the variance for a simple random sample (Dargatz and Hill 1996). Their software converts the design effect into an estimate of ρ using the following formula:

$$\rho = \frac{(deff - 1)}{(\overline{m} - 1)}$$

where deff=design effect and $\overline{m}$ = average cluster size.

The design of the swine *Salmonella* sampling raises the issue of how to design sampling that is multi-stage and not just two-stage. Similar circumstances will arise in other species such as feedlot cattle in pens on feedlots and layers in houses on a layer operation. Clustering can occur within any of the levels of interest. In this study, more than 80 percent of the variation in the occurrence of *Salmonella* occurred at the farm and animal level. Thus, a two-stage design may provide an approximate answer to the sample size of farms and animals within the farm but does not provide a mechanism for deciding on the number of buildings and number of pens to sample on a farm. Dunlop et al. (1999) found that a large proportion of the variation of resistance (>0.8) of *Escherichia coli* to two antimicrobials was at the between-pig, within-pen level. They recommended that sampling emphasis be placed on individual pigs rather than on the pen or room

levels. Because their study was conducted on a single farm, variation at the farm-level was not a factor.

A common approach to determining whether a herd is "free" from disease is to design a sampling protocol to detect a minimum within-herd prevalence (Garner et al. 1997; Cameron and Baldock 1998b; Cannon 2001). For example, Garner et al. (1997) assumed that, in infected swine herds, at least 25% of the finisher pigs would have antibodies to the porcine reproductive and respiratory syndrome virus. The minimal value can be considered to be the lower tail of the distribution of within-herd prevalence in a population that is not free from disease such as would be expected in a herd-level disease estimation problem. Assuming a minimal value will result in a conservative estimate of HSE. If the distribution of the within-herd prevalence follows a "bathtub" shape, such as when ρ is relatively high, HSE may be liberal if the design is based on the average within-herd prevalence. Thus, modeling the distribution of within-herd prevalence may provide a more accurate assessment of survey-level test characteristics for sample size determination than does a detection level. Regardless of the approach, assumptions regarding the distribution of within-herd prevalence will have an impact on the design. The assumption should be carefully considered and well documented to support the selected design.

In the hypothetical and applied examples for designing studies to estimate herd-level prevalence, it is clear that there are trade offs in choosing design strategies. With the hypothetical example the variable cutpoint design substantially improved the survey-level

specificity and the positive predictive value while decreasing the number of herds that needed to be visited and increasing the number of animals for testing by approximately 30 to 50 percent depending on the alternatives. Similar results were observed for OPP. The two *Salmonella* designs compared only herd-level sensitivity, but as the number of animals required for testing increased the number of herds decreased. These trade offs give the designer an option for evaluating the costs and objectives of proposed studies.

Studies conducted by NAHMS often have the dual purpose of estimating disease (animal and herd levels) and investigating risk factors. Herd-level prevalence estimation can be adjusted using the Rogan and Gladen method or through Bayesian techniques that are available (Enoe et al. 2000; Johnson et al. 2001). The potential for misclassification must be considered when risk factors are being evaluated. Greiner and Gardner (2000) discuss the misclassification at the individual-test level when the risk factor and the diagnostic classification both are not readily observable and surrogate measurements must be used. Unadjusted odds ratios systematically underestimate the true odds ratio when misclassification is non-differential. When differential misclassification occurs, the odds ratio may be biased in either direction. However, differential misclassification can be a product of the design. For example, in the OPP variable cutpoint design with the HSE close to 80%, the HSE was relatively stable by design, but the HSP fell between 48% for large herds to 82% for smaller herds. This variability is reflected in the negative predictive values. Christensen and Gardner (2000) suggest that it is theoretically possible to adjust odds ratios in a risk-factor study if HSE and HSP are known.

Alternatively, when the primary objective is to examine risk factors, some of the bias may be removed by adjusting the design to minimize differential misclassification.

The hypothetical population example allowed for examination of different selection probabilities for the herds in different herd size categories. The survey-level sensitivities were consistent whether the herds were sampled in proportion to the population or in equal proportions across the herd size categories. This effect is due to the relatively homogenous herd-level sensitivities and specificities across the herd size categories. If HSE and HSP varied dramatically across the herd size categories then altering the sampling proportions within herd size categories could have an impact on the survey-level test characteristics. The prevalence estimates used for the sample size estimates were slightly affected by the sampling protocol. This bias can be removed with a design-based analysis that assigns sampling weights to observations (Dargatz and Hill, 1996).

