THESIS

This is to certify that the
thesis entitled

IDENTIFYING SOURCES OF SURFACE WATER CONTAMINATION BY DISCRIMINANT ANALYSIS OF PATTERNS OF ANTIMICROBIAL RESISTANCE IN *E. COLI*

presented by

Raida Sayah Sayah

has been accepted towards fulfillment
of the requirements for the

M.S. degree in Large Animal Clinical Sciences (Epidemiology)

Major Professor's Signature

August 18, 2004

Date

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

6/01 c:/CIRC/DateDue.p65-p.15

IDENTIFYING SOURCES OF SURFACE WATER CONTAMINATION BY DISCRIMINANT ANALYSIS OF PATTERNS OF ANTIMICROBIAL RESISTANCE IN *E. COLI*

By

Raida Sayah Sayah

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Large Animal Clinical Sciences
(Epidemiology)

2004

ABSTRACT

IDENTIFYING SOURCES OF SURFACE WATER CONTAMINATION BY DISCRIMINANT ANALYSIS OF PATTERNS OF ANTIMICROBIAL RESISTANCE IN *E. COLI*

By

Raida Sayah Sayah

A repeated cross–sectional study was conducted to investigate whether discriminant analysis of antimicrobial resistance profiles of *E. coli* would be useful in identifying sources of fecal contamination of surface water from the Red Cedar (Michigan) watershed. Fecal samples were collected from livestock, companion animals, human septic tanks, farmed deer and wild geese from the watershed, and water was collected from several different sites on the Red Cedar River. Disc diffusion was used to test for susceptibility to neomycin, gentamicin, streptomycin, chloramphenicol, ofloxacin, trimethoprim/sulfamethoxazole, tetracycline, ampicillin, nalidixic acid, nitrofurantoin, cephalothin, and sulfisoxazole. Both resubstitution and cross-validation discriminant function analysis methods were applied to the data.

Resistance to at least to one antimicrobial was seen in isolates from domestic animals, wildlife, surface water, and human septic tanks. Overall, *E. coli* isolates from food animals showed resistance to the largest number of antimicrobials, followed by horses, companion animals, human septic tanks, farmed deer, wild geese and surface water. Discriminant analysis using the resubstitution method produced a higher but biased average rate of correct classification (ARCC), while cross-validation produced lower but valid ARCC. Also, the ARCC improved as the number of source categories decreased.

ACKNOWLEDGMENTS

I would first like to thank USDA and Population Medicine Center for their financial support of this project.

I am grateful to my major advisor Dr. John B. Kaneene for all the encouragement and tremendous support that he provided from the beginning to the end of this project. Without his epidemiological experience, wise guidance and strong determination to keep meeting the challenges and working through the obstacles that come along with this venture, my success would not be possible.

I would like to thank Dr. Joseph Gardiner, Dr. Michellel Kopcha and Dr. Scott Witter for being a part of participating on my committee.

A special thank you goes to Dr. Yvette Johnson for her enormous support and help throughout the study.

I would like to thank Dr. Mark Hansen; Ingham County Extention Agent and Dr. Betsy Deerburger; Livingston county Extention Agents for assisting us reach out to the farmers.

I would like to thank all the participated farmers in this study and the lab technicians: Katie May at Michigan State University, Janice Donohoe at the University of Maryland and all the lab staff at Michigan Department of Environmental Quality, I would like to thank Mr. Barry Looper from the Department of Natural Resources, Mr. Randy Abbot from Ingham County Drain Commissioner, Mr. Bob Godbold and Mr. Rene Franco from Ingham County Environmental Health Department, Mr. Don Hayduk from Livingston County Environmental Health Department, RoseAnn Miller from the Population Medicine Center and Betty Wernette-Babian for all the help in collecting and processing

the fecal and water samples, and analyzing the data. Without their assistance this project would not be possible.

Finally, my deepest thank to my family especially my husband and children who give me the strength on a day-to-day basis to undertake my responsibilities, and make everyday worthwhile; they are my inspiration in all I do.

TABLE OF CONTENTS

LIST OF TABLES.....vii

LIST OF FIGURES.....viii

KEY TO APPREVIATIONS.....x

INTRODUCTION.....1
Purpose.....1
Hypotheses tested.....2
Objectives.....2
Overview.....2

CHAPTER 1
LITERATURE REVIEW: USE OF DISCRIMINANT ANALYSIS FOR THE CLASSIFICATION OF ECOLOGICAL DATA.....4
Definition.....4
History of discriminant analysis.....5
Types of discriminant analysis.....6
Measuring the performance of discriminant analysis.....7
Use of discriminant analysis in ecological studies.....9
Assumptions.....21
Interpretation of discriminant analysis.....24
Logistic regression and discriminant analysis.....25
Cluster analysis and discriminant analysis.....25
MANOVA and discriminant analysis.....26
Conclusion.....27

CHAPTER 2
PATTERNS OF ANTIMICROBIAL RESISTANCE OBSERVED IN *E. COLI* ISOLATES OBTAINED FROM DOMESTIC AND WILD ANIMAL FECAL SAMPLES, HUMAN SEPTAGE, AND SURFACE WATER IN MICHIGAN, USA....30
Abstract.....30
Introduction.....31
Hypotheses tested.....35
Objectives.....35
Materials and Methods.....36
Results.....47
Discussion.....58

CHAPTER 3
DISCRIMINANT ANALYSIS AS A TOOL FOR CLASSIFYING THE SOURCES OF *E. COLI* CONTAMINATION OF SURFACE WATER IN MICHIGAN, USA.....80
Abstract.....80

Introduction....81
Hypotheses tested....87
Objectives....88
Materials and Methods....88
Results....92
Discussion....97

OVERALL SUMMARY AND CONCLUSIONS....107

APPENDEX....110
Informed consent form....111
Letter from Population Medicine Center to farmers....112
Letter from Ingham County Extension Agent to farmers....113
Letter from Livingston County Extension Agents to farmers....114

REFERENCES....115

LIST OF TABLES

Table 2-1: Concentrations and diffusion zone break points for resistance for antimicrobial agents tested in this study, by class of antimicrobial agent 46

Table 2-2: Antimicrobial agents reported as having been used on the farms in the different species of animals sampled 48

Table 2-3: Percentage of resistant isolates for specific antimicrobial agents, by species exposure class 55

Table 2-4: Percentage of multi-drug resistant isolates, by species exposure class 56

Table 2-5: Most commonly identified combinations of agents in multi-drug resistant isolates from food animals.....57

Table 3-1: Comparison of classification tables for classification rules generated by non-parametric discriminant analysis of antimicrobial resistance profiles (neomycin, streptomycin, tetracycline, ampicillin, sulfamethoxazole, cephalothin, sulfisoxazole) using two different rule development methods, for five animal classes 94

Table 3-2: Comparison of classification tables for classification rules generated by non-parametric discriminant analysis of antimicrobial resistance profiles (neomycin, streptomycin, tetracycline, ampicillin, sulfamethoxazole, cephalothin, sulfisoxazole) using two different rule development methods, for two animal classes.....95

Table 3-3: Comparison of average rate of correct classification (ARCC) values between re-substitution and cross-validation rule development for different animal classification systems 95

Table 3-4: Classification of *E. coli* from water samples using discriminant analysis.....96

LIST OF FIGURES

Figure 2-1: Michigan State map showing the location of Red Cedar Watershed 37

Figure 2-2: The Red Cedar Watershed, Michigan, USA 37

Figure 2-3: The sampler used to collect water samples 42

Figure 2-4: Disc diffusion test plate showing the zones of growth inhibition of *E. coli* around an antimicrobial agent.....45

Figure 2-5: Antimicrobial resistance of *E. coli* from cattle 69

Figure 2-6: Antimicrobial resistance of *E. coli* from cattle farms 69

Figure 2-7: Antimicrobial resistance of *E. coli* from sheep 70

Figure 2-8: Antimicrobial Resistance of *E. coli* from sheep farms 70

Figure 2-9: Antimicrobial resistance of *E. coli* from swine 71

Figure 2-10: Antimicrobial resistance of *E. coli* from swine farms 71

Figure 2-11: Antimicrobial resistance of *E. coli* from horses 72

Figure 2-12: Antimicrobial resistance of *E. coli* from horse farms 72

Figure 2-13: Antimicrobial resistance of *E. coli* from poultry 73

Figure 2-14: Antimicrobial resistance of *E. coli* from poultry farms 73

Figure 2-15: Antimicrobial resistance of *E. coli* from companion animals 74

Figure 2-16: Antimicrobial resistance of *E. coli* from farmed deer and wild geese 74

Figure 2-17: Antimicrobial resistance of *E. coli* from human septage 75

Figure 2-18: Antimicrobial resistance of *E. coli* from water 75

Figure 2-19: Antimicrobial resistance of *E. coli* from a farm with pigs and chickens.....76

Figure 2-20: Antimicrobial resistance of *E. coli* from a farm with cattle, horses and sheep.....76

Figure 2-21: Antimicrobial resistance of *E. coli* isolates from chicken and food animals 77

Figure 2-22: Antimicrobial resistance of *E. coli* isolates from companion animals and food animals.....77

Figure 2-23: Antimicrobial resistance of *E. coli* isolates from companion animals and horse.....78

Figure 2-24: Antimicrobial resistance of *E. coli* isolates from horse and food Animals.....78

Figure 2-25: Multi-drug resistant *E. coli* isolates 79

KEY TO ABBREVIATIONS

AFLP	Amplified Fragment Length polymorphism
ANERR	Apalachicola National Estuarine Research Reserve
ARCC	Average Rate of Correct Classification
ATCC	American Type Culture Collection
CFU	Colony Forming Unit
DA	Discriminant Analysis
DNR	Department of Natural Resources
E. coli	*Escherichia coli*
IMVIC	Indole Methyle red Voges proskauer Simmon Citrate
LDA	Linear Discriminant Analysis
MANOVA	Multiple Analysis Of Variance
MAR	Multiple Antibiotic Resistance
MIC	Minimum Inhibitory Concentration
NARMS	National Antimicrobial Resistance Monitoring System
NCCLS	National Committee for Clinical Laboratory Standards
PFGE	Pulsed field Gel Electrophoresis
RNA	Ribonucleic Acid
TMP/SMX	Trimethoprim Sulfamethoxazole
TSA	Triple Sugar Iron
TSA	Tryptic Soy Agar
TSB	Tryptic Soy Broth
USA	United States of America
USDA	Untied States Department of Agriculture
USEPA	United States Environmental Protection Agency
VRB	Violet Red Bile

INTRODUCTION

Purpose

Water plays a major role in sustaining the natural systems on and under the earth's surface. Contamination of surface water poses serious risks not only to human and animal health but also to environmental health on a global scale.

Chemical fertilizers, livestock manure, and sewage may lead to deterioration of water quality. However, equally threatening to both quality, and human and animal health is the potential damage to water caused by bacterial contamination from manure and sewage. Bacteria such as *Salmonella*, *Campylobacter* and *E. coli* are considered a major threat to human health and are routinely isolated from animal and human feces. Efforts to monitor water quality by determining the level of contamination with coliform bacteria have been undertaken for more than a century. Attempts have also been made to determine the source of fecal contamination. Discriminant function analysis has been reported to provide a promising tool that could offer a low cost and statistically valid means of identifying the source species of water contamination based on the antibiotic sensitivity profile of the *E. coli* isolates from different sources. Since the magnitude of threat posed by bacterial contamination of surface water in Michigan is unknown and protecting water quality is a vital issue from the perspective of environmental protection and human health, epidemiological research is needed to determine the extent to which human, livestock, companion animals, and wildlife contribute to contamination of surface water in the state. The overall goal of the studies in this thesis was to investigate whether combining two techniques (determining antimicrobial resistance profiles of *E. coli* and conducting discriminant function analysis on such profiles) would be useful in

identifying the source of fecal contamination of surface water in Michigan. The Red Cedar Watershed was used as a case site. It is hypothesized that such techniques would be useful tool in programs aimed at reducing the risk of surface water contamination.

Hypothesis Tested

The general hypothesis that was tested in this thesis was that, "Domestic animal, human, and wildlife sources of fecal contamination of surface water can be identified using discriminant function analysis of antimicrobial resistance patterns of *E coli* isolated from such sources."

Objectives

1. Identify patterns of antimicrobial resistance in *E. coli* isolates obtained from human septage, domestic and wild animals fecal samples in Red Cedar Watershed- Michigan, USA.
2. Identify patterns of antimicrobial resistance in *E. coli* isolates obtained from surface water samples in Red Cedar Watershed- Michigan, USA.
3. Apply discriminant function analysis of antimicrobial resistance patterns in identifying the source of fecal contamination of surface water.

Overview

The thesis is arranged in chapters as outlined below. In addition to the overall hypothesis stated above, specific hypotheses are presented for chapter 2 and chapter 3. Chapter 1 is a literature review of the use of discriminant function analysis in ecological studies. Particular attention is paid to its successful use in classification the ecological data. Chapter 2 discusses patterns of antimicrobial resistance in *E. coli* isolates obtained from domestic animal, human and wildlife fecal samples, and surface water samples in

Michigan, USA. Chapter 3 describes the application of the discriminant function analysis to the anti-microbial resistance profiles of *E. coli* isolated from different sources in order to classify and identify the source species of microbial surface water contamination in Red Cedar Watershed in Michigan, USA. Chapter 4 presents an overall summary for the thesis and recommendations for future research.

CHAPTER 1

LITERATURE REVIEW: USE OF DISCRIMINANT ANALYSIS FOR THE CLASSIFICATION OF ECOLOGICAL DATA

Definition

Discriminant analysis or discriminant function analysis is a multivariate statistical method entails separating sets of observations and allocating new observations to previously defined groups (Tatsuoka, 1970; Morrison, 1990; Johnson and Wichern, 1992; Lattin et al., 2003). It is exploratory in nature. As a separator procedure, it is often employed on a one-time basis in order to investigate observed differences when causal relationships are not well understood. A classification procedure is less exploratory in the sense that it leads to well-defined rules, which can be used for assigning new observations (Johnson and Wichern, 1992). The two goals for discriminant analysis are to determine which variables discriminate between two or more naturally occurring groups and to classify cases into the values of categorical dependent groups (Tatsuoka, 1970; Morrison, 1990; Johnson and Wichern, 1992; Lattin et al., 2003). For example, an educational researcher may want to investigate which variables discriminate between high school graduates who decide to go to college, to attend a trade or professional school or seek no further training or education. For such an investigation, the researcher could collect data on numerous variables prior to students' graduation such as achievement motivation, academic performance, and personality. After graduation, most students will fall into one of the three categories. Discriminant analysis could be used to determine which variables are the best predictors of the students' subsequent educational choice. Measurement of height in a random sample of 50 males and 50 females provides another

example. Females are, on the average, not as tall as males and this difference will be reflected in the difference in means for the variable height. Therefore, the variable height allows us to discriminate between males and females with a better than chance probability: if a person is tall then he is likely to be a male, if a person is short then she is likely to be a female. The basic idea behind discriminant function analysis is to determine whether groups differ with regard to the mean of a variable, and then to use that variable to predict group membership. So that rather than test the usual hypothesis of equal mean vectors, we wish to construct a linear compound or index for summarizing observations from the groups on a one-dimensional scale that discriminates between the populations by some measure of maximal separation.