The model developed here has many similarities to models developed by Audige and Beckett (1999) and Jordan and McEwen (1998) but contains a combination of features not available in either of the other models. Our model allows for the use of a beta-binomial distribution for prevalence as does the Jordan model, but our model calculates the beta distribution parameters based on intracluster correlation coefficients and animal-level prevalence. Additionally, several intracluster correlation coefficients can be evaluated simultaneously. Unlike Jordan's model, we do not incorporate the flexibility to examine different prevalence distributions, but the model could be modified to accommodate those options. Our model differs from Audige's model in that our

objective is to determine herd-level sensitivity and specificity under the assumption of disease clustering for use in estimating herd-level disease prevalence while their objective is to assess disease freedom. Another difference between our model and Audige's is the use of herd-level sensitivity and specificity. We use the herd-level sensitivity and specificity to determine the number of herds to sample. Audige and Beckett (1999) developed herd-level test characteristics and then applied them to infected and non-infected herds in the second part of their model to build likelihood ratios for freedom versus a specified prevalence. Also, our model was developed to output individual herds so that herd-level testing strategies could be assessed.

Problems with using the beta-binomial distribution, of which the parameters are estimated ρ and animal-level prevalence, to model clustered populations have been noted by others. The probability of zero prevalence in the continuous distribution is zero so that non-infected herds are not probabilistically plausible. Donald (1993) and Donald et al. (1994) recommended dividing the population under study into infected and non-infected populations. Herd-level specificity applies only to non-infected herds and herd-level sensitivity applies to the infected herds. A minimum of a single positive animal is forced into infected herds if the modeled prevalence is too low to assign one based on binomial probabilities. Audige and Beckett (1999) used a similar strategy while Jordan and McEwen (1998) offered an option for forcing the minimum of a single positive into infected herds. We used the approach of dividing the population into infected and non-infected for the theoretical modeling and chose not to force a single positive animal into "positive" herds. The model has the flexibility to adopt either choice. The modeling of

OPP and *Salmonella* did not assume the presence of an infected and non-infected population.

References

Ariwan, I. and R.R. Frerichs. User's manual: CSurvey version 1.5. 1996. Univ. of Indonesia, Indonesia and Univ. of Calif., Los Angeles.

Audige, L. and Beckett, S. 1999. A quantitative assessment of the validity of animal-health surveys using stochastic modelling. Prev Vet Med **38:** 259-76.

Bohning, D. and Greiner, M. 1998. Prevalence estimation under heterogeneity in the example of bovine trypanosomosis in Uganda. Prev Vet Med **36:** 11-23.

Cameron, A.R. and Baldock, F.C. 1998a. A new probability formula for surveys to substantiate freedom from disease. Prev Vet Med **34:** 1-17.

Cameron, A.R. and Baldock, F.C. 1998b. Two-stage sampling in surveys to substantiate freedom from disease. Prev Vet Med **34:** 19-30.

Cannon, R.M. 2001. Sense and sensitivity - designing surveys based on an imperfect test. Prev. Vet. Med. **49:** 141-163.

Carpenter, T.E. and Gardner, I.A. 1996. Simulation modeling to determine herd-level predictive values and sensitivity based on individual-animal test sensitivity and specificity and sample size. Prev. Vet. Med. **27:** 57-66.

Christensen, J. and Gardner, I.A. 2000. Herd-level interpretation of test results for epidemiologic studies of animal diseases. Prev. Vet. Med. **45:** 83-106.

Dargatz, D.A. and Hill, G.W. 1996. Analysis of survey data. Preventive Veterinary Medicine **28:** 225-237.

Donald, A. 1993. Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm. Preventive Veterinary Medicine **15:** 125-145.

Donald, A.W., Gardner, I.A., and Wiggins, A.D. 1994. Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. Prev. Vet. Med. **19:** 167-187.

Dunlop, R.H., McEwen, S.A., Meek, A.H., Friendship, R.M., Black, W.D., and Clarke, R.C. 1999. Sampling considerations for herd-level measurement of faecal *Escherichia coli* antimicrobial resistance in finisher pigs. Epidemiol. Infect. **122:** 485-496.

Enoe, C., Georgiadis, M.P., and Johnson, W.O. 2000. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Prev. Vet. Med. **45:** 61-81.