History of discriminant analysis

Discriminant analysis was first proposed by Fisher in1936 as a statistical tool for use in taxonomic problems originally confined to two-group situations (Fisher, 1936). He suggested finding a linear combination of observations that would maximize the difference between groups relative to the standard deviation within groups. Then discriminant analysis involved more than two groups in which more than one discriminant function is found (Rao, 1948; Tukey, 1949). The number of discriminant functions in multi-group case is one less than the number of groups (Tatsuuoka, 1970). Wald (1944) and Anderson (1951), created the classification statistic (W), in order to assign an observation to population I if classification statistic W is usually greater than 0, otherwise to population II. Mahalanobis (1936) originated the Mahalanobis squared distance, which is the distance between a case and the centroid for each group of the

dependent. The smaller the Mahalanobis distance, the closer the case to the group centroid, and the more likely it is to be classified as belonging to that group (Mahalanobis, 1936). When the parameters are unknown, they may be replaced by their unbiased sample estimators to give the sample quadratic discriminant function, which was originally proposed by Smith (1947). This resulted in discriminant analysis being widely used in many types of ecological studies.

Types of discriminant analysis

Data for discriminant analysis consist of observations for which there are a grouping index and associated vector of measurements. One objective of the analysis is to predict the group to which an observation belongs based on its measurement value. Such a formulation is called predictive discriminant analysis, and the prediction equations are called classification or discriminant functions (Lattin et al., 2003). Alternatively, the objective may be to exhibit optimal separation of groups, based on certain linear transforms of the measurement variables, and this is called descriptive discriminant analysis (Lattin et al., 2003). The associated linear functions are known as canonical variates, the structure of which is often of primary concern to ecologist (Hix, 1988). Both methods have been used in ecological studies but most have used the descriptive orientation (Williams, 1983, 1988). However, there are studies that used the predicative methods (Rice et al., 1983; Verner et al., 1986).

When the distribution within each group is assumed to be multivariate normal, a parametric method can be used to develop a discriminant function. It is also known as classification criterion and determined by a measure of generalized square distance (Rao,

1973). The classification criterion can be based on individual within group covariance matrices yielding a quadratic function of the pooled covariance matrix producing a linear function. When no assumption can be made about the distribution within each group, or when the distribution is assumed not to be multivariate normal, nonparametric methods can be used to estimate the group specific densities. These methods include the kernel and k-nearest neighbor methods (Rosenblatt, 1956).

Measuring the performance of discriminant analysis

The performance of discriminant analysis can be evaluated by its ability to classify future observations using error count estimates and posterior probability error rate estimates. The error count estimate can be calculated by applying the classification criterion derived from the training sample data set that discriminant analysis uses to derive the discriminant function- to a test set, and then counting the number of misclassified observations. The group specific error count estimates are the proportion of misclassified observations in the group. It has an optimistic bias called "apparent error rate." The estimation of error rate is very important, especially in medical application. For example, in the field of electrocardiography, a patient is diagnosed as being healthy or unhealthy on the basis of the results of an electrocardiogram; for this particular application of W it is vital not to underestimate the error of classifying a patient as healthy when the patient actually does have a heart disease.

Many techniques are used to estimate the error rates of sample discriminant functions and to reduce the bias (Lachenbruch and Mickey, 1968). Most of the time, these methods are not satisfactory in cases where the sample sizes are smaller than the variables. The

first one, the ***hold out*** method, is often used for validation of the discriminant function (Lattin et al., 2003; Wiggins et al., 2003). This is a split halves test, where a portion of the cases are assigned to the analysis sample for purposes of training the discriminant function, then it is validated by assessing its performance on the remaining cases in the hold-out sample. In other words, the training sample is used to construct the classification function, and the validation sample is used to evaluate it. The validation sample is not biased and also possesses the same properties as the analysis sample. The proportion misclassified in the validation sample determines the error rate. Although this method overcomes the bias problem by not using the same data to both build and judge the classification, there are drawbacks to this method.

First, in many applications, large samples are not available especially when the data are expensive to obtain. Second, if the size of the holdout sample is large, a good estimate of the performance of the discriminant function will be obtained, but that function is likely to be poor. On the other hand, if the size of the holdout sample is small, the discriminant function will be better, but the estimate of its performance will be highly variable (Lachenbruch and Mickey, 1968). Third, the function evaluated is not the function of interest. Ultimately, almost all of the data must be used to construct the classification function; if not, valuable information may be lost (Johnson and Wichern, 1992). Another way to reduce the bias is ***cross validation*** (Lachenbruch and Mickey, 1968; Lattin et al, 2003; Wiggins et al., 2003). Cross validation treats n-1 out of n training observations as a training set. It determines the discriminant functions based on n-1 observations and then applies them to classify the one observation left out. This is done for each of the training observations. The misclassification rate for each group is

the proportion of sample observations in that group that are misclassified (Lattin et al., 2003; Wiggins et al., 2003).

The third method is the ***resubstitution*** method in which the sample used to compute the discriminant function would be reused to estimate the error (Smith, 1947; Lattin et al., 2003; Wiggins et al., 2003). When the test set is independent of the training sample, the estimate is unbiased. However, it can have a large variance, especially if the test set is small. The good classification (low error rates) will depend upon separation of the populations. The farther apart the groups, the more likely a useful classification rule can be developed (Johnson and Wichern, 1992).

Several studies addressed the appropriate statistical assumptions (Williams, 1983) and the potential prediction bias (Verbyla, 1986; Hix, 1988), involved when applying discriminant analysis in ecological studies. The prediction bias is likely to occur when the number of independent variables in the model is large relative to the sample size, or when many different sets of independent variables are tested by a stepwise procedure (Verbyla, 1986), or when multivariate normality cannot be assumed (Hix, 1988). Biased results could occur by using categorical and continuous data sets; mixed data sets have been used in several ecological studies (Pregitzer and Barnes, 1984; Spies and Barnes 1985a; Hix, 1988).

Use of discriminant analysis in ecological studies

Discriminant analysis is used widely in ecological studies in which multiple measurements are made on samples of observations possessing an identifiable group structure. Its application would focus on the structure of plant or animal communities

indexed by geographically distinct habitat. In this case, the samples in each habitat consist of the abundance of species and the objective is to highlight differences in community structure (Matthews, 1979; Tonn and Magnuson, 1982). On the other hand, discriminant analysis could be used to highlight habitat differences separating different animal species. In this case, the samples of each species would consist of multiple habitat measurements and the objective is to highlight differences in habitat use (Munro and Rounds, 1985; Seagle, 1985). Matthew (1979) applied multiple discriminant analysis to a number of plant assemblage-types in order to differentiate, as clearly as possible, and give information about within and between type variability. Multiple discriminant analysis is the classification that derives discriminant functions which maximize the between group variance and minimize the within group variance on the discrimiannt function scores. Norris and Barkham (1970) found multiple discriminant analysis to be a useful technique in the analysis of differences between some beechwood ground floras in the Cotswold Hills, England, but did not extend their study to include an analysis of within wood variability. Grigal and Golstein (1971) gave graphical representation of variability within and between four types of woodland in the oak hickory forest of eastern Tennessee, USA, and concluded that the four types were distinct.

In Tonn and Magnuson's (1982), study discriminant analysis was used to discern two assemblage types of fish in eighteen lakes. All the *Umbra*-cyprinid and centrarchid –*Esox* assemblages groups were correctly classified in the study, and each had a distinctive species composition and seasonal change in composition. Also, discriminant analysis was applied to the log transformed environmental data on two of the eighteen lakes. The purpose was to evaluate the environmental distinctness between the two groups of lakes

to help identify environmental factors contributing to their separation (Tonn and Magnuson, 1982). This aided the ecologist to properly manage the lakes and maintain these assemblies.

When forestland classification is needed to develop a framework for managing forests, an ecological multifactor approach integrating climate, physiography (physical geography or landforms), soils, and vegetation was used to develop a classification of the upland hardwood forest ecosystems of the Kickapoo River Watershed of southwestern Wisconsin (Hix, 1988). After applying discriminant analysis on the dataset, it was found that a combination of physiographic, soil, and vegetational variables resulted in the highest overall percentage of correct classification about 97%. It is similar to results obtained by Pregitzer and Barnes (1984) and Spies and Barnes (1985a). The overall probabilities of correct classification were 81% and 91%, respectively.

Discriminant analysis has been used to classify trees as decayed versus sound (LeMay et al, 1994). Age, size, and the presence or absence of external indicators of possible internal decay are variables used in deriving various rules for classifying trees. The success of each developed classification rule was evaluated using the misclassification error rates.

Discriminant analysis has been used to differentiate between human and animal sources of fecal pollution in natural waters (Wiggins 1996; Wiggins et al., 1999; Hagedorn et al., 1999; Harwood et al., 2000) using a set of antimicrobial resistance profiles for fecal streptococci in the USA, and using a set of antimicrobial resistance profiles in *E. coli* in one study in Canada (Guan et al., 2002) and another study in USA (Parveen et al., 1998), and to a set of antibiotic resistance profile in *Enterococcus* isolates

in Virginia Watershed, USA (Graves et al., 2002). Discriminant analysis has also been applied on ribotyping profile for *E. coli* isolates to differentiate human and non-human source of fecal pollution of water (Parveen et al., 1999; Carson et al., 2001; Troy et al., 2003). Wiggins (1996) first addressed the application of discriminant analysis of antimicrobial resistance patterns of streptococci to identify fecal pollution sources of water. A total of 1,435 streptococcus isolates from 17 samples of cattle, poultry, human and wildlife were used to create the antibiotic resistance profile of five antibiotics (chlortetracycline, halofuginone, oxytetracycline, salinomycin, and streptomycin) using four concentrations of each antibiotic. After analysis, 74% of known isolates were correctly classified into one of six possible sources (beef, chicken, dairy, human, turkey and wildlife). Ninety-two percent of human isolates were correctly classified. When isolates were pooled into four categories, cattle, human, poultry and wildlife, the average rate of correct classification (ARCC) increased to 84%. Human versus animal isolates were correctly classified on average of 95%. Human and wildlife had an ARCC = 98%, cattle versus poultry had ARCC= 92%. However, 72% of isolates from surface waters that received fecal pollution from unknown sources from one site were classified as cattle isolates, and 68% isolates from the other sites were classified as cattle also. The data strongly suggest that discriminant analysis can be used to differentiate among isolates from several sources of fecal pollution and to determine the sources of fecal pollution in natural waters. The use of several concentration of each antibiotic to establish antimicrobial resistance profiles coupled with discriminant analysis, provides the predictive power necessary to provide useful information about the sources of isolates from surface water. Another study performed by the same author (Wiggins et al., 1999),

used a larger sample size. Four sets of isolates of fecal streptococci (from 2,635 to 5,990 isolates per set) were obtained from 236 samples of human sewage, cattle and poultry feces and pristine waters. The patterns of antimicrobial resistance of the isolates to each four concentrations of up to nine antibiotics were analyzed by discriminant analysis. The ARCC in four possible groups (human, cattle, poultry and wildlife) ranged from 64% to 78%. It is confirmed that there are measurable and consistent differences in the antimicrobial resistance patterns of fecal streptococci isolated from various sources of fecal pollution, and it could be used to classify and identify these sources. Hagedorn et al. (1999) validated the Wiggins's methods by using larger database (7,058 fecal streptococcus isolates) of known sources including human, livestock and wildlife from wide geographical region and thirteen antibiotics, the five reported by Wiggins plus amoxicillin, ampicillin, chloramphinicol, erythromycin, neomycin sulfate, rifampin, tetracycline and vancomycin hydrochloride. The correct fecal streptococcus identification averaged 87% for the entire database and ranged from 84% for deer to 93% for human isolates. In order to test the database, and additional 892 streptococcus isolates were compared against the database and resulted in ARCC of 88%. When all animals' isolates were combined the ARCC increased to greater than or equal to 95%. Stream samples from three highly contaminated collection sites, found that 78% of the fecal streptococcus isolates from these sites were classified as cattle. These results were consistent and support the Wiggins's results. Based on these results, the cattle access to the river was prohibited by installation of fencing and in- pasture watering stations. These interventions lead to less than 45% of fecal streptococcus isolates being classified as

cattle. Additionally, the fecal coliforms counts were reduced from 15,900 per 100ml to 960 per 100ml.

The use of thirteen antibiotics was necessary in Hagedorn's (1999) study in order to find those that did provide levels of separation that were as high as possible. Hagedorn et al. (1999) supported his results by using cluster analysis with discriminant analysis to get additional confidence when the two methods provide the same answers.

Another study by Harwood et al. (2000) also supported the research done by Wiggins (1996) and Hagedorn et al. (1999). In his study, Harwood et al. (2000) used two databases; one for 4619 fecal streptococcus isolates and one for 6144 fecal coliform isolates, from a large geographical area in Florida. The antimicrobial resistance profiles were established using nine antibiotics (ampicillin, amoxicillin, cephalothin, chlortetracycline, oxytetracycline, streptomycin, tetracycline, erythromycin, and vancomycin). Only eight antibiotics were used for fecal coliforms because gram-negative bacteria are not susceptible to vancomycin. In the previous studies, Wiggins (1996) and Hagedorn et al. (1999), the highest correct classification rates were obtained by the use of a subset of antibiotics tested for analysis while in the Harwood et al. (2000) study, omission of any antibiotic resistance data resulted in lower correct classification rates. The ARCC for fecal streptococcus database was 62.3% and that for fecal coliform database was 63.9%. These results are lower than Wiggins' original study results in 1996, because the database was composed of bacteria isolated from sources within a limited geographical area, and the sample sizes were relatively small. However, the second study by Wiggins (1999) was less homogeneous (more sampling sites) and samples sizes were larger and resulted in correct classification rates comparable to

Harwood's (2000) results. The lower correct classification rate for some sources reflected the geographic diversity. Hagedorn's et al. (1999) study resulted in higher classification rates than Harwood's et al. (2000) study because the sources of isolates designated human in the two studies probably contributed to the differences in correct classification rates. In Harwood's et al. (2000) study, the human isolates were obtained from domestic wastewater, which provides a cross-section of human antibiotic resistance profile and, thus, high variability in antibiotic resistance. In Hagedorn's et al. (1999) study, the human isolates were obtained from experimental domestic wastewater treatment systems from individual homes providing lower sample variability. The sample is not likely to be representative of a large human population, which yielded higher correct classification rates because antibiotic resistance variability is lower.