Garner, M.G., Gleeson, L.J., Holyoake, P.K., Cannon, R.M., and Doughty, W.J. 1997. A

national serological survey to verify Australia's freedom from porcine reproductive and respiratory syndrome. Aust Vet J **75:** 596-600.

Greiner, M. and Gardner, I. A. Application of diagnostic tests in veterinary epidemiologic studies. Prev. Vet. Med. 45[1-2], 43-59. 2000.

Johnson, W.O., Gastwirth, J.L., and Pearson, L.M. 2001. Screening without a "gold standard": the Hui-Walter paradigm revisited. Am J Epidemiol **153:** 921-4.

Jordan, D. and McEwen, S.A. 1998. Herd-level test performance based on uncertain estimates of individual test performance, individual true prevalence and herd true prevalence. Prev Vet Med **36:** 187-209.

Martin, S.W., Shoukri, M., and Thorburn, M.A. 1992. Evaluating the health status of herds based on tests applied to individuals. Preventive Veterinary Medicine **14:** 33-43.

McDermott, J.J. and Schukken, Y.H. 1994. A review of methods used to adjust for cluster effects in explanatory epidemiological studies of animal populations. Preventive Veterinary Medicine **18:** 155-173.

Rogan, W.J. and Gladen, B. 1978. Estimating prevalence from the results of a screening test. Am. J. Epidemiol. **107:** 71-76.

Stark, K.D., Mortensen, S., Olsen, A.M., Barfod, K., Botner, A., Lavritsen, D.T., and Strandbygard, B. 2000. Designing serological surveillance programmes to document freedom from disease with special reference to exotic viral diseases of pigs in Denmark. Rev Sci Tech **19:** 715-24.

Wineland, N.E. and Dargatz, D.A. 1998. The National Animal Health Monitoring System. A source of on-farm information. Vet Clin North Am Food Anim Pract **14:** 127-39.

Appendix

The variance of the Rogan and Gladen estimator can be rewritten as follows:

$$Var(\hat{\theta}_{RG}) = \frac{Var(\hat{p})}{(HSE + HSP - 1)^2} = \frac{(\hat{p}(1-\hat{p}))/n}{(HSE + HSP - 1)^2} = \frac{\hat{p}(1-\hat{p})}{n(HSE + HSP - 1)^2}$$

where $\hat{p}$ = unadjusted estimate of prevalence, HSE=herd-level sensitivity, HSP=herd-level specificity and n=within-herd sample size.

The variance estimate can be used to estimate the bounds on error associated with an estimate. The usual formula for the bound on error for a sample proportion, $\hat{p}$, is:

$$Error \leq Z_c \sqrt{\frac{p(1-p)}{n}}$$

where Z_c is the value from standard normal distribution corresponding to the confidence level c. Thus the bound in error of the Rogan and Gladen estimator provides the following formula:

$$Error =\leq Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n*(HSE + HSP - 1)^2}}$$

Solving for n results in the following formula for estimating sample size

$$n = \frac{p(1-p)}{(HSE + HSP - 1)^2} * \left(\frac{Z_{\alpha/2}}{Error}\right)^2$$

CHAPTER 8

OVERALL CONCLUSIONS

- Antimicrobial resistance is a global problem involving humans, companion animals, agricultural production animals, crops, and the environment. Analytical methods are critical for developing an understanding of the complexities of the issue. A review of veterinary and human antimicrobial resistance literature was conducted to identify common data analysis techniques. Additional techniques that have not been applied were proposed.
  - Descriptive statistical techniques are widely used in the literature. A cumulative distribution of minimum inhibitory concentrations, standardized for resistance breakpoints, was proposed for visually comparing resistance data for multiple antimicrobials. Simple indices have been used to summarize antimicrobial resistance data and other indices may have application. Categorical and nonparametric methods have been used primarily to analyze susceptible/resistant categorized data. Multivariate and cluster analytical techniques have not been widely used but have potential application.
  - Advantages and disadvantages of the analytical methods must be considered when selecting analytical techniques for antimicrobial

resistance data. Minimum inhibitory concentration (MIC) data potentially retains more information than does collapsing into susceptible/resistant categories. Use of MIC data is encouraged. Observations may not be independent because of clustering of animals or the selection of multiple isolates from the same single source. Analytical methods should account for the clustering. Multivariate techniques can be an important tool when susceptibility is tested for several antimicrobials on a single isolate.