A study done in Canada (Guan et al., 2002) in which a collection of 319 *E. coli* isolates from feces of cattle, poultry, swine, deer, goose, and moose as well as human sewage and clinical samples were used. Fourteen antibiotics, ampicillin, cephalothin, streptomycin, neomycin, kanamycin, gentamicin, tetracycline, chloramphinicol, sulfathiazole, cotrimoxazole, apramycin, ceftiofur, spectinomycin, and tilmycosin were used in this study. By applying discriminant analysis on all *E. coli* isolates, the ARCC was 33.9% when all isolates were classified into nine groups (human, beef, dairy, chicken, pig, turkey, deer, goose, and moose). The ARCC for moose was 100% and 80% for chicken, while goose, beef - dairy cattle and deer were poorly classified (0,0, and 14.8% respectively). It was suggested that most goose and beef-dairy isolates were misclassified into moose categories because these groups displayed similar multiple antibiotic resistance (MAR) profiles. When deer, goose, and moose isolates were pooled

together as the wildlife group, and swine, turkey, chicken, and bovine isolates were pooled together as livestock group, the ARCC for wildlife and livestock were 95% and 46%, respectively, and 55% for human isolates. When all nonhuman sources were pooled and discriminant analysis applied on human and nonhuman isolates, the ARCC for human increased to 56.25% and to 92.38% in nonhuman. These results are different from Wiggins (1996) and Hagedorn et al. (1999) studies due to several factors. First, *E. coli* is the microorganism that was investigated in this study, not streptococcus that was investigated in the previous studies. Second, different antibiotics were used in all the studies, and third, the collection and isolation of *E. coli* is different in the three studies. In Guan et al. (2002), study samples were collected from farms and parks over a wide geographical area, and only one *E. coli* isolate was selected from each animal's fecal sample so that repetitious selection of the same clone of *E. coli* was avoided. The sampling protocol could produce heterogeneous collection of bacterial isolates than other protocols in which multiple bacterial isolates were derived from each fecal sample.

In the Guan et al. (2002) study, three methods were evaluated for their ability to differentiate *E. coli* isolates from various sources. Multiple antibiotic resistance (MAR), amplified fragment length polymorphism (AFLP) and 16S r RNA sequences. It was found that AFLP provided the greatest discriminatory power, the highest rate of correct classification and ease of standardization and automation, but it is expensive.

Another study done by Parveen et al. (1998) in which discriminant analysis was used on a MAR profile of *E. coli* isolates from the Apalachicola National Estuarine Research Reserve (ANERR) found that 82% of human source isolates and 68% of non-human isolates were classified with an ARCC of 75%.

Grave et al. (2002), built a library of 1174 known source *Enterococcus* isolates, and then applied the discriminant analysis on antibiotic profiles from these isolates. She found that the ARCC was 94.6% for 203 human isolates, 93.7% for 734 livestock isolates and 87.8% for 237 wildlife isolates. The antibiotic resistance analysis of enterococcal isolates recovered from the stream samples indicated that isolates of human origin appeared throughout the stream, but it was small compared with the proportion of isolates from the livestock, which was 50%. The results presented in this study, affirm the use antimicrobial resistance analysis and discriminant analysis in the previous studies (Wiggins, 1996; Wiggins et al.1999; Hagedorn et al., 1999; and Harwood et al., 2000). Today, with the fast development in the genotypic methods in determining the source of fecal contamination in water, the applicability of ribotyping to predict the source of *E. coli* pollution has been tested using discriminant analysis. (Parveen et al., 1999; Carson et al., 2001; Troy et al., 2003). These studies strongly indicated that discriminant analysis of ribotypes profiles could be used to differentiate human sources and non-human sources of fecal pollution. Discriminant analysis of ribotype profiles of 238 *E. coli* isolates showed that 97% of the nonhuman isolates and 100% of animal fecal isolates were correctly classified, and the ARCC for both was 82% (Parveen et al., 1999). The advantage of discriminant analysis is that it generates a classification rule based on all isolates; that rule can then be used to classify each individual isolates into one of many possible sources. Once a discriminant analysis classification rule is developed by using ribotyping profiles of *E. coli* isolates from known sources, discriminant analysis can then be used to classify an unknown isolates to one of the known sources by using unknown organism's ribotypes profiles (Parveen et al., 1999). According to Parveen et al. (1999)

the MAR has disadvantages that antibiotic resistance patterns of bacteria are influenced by selective pressure, and thus may be different in other geographical areas and may vary over time. But ribotype profiles are a genetic characteristic and are not as easily influenced by the selective pressure. The discriminant analysis of ribotype profiles provides a strong method for differentiating human source and non-human source fecal pollution, and may enhance efforts to improve the natural quality of estuarine ecosystems and to assess the importance of upstream activities, local storm water runoff and marine activities (Parveen et al., 1999).

In Carson's et al. (2001) study, a total of 287 known host ribotype patterns were generated for *E. coli* strains isolated from human, cattle, pigs, horses, chickens, turkeys, geese and dogs. Discriminant analysis of riboprint of human and nonhuman isolates resulted in 95.0 and 99.2% correct classification, respectively. The ARCC for riboprints compared to all eight hosts classes was 73.6% and when the comparison was made between more limited numbers of classes, the ARCC improved.

Another study by Troy et al. (2003), using discriminant analysis on ribotype of *E. coli* isolates supported the idea that it may be possible to differentiate human from animal derived *E. coli* over a broad geographic region via the single-enzyme (HindIII) ribotype procedure. However, it is more difficult to differentiate between *E. coli* isolated from multiple non-human sources by using the same method- ribotype profiles. It showed that over all, the correct classification of animal derived *E. coli* isolates, as being either human or animal - derived was greater than 78%, while human - derived isolates were correctly classified greater than 85% of the time. Troy et al. (2003) also found that an overlap of ribotype profiles within and among animal groups was significant; the reasons

for the overlapping are unknown and the overlapping make it hard to differentiate sources of *E. coli*. It was thought that fecal material from birds could be present in the samples and cause the overlapping. Troy et al. (2003) concluded that the ribotyping procedure continue to be a viable molecular tool to be used for determining whether the source of *E. coli* is from human or animal sources. This conclusion supports the previous work reported by Parveen et al., 1999; and Carson et al., 2001.

Troy's et al. (2003) study is different from Carson's et al. (2001) study because of the diversity of sample collection and the type of the sample collected. In Carson's study, the samples collected were from central Missouri and from a small number of animals while in the Troy's study, composite fecal samples were collected from poultry, swine, dairy and beef cattle, over a wide geographic region. This type of sampling protocol was chosen because it is more likely that these samples would contain isolates that have the most environmental impact, subjected to external stressor, more likely to survive and more representative of those one would expect to find in the environment (Troy et al., 2003).

Although there is not an established standard of accuracy defined for any bacterial source tracking method, any method with a correct rate of classification of over 50% when there are five or more possible source categories has been considered as a worthwhile tool for predicting the potential sources of fecal pollution in environmental water (Harwood et al., 2000).

Discriminant analysis is used in human medicine to differentiate basal cell carcinoma and other skin neoplasms from normal skin by using infrared spectroscopy as a screening tool for cutaneous neoplasia with correct classification 93.5%. Linear discriminant

analysis characterized spectra, as arising from basal cell carcinoma, epidermis, or follicle sheath was 98.7% accurate. In addition, linear discriminant analysis accurately classified spectra as arising from epidermis overlying basal cell carcinoma versus epidermis overlying nontumor-bearing skin in 98.0% of cases (McIntosh, 1999). In this study, non-subjective classification of normalized spectra was performed using linear discriminant analysis (LDA). The genetic algorithm was implemented to select the regions of spectra that contained the most useful information. These spectral regions were then subjected to LDA, which determines the boundaries that best separate classes by computing the distance of each spectrum from all class centroids. It then allocates each spectrum to the class whose centroid is nearest. Data were split into a training set, which was used to train the LDA algorithm to find the discriminatory patterns in the data, and a test set to assess the accuracy of the trained algorithm (McIntosh, 1999).

The applicability of discriminant analysis to the interpretation of skin images was tested by Josef (2000) where the classification of tissue elements into 5 classes (background, epidermis, papillary dermis, reticular dermis and dermal infiltration) yielded a correct classification in 98.4% of all elements. Cluster analysis was used also, and when reclassification of cluster analysis was done by discriminant analysis it yielded 100% correctly classified elements. Josef (2000) concluded that discriminant analysis might be a helpful technique for an independent, and subjectively unbiased measurement system of skin structures in digital images.

Assumptions

Ecologists almost always use linear discriminant analysis; therefore, it is important to recognize the assumptions that are critical for its use.

Sample size: Unequal sample sizes are acceptable. The sample size of the smallest group needs to exceed the number of predictor variables. As a rule, the smallest sample size should be at least 20 for a few (4-5) predictors. The total sample size should be at least two or preferably three times as large as the number of variables measured (Tatsuoka, 1970). The maximum number of independent variables is n-2 where n is the sample size. While this low sample size may work, it is not encouraged, and generally it is best to have 4 or 5 times as many observations and independent variables (Johnson and Wichern, 1992).

The sample means and covariance must be estimated from the data. When the number of parameters to be estimated approaches the number of samples, there is a good likelihood that any patterns exhibited by individual coefficients are fortuitous, and therefore, of no ecological consequence. Therefore, data splitting, balanced repeated replication should always be used when sample size is small relative to dimensionality (Williams, 1983).

Normal distribution: It is assumed that the data (for the variables) represent a sample from a multivariate normal distribution (Johnson and Wichern, 1992; Lattin et al., 2003). One can examine whether or not variables are normally distributed with histograms of frequency distributions. However, violations of the normality assumption are not fatal, and the resultant significance tests are still reliable as long as non-normality is caused by skewness and not outliers (Tabachnick and Fidell, 1996). Klecka (1980) points out that

dichotomous variables, which often violate multivariate normality, are not likely to affect conclusions based on discriminant analysis.

Priori: Prior probabilities could influence the forms of discriminant functions so that incorrect specification of priors can distort or obscure any underlying structure of the data. In ecological studies, it is impractical or even impossible to sample in such a way that reasonable estimates of priors can be obtained. Instances involving rare species demonstrate such a lack and the need to use historical data to supplement samples, which also entails high costs associated with obtaining random samples. The solution is collecting systematic or stratified samples and priors are guessed, determined from ancillary information, or assigned arbitrarily. When priors are replaced by relative sample sizes that bear no direct relationship to them, an uncontrolled and largely inscrutable amount of arbitrariness is introduced into the discriminant analysis, so that initial decision of sample size can affect the resulting mathematical forms irrespective of underlying statistical properties (Williams, 1983).

Homogeneity of variances/ covariances: Discriminant analysis is very sensitive to heterogeneity of variance –covariance matrices (Lattin et al., 2003). It is assumed that the variance –covariance matrices of variables are homogeneous across groups, but researchers should perform some tests on the covariances to better understand their data. This includes tests for homogeneity of within class covariance matrices before accepting final conclusions for a study. In particular, a scatter plot matrix can be produced and can be very useful for this purpose. Lachenbruch (1975) suggested that discriminant analysis is relatively robust even when there is modest violation of this assumption.

Equal dispersion is a very important assumption since several properties results from it including linearity of discriminant functions, improved efficiency of estimation, and invariance of posterior probabilities in canonical space. This is important for interpretation of canonical plots, as it enable one to display data in canonical space without distortion. Most ecologists base their interpretation on canonical function, but their data exhibit heterogeneous dispersions. Heterogeneity of dispersion is manifested by the nonuniform patterns of dispersion in canonical space (Williams, 1983).

Outliers: Discriminant analysis is very sensitive to the inclusion of outliers. A test for outliers should be performed for each group, and outliers should be eliminated. If one group in the study contains extreme outliers that impact the mean, they will also increase variability. Overall significance tests are based on pooled variances, that is, the average variance across all groups. The significance tests of relatively larger means (with larger variances) would be based on the relatively smaller pooled variances, resulting erroneously in statistical significance.

Non-Multicollinearity: If one independent variable is very highly correlated with another, or one is a function (e.g., the sum) of other independents, then the tolerance value for that variable will approach 0, and the matrix will not have a unique discriminant solution. Such a matrix is said to be ill-conditioned, especially if any one of the variables is completely redundant with the other variables. There must also be low multicollinearity of the independents to the extent that independents are correlated; the standardized discriminant function coefficients will not reliably assess the relative importance of the predictor variables. The tolerance value is computed as *1 minus R-*

square of the respective variable with all other variables included in the current model. Thus, it is the proportion of variance that is unique to the respective variable.
When several assumptions are violated, statistical and interpretive problems are compounded. Discriminant analysis in this case, like any other mathematical technique, becomes a data-exploratory procedure. It may provide fruitful insights into the data, but without it, additional testing and evaluations are needed (Williams, 1983). However, researchers should know the difference between exploratory and confirmatory analysis. Statistical procedures can be used to explore data whether assumptions are met or not.

Interpretation of discriminant analysis

Ecologist are generally interested in parameters that separate populations, on the assumption that the operation of natural selection will be reflected in among-group differences of these parameters. A stepwise procedure is used to select variables that are highly separating groups, and then canonical transformations of these variables are determined. The canonical transformations are interpreted through the signs and magnitudes of the associated coefficients (Williams, 1983; 1988). In other words, the importance of a variable in a discriminant analysis is related to the magnitude and sign of its function coefficients (Hix, 1988), and by means of their correlations with the original variables (Williams, 1983). However, these coefficients do not tell us between which of the groups the respective functions discriminate. We can identify the nature of discrimination for each discriminant (canonical) function by looking at the means for the functions across the groups. Also, we can visualize how the two or more functions discriminate between groups by plotting the individual scores for the discriminant

functions (Williams, 1983). One should distinguish between discriminant analysis and logistic regression, cluster analysis, and Multiple Analysis Of Variance MANOVA.

Logistic regression and discriminant analysis

Discriminant analysis is a close kin to logistic regression. Although logistic regression answers the same questions as discriminant analysis, computing the functions is quite different. Logistic regression is often a preferred as an alternative procedure to discriminant analysis as it is more flexible in its assumptions, and the types of data to which it can be applied. It can handle both categorical and continuous variables, and the predictors do not have to be normally distributed, linearly related or of equal variances within each group (Tabachnick and Fidell, 1996). In their study, Press and Wilson (1978) compared logistic regression and parametric discriminant analysis, and concluded that logistic regression is preferable to parametric discriminant analysis in cases for which the variables do not have multivariate normal distributions within classes.
The interpretation of the results of a two-group problem is straightforward and closely follows the logic of multiple regression: Those variables with the largest (standardized) regression coefficients are the ones that contribute most to the prediction of group membership.