- Understanding the processes that either promote or inhibit development of antimicrobial resistance in feedlot cattle is critical for determining potential management tools and interventions. The ability to measure antimicrobial susceptibility depends on appropriate sampling methodologies. A study was conducted to compare antimicrobial susceptibility patterns of *Escherichia coli* isolated from fecal samples obtained per rectum versus pen floor samples and individual animal samples versus pooled samples.
    - Patterns of antimicrobial resistance for fecal samples obtained from the rectum of cattle did not differ from those for fecal samples obtained from pen floors.
    - Resistance patterns from pooled samples differed from patterns observed for single fecal samples. Pooling of fecal samples yielded resistance patterns that were consistent with those of single fecal samples when the prevalence of resistance to an antimicrobial was > 2%. Pooling may be a practical alternative when investigating patterns of resistance that are not rare.

- Little pen-to-pen variation in resistance prevalence was observed. Clustering of resistance phenotypes within samples was detected. Apparent clustering of resistance phenotypes within samples argues for examining fewer isolates per fecal sample and more fecal samples per pen.
- Describing and comparing resistance patterns requires assessing multiple characteristics of the data. A variety of analytical approaches is appropriate for evaluating these characteristics.

- Short-term stability of measurements of antimicrobial susceptibility of *Escherichia coli* organisms isolated from feces of feedlot cattle is important in developing monitoring and surveillance programs. Frequent evaluations (e.g. daily) are resource intensive and in some situations may be impractical for long term sampling protocols. Consequently, a point-in-time measurement will need to be used to represent conditions in the peri-sampling period.
  - In this study, 30 fecal samples were collected from each of six cattle pens on a commercial cattle feedlot on two occasions separated by 48 hours. *Escherichia coli* were isolated from single and pooled samples. Resistance to five antimicrobials (ampicillin, nalidixic acid, streptomycin, sulfamethoxazole, and tetracycline) was detected in single and pooled samples from both sampling periods (Day 0 and Day 2). The prevalence of isolates resistant to these five antimicrobials was two percent or higher in all treatment combinations except for pools obtained from Day 2 samples. Lower levels of resistance to six more antimicrobials were detected inconsistently across the single and pooled samples.

- The apparent inconsistencies in detection and prevalence do not appear to be related to external factors but rather to sampling intensity. Short-term stability is a plausible assumption under sampling strategies that are designed to detect specific levels of prevalence. However, when resistance levels fall below these levels, there will likely be fluctuations in the presence/absence of rare resistance phenotypes and in their prevalence and central tendency measures.

- Panels of antimicrobials are often used to test the susceptibility profile of bacterial isolates. Often univariate susceptible/resistant proportions are used to describe testing results. Antibiograms, a tabulation of observed resistance phenotype combinations, is an attempt to address the co-occurrence of resistance to multiple antimicrobials. However, the complexity of interrelationships can be difficult to describe using an antibiogram, and the use of susceptible/resistant categorization can result in the loss of information.
  - Factor analysis was used to assess relationships in the minimum inhibitory concentration among 17 antimicrobials tested on isolates of *Escherichia coli* isolated from 360 fecal samples obtained from feedlot cattle. Six factors were extracted using maximum-likelihood factor analysis. The factors were interpretable antimicrobial groupings based on class of antimicrobial and previously described associations. New generation cephalosporins, older generation beta-lactams, fluoroquinolones, and aminoglycosides grouped separately as classes of antimicrobials on four of the six factors. One of the remaining factors was a grouping of

antimicrobials that had been identified as being related in previous feedlot studies that examined antimicrobial resistance in *Salmonella*. The last factor was a grouping of three of the five antimicrobials that comprise the antimicrobials found in penta-resistant strains of *Salmonella* Typhimurium.

- The factor analysis described patterns in the MIC data that would not have been apparent if only susceptible/resistant categorized antimicrobial-resistance data had been analyzed.

- Production animals often are raised in clusters such as a herd or flock. Determination of the health status of the cluster often requires a diagnostic test. The use of imperfect diagnostic tests to determine health status of a cluster or a group of clusters has received much attention in the veterinary literature. Important considerations in evaluating herd-level tests include animal-level test sensitivity and specificity, sample size, cutpoint, within-herd prevalence, and herd size. The notion of herd-level sensitivity and specificity arises from characterizing the effect of the combination of these considerations on the ability to correctly classify the status of a cluster.
  - Two-stage sampling approaches for establishing freedom from disease have been based upon the relationships between within-herd and herd-level sampling, as well as individual animal diagnostic and herd-level test characteristics. Sampling at both stages is focused on detection of disease rather than disease prevalence estimation.