Cluster analysis and discriminant analysis

It is important not to confuse discriminant analysis with cluster analysis. All varieties of discrimiannt analysis require prior knowledge of the classes, usually in the form of a sample from each class. In cluster analysis, the data do not include information on class membership, and the purpose is to construct a classification. In other words, in

discriminant analysis the groups are determined beforehand, and the objective is to determine the linear combination of independent variables which best discriminates among the groups. In cluster analysis, the groups are not predetermined and in fact, the purpose is to determine the best way in which cases may be clustered into groups (Johnson and Wichern, 1992).

MANOVA and discriminant analysis

The close association between separation and multivariate analysis of variance has led to confusion about the nature of discriminant analysis. At first glance, discriminant analysis appears to do what MANOVA does, since both deals with differences among group of observations. Discriminant analysis addresses a statistical mixture of population and MANOVA does not. In fact, it is multivariate analysis of variance reversed, and computationally very similar to MANOVA. In MANOVA, the independent variables are the groups and the dependent variables are predictors. In discriminant analysis, the independent variables are the predictors (discriminating variables) and the dependent variables are the groups. The focus is on the groups as the outcome, and the main purpose is to develop a function that will maximize the distance among the groups (Morrison, 1990). MANOVA can be used to see the effect on multiple dependents of a single categorical independent, while discriminant analysis can be used to see the effect on a categorical dependent of multiple interval independents

In this case, we have a matrix of total variances and covariances; likewise, we have a matrix of pooled within-group variances and covariances. Those can be compared via multivariate F tests in order to determine whether or not there are any significant differences with regard to all groups and between groups. It is the same in MANOVA.

One could first perform the multivariate test and if it is statistically significant, proceed to see which of the variables have significantly different means across the groups. Another alternative to discriminant analysis is to perform a series of univariate, one - way Analysis of Variance ANOVAs. However, the advantage of the multivariate approach is that two or more classes that overlap considerably when each variable is viewed separately may be more distinct when examined from a multivariate point of view. According to Williams (1983), discriminant analysis has frequently been used improperly in ecology. The reasons are that multivariates procedures are complex, and require complex statistical analysis, and the documentation of these are highly technical.

Conclusion

Discriminant analysis is a multivariate statistical test used to perform three operations: a) classify cases into groups; b) determine which variable discriminates between groups by creating a classification rule based on the measurement of that variable; and c) use the discriminating variable to predict the group membership for an unknown sample. Discriminant analysis has two approaches, which are descriptive or predictive. The objective of discriminant analysis is to exhibit the maximum separation between groups based on linear transformation of the measurement variables (descriptive), and predicts the group to which an observation belongs based on its measurement value (predictive) respectively.

A reliable classification procedure should results few misclassifications and less error rate, so that the performance of discriminant analysis can be evaluated by estimating the error rate by either the ***hold-out***, ***cross validation*** or ***resubstitution*** methods.

Discriminant analysis is usually applied to a case of two populations so that effective allocation of an observation is probably not possible unless the populations are well separated. The same is true for many populations, which means the populations' mean vectors differ significantly from one another. Although apparent and significant differences do not automatically imply effective classification, testing is a necessary first step. If no significant differences are found, constructing classification rules will be a waste of time. Discrimination is often attempted with large number of variables, some of which are qualitative. In this situation, the multivariate normality may not be a sensible assumption, in which case the results of discriminant analysis may be affected. Therefore testing for normality is a very important issue. The distribution should be multivariate normal with equal covariance matrices. Discriminant analysis is very sensitive to sample size, and the sample size of the smallest group needs to exceed the number of predictor variables.

Usually, researchers deal with more than one population and many variables. This results in a very complicated situation requiring computer software programs to perform the calculations and have the capability for stepwise discriminant analysis. Selection of a subset of variables on the basis of minimizing the apparent error rate or maximizing discriminatory power may perform poorly in future samples, especially if there are large correlations among variables.

According to the articles reviewed in this chapter, discriminant analysis has been applied to a wide range of ecological problems in which multiple measurements are made on samples of observations possessing an identifiable group structure. More research has been conducted in the forestry industry where it focused on the structure of the plant, the

animal communities indexed by geographically distinct habitat, and classify trees into decayed versus sound where forestland classification is needed in order to apply the proper management plans and maintain the forest structure.

Discriminant analysis is a successful method to identify the sources of fecal pollution in water. It has been applied to the antimicrobial resistance patterns of fecal *Streptococcus* and *E. coli* bacteria that are considered indicator bacteria. The advantage of discriminant analysis method is that it generates a classification rule based on all bacterial isolates, then employs this rule to classify each isolate into one of many possible sources. According to Wiggins (1996), who first introduce this method, the high classification rates could be achieved when two sources are compared, supporting the hypothesis that different animal species will harbor different bacteria with different patterns of antibiotic resistance. Sometimes, one species of animal was misclassified into other animal species categories. For example chicken isolates were misclassified as turkey isolates indicating that pooling of the sources into poultry category would be helpful for the classification process. Usually, there is higher ARCCs when discriminant analysis is performed with pooled sources (poultry, cattle, wild animal and human) than all sources (chicken, turkey, beef, dairy, wild animal and human), or human and animal sources as two categories.

The use of discriminant analysis in research is simple, cost effective, and provides a strong method for differentiating human source and non-human source of fecal pollution of surface water. It will be suitable for surveillance purposes and routine monitoring that enhance the efforts to improve the quality of water, and it promises to accurately evaluate the risk to the public health.

CHAPTER 2

PATTERNS OF ANTIMICROBIAL RESISTANCE OBSERVED IN *E. COLI* ISOLATES OBTAINED FROM DOMESTIC AND WILD ANIMAL FECAL SAMPLES, HUMAN SEPTAGE, AND SURFACE WATER IN MICHIGAN, USA

Abstract

A repeated cross-sectional study was conducted to determine the patterns of antimicrobial resistance in *E. coli* isolated from fecal samples collected from domesticated animals, companion animals, farmed deer and wild geese. *E. coli* also isolated from human septic tanks, environmental samples on farms and surface water samples in Michigan, USA. Thirty-one farms, and three septic companies in the Red Cedar Watershed participated in the study. Isolation and identification of *E. coli* were conducted done using enrichment media, selective media, and biochemical tests. Antimicrobial susceptibility testing was conducted using the disc diffusion method. Antimicrobial agents tested included neomycin, gentamicin, streptomycin, chloramphenicol, ofloxacin, trimethoprim/sulfamethoxazole, tetracycline, ampicillin, nalidixic acid, nitrofurantoin, cephalothin, and sulfisoxazole. Resistance to at least one antimicrobial was demonstrated in isolates from food animals, wild geese, farmed deer, surface water, and human septic tanks. In general, *E. coli* isolates from food animal species showed the highest resistance to most antimicrobial agents, followed by those from horses, companion animals, wild geese, farmed deer, human septic tanks, and surface water. The top three-antimicorbial agents for which resistance was demonstrated were cephalothin, tetracycline, and streptomycin. At least one *E. coli* isolate from food animal species, wild gees, farmed deer, human septic tanks and water samples were

resistant to cephalothin. Additionally, the antimicrobial resistance profiles in *E. coli* isolated from companion animals and human septic tanks were closer to that from surface water than those from domesticated animals, suggesting that companion animals and human septic tanks may play a major role in contamination of surface water in the Red Cedar Watershed than previously thought.