- Studies of the distribution of disease in a population may be concerned with estimating herd-level prevalence rather than disease detection. We propose the use of a herd-level sample size formula based on a common adjustment for prevalence estimates when diagnostic tests are imperfect. The formula depends on an estimate of herd-level sensitivity and specificity. A Monte Carlo simulation model is developed to calculate these herd-level test characteristics when disease may be clustered in herds.
    - Simulations explore the effect of varying levels of intracluster correlation on herd-level sensitivity and specificity. At low prevalence, herd-level sensitivity increased with increasing intracluster correlation but the sensitivity was less affected at higher prevalence. Also at low prevalence many herds are being classified as positive based only on false positive test results. Positive predictive values drop sharply with increasing intracluster correlation.
    - A hypothetical example of two-stage sampling was created to examine the impact on herd and within-herd sample sizes of a single test positive versus a variable number of positives for declaring a herd positive. More herds but fewer animals within herd were needed for the single test positive design versus the variable cutpoint design to attain the same precision on the herd-level prevalence estimate.
    - A comparison of designs for estimating the flock-level prevalence of Ovine Progressive Pneumonia in sheep also demonstrated that variable

cutpoints resulted in more sheep needed per flock and fewer flocks than did a fixed cutpoint of one.

- A second example considered two designs for two-stage sampling of *Salmonella* spp. in finishing swine herds. The assumption of 100% specificity negated the need for consideration of cutpoints other than one. Increased sample size within herds resulted in increased herd-level sensitivity while decreasing the number of herds needed to estimate herd-level prevalence with desired precision. Allocation of the within-herd sample size to a multi-level sampling strategy (farm, building on farm, pen within building, and animal) was based on the proportion of the variance from each level.
- The two-stage sampling approach for estimating herd-level prevalence is a flexible methodology that allows researchers to address sampling objectives. The use of a distribution for within-herd prevalence results in a conservative estimate of herd-level test characteristics. The model allows for researchers to trade off between the number of herds and the number of animals sampled by manipulating herd-level test characteristics.
- Sampling designs can result in problems with differential misclassification which may lead to biased prevalence estimates and unreliable risk factor analyses. The model can be used to adjust designs to minimize the potential for differential misclassification of herd status.

Future directions for research

- Research studies start with questions and hypotheses but also end with new questions and research hypotheses. There are several areas of research that should be pursued as outgrowths of this research.
  - The model for estimating herd-level prevalence utilized a beta-binomial distribution that is intended to be derived from estimated animal-level prevalence and the intracluster correlation coefficient. The intracluster correlation coefficient was estimated in the Ovine Progressive Pneumonia and the *Salmonella* examples using empirical data. However, just as with prevalence estimates, sampling designs and animal-level test characteristics would most likely affect the estimate of the intracluster correlation coefficient. The Bayesian methods used by researchers at UC Davis may well provide more accurate estimation of the intracluster correlation.
  - The herd-level sample size equation assumed that herd-level sensitivity and specificity are measured without error. The impact of this assumption should be investigated. Perhaps the sample size equation can be modified to allow for variability in the herd-level test characteristics.
  - Disease freedom models depend on two-stage sampling designs which require a prediction of the herd disease status. This step of requiring herd-level disease status is intuitively appealing and, in fact, addresses some international requirements for declaring disease freedom. However, the

issues which complicate herd-level disease detection (cutpoint, distribution of within-herd prevalence [clustering], and animal-level test characteristics) can potentially be circumvented. Traditional two-stage sampling designs do not attempt to estimate first-stage parameters. A simpler approach for assessing freedom may be to rely on estimates of disease at the animal-level.

- The National Animal Health Monitoring System has conducted a number of studies in which the biological sampling design was based on the assumption of perfect diagnostic tests. Risk factor analyses were performed based on cutpoints of a single positive test result. The likelihood of a single positive likely varied according to sampling intensity based on herd size. I believe a different risk factor profile might emerge if these analyses were revisited. One approach would be to use Poisson regression analysis to allow the sample size to be incorporated into the analysis. Another approach would be to construct a different sampling plan that might reduce some of the specificity problems with larger herds. This approach could mean reducing, after the fact, the sample size in larger herds or changing the cutpoint for herd-level designation (positive/negative). Results from these simulated designs could be compared to the original results to assess the impact of the design.