Introduction

Antimicrobial resistance has been recognized as an emerging worldwide problem, and a serious problem in both human and veterinary medicine (Neu, 1992; Witte, 1998). Antimicrobial use is considered the most important factor for emergence, selection and dissemination of antimicrobial resistant bacteria in both human and veterinary medicine (Neu, 1992; Witte, 1998). Antimicrobial agents are used in animals and humans for therapy and control of bacterial infections. In food animals, it may be incorporated in feed and fed continuously to the whole population rather than to an individual animal as growth promoters, which may lead to selection of resistant bacteria. (Van Den Bogaard et al., 2001)

Multiple antimicrobial resistance testing is based on detection of bacterial resistance to a panel of antimicrobial agents. The principle behind developing resistance is that bacterial flora in the gut of humans and animals are subjected to different types, concentrations, and frequencies of antimicrobial agents. Over time, selective pressure will select resistant bacteria that have specific fingerprints against the antibiotics that have been used (Prescott et al., 2000; Troy et al., 2002). This test is simple, cost effective and suitable for surveillance (Troy et al., 2002).

With the exception of the large number of studies that have dealt with the antimicrobial resistance of *E. coli* isolated from food, various species of animals and humans (Krumperman, 1983; Langlois et al., 1983; Ginns et al., 1996; Blanco et al., 1997; Meng et al., 1998; Galland et al., 2001; Van Den Boggard et al., 2001; Schroeder et al., 2002a; Schroeder et al., 2002b), little is known about the antimicrobial resistance patterns of *E. coli* isolated from surface water.

Multiple antimicrobial resistance (MAR) profiling has been used to identify the sources of fecal contamination in water (Kasper et al., 1990; Wiggins, 1996; Parveen et al., 1997; Hagedorn et al., 1999; Wiggins et al., 1999; Harwood et al., 2000; Graves et al., 2002; Guan et al., 2002). The MAR profile may be a useful tool, from a public health and environmental protection standpoint, in establishing standards to determine water quality and facilitate detection of sources of water contamination. Efficient use of resources for water quality improvement needs to be based on accurate identification of the sources of fecal pollution. Hagedorn et al., (1999) reported that antibiotic resistance analysis proved to be accurate in identifying the source of fecal pollution in the Page Brook Watershed in Virginia. He found that livestock contributed more than humans to the contamination of the stream. Fencing part of the stream to reduce livestock's access to it reduced the fecal coliform population by an average of 94%.

The MAR profile is used to distinguish between *E. coli* that comes from point sources, such as industrial, municipal effluents, and meat-processing plants wastes; as well as non-point sources as in cases of soil erosion and runoff over a wide area of land. A study by Parveen et al., (1997) found that more than 80% of *E. coli* strains isolated from municipal waste, and river and estuarine water show antibiotic resistance. Kaspar et al., (1990)

showed that urban waters have higher percentage of resistant *E. coli* strains than rural waters, and antibiotic resistance *E. coli* may offer an index of water quality related to the source.

Multiple antibiotic resistance indices have been used to identify and differentiate *E. coli* of high-risk sources (human, poultry and swine) of fecal contamination of food from those of low-risk sources (wild animals) (Krumperman, 1983). Krumperman (1983) showed that multiple antibiotic resistance index of *E. coli* from wild animals (low risk) was generally low while human and poultry *E. coli* isolates had higher MAR indices (high risk) suggesting that multiple antibiotic resistance *E. coli* exist in large numbers within the major reservoirs of enteric diseases for humans while present in low number elsewhere. It is interesting that MAR index for isolates from human raw sewage was greater than the index found among isolates recovered by direct anal swabbing and this could be due to plasmid exchange that occurs in sewage systems (Krumperman, 1983).

Bacteria could gain resistance to antimicrobial agents by many ways including 1) acquisition of antibiotic resistance genes through mobile elements such as plasmids and insertion sequences (Rubens et al., 1979), which tend to code for enzymes that metabolize antibiotics (Prescott et al, 2000); 2) mutations in genes responsible for antibiotic uptake or binding sites (Spratt, 1994) and it often produces changes in bacterial cell, and 3) activation of MAR locus *(mar)* in the bacterial chromosome (Hachler et al., 1991; Alekshun and levy, 1999).

There are four mechanisms of resistance 1) enzymatic inactivation, or modification of antibiotics; 2) impermeability of the bacteria cell wall or membrane; 3) active expulsion

of the drug by the cell efflux pump, and 4) alteration in target receptors (Prescott et al., 2000).

The ability of resistance factor containing R+ bacteria, especially *E. coli*, to transfer drug resistance is well known (Richmond, 1972), with a large number of studies that discussed the antibiotic resistance patterns of *E. coli* isolated from fecal samples. Information describing antibiotic resistance patterns of *E. coli* (or other bacteria) isolated from the surface water in the environment is limited. Kelch and Lee (1978) suggested that resistant fecal coliform bacteria survive better than sensitive bacteria in surface water because R factor could increase the survival ability of coliform bacteria, supporting Grabow's findings (Grabow et al., 1973; Grabow et al., 1976). On the other hand, Anderson (1974) suggested that R factor-mediated antibiotic resistance might reduce survival ability of *E. coli*. Smith et al., (1974), however, suggested that R factor – mediated antibiotic resistance had no effect on *E. coli* survival.

It is widely speculated that the use of antimicrobial agents in food animals may be contributing to antimicrobial resistance in humans (Barton, 1998; Witte, 1998). Antibiotics fed at low levels as part of the feed have been used as growth promoters in domestic animals. Using antibiotics in this manner, will produce a substantial difference in the gut microflora of these animals, increasing the incidence of resistant bacteria especially when these bacteria are resistant to antibiotic used clinically for treating diseases in humans (Van den Boggard and stobberingh, 1999; Casewell et al., 2003). One strategy recommended in order to minimize this problem was to stop using antibiotics needed for human treatment as food additives (Swann Committee Report, 1969; Richmond, 1972; World Health Organization, 1997). Nevertheless, there is ongoing

debate whether and to what extent antibiotic use as feed additives contribute to resistance development in human bacterial pathogens (Prescott et al., 2000; Casewell et al., 2003). Antimicrobial resistant bacteria have been isolated from a variety of sources such as hospitals, domestic sewage, drinking water, rivers and lakes (Mulamattathil et al., 2000; McKeon et al., 1995; Kasper et al., 1990). The level of antibiotic resistance reported by Mulamattathil et al., (2000) was 72% lower than that reported by McKeon et al., (1995) which was 100 and 87% for fecal and non-fecal coliforms respectively.

Objectives

The two major objectives of this study were to 1) identify patterns of antimicrobial resistance in *E. coli* obtained from human septage, domestic animals and wildlife living in the Red Cedar Watershed in Michigan, USA; and 2) compare such patterns with that from *E. coli* obtained from surface water of the same watershed.

Hypotheses tested

- Prevalence of antimicrobial resistant strains of fecal *E. coli* isolates should be significantly lower in isolates obtained from wildlife fecal samples than those from domestic animals and humans.
- Prevalence of antimicrobial resistant strains of fecal *E. coli* isolates should be significantly lower in isolates obtained from domestic pets and human septage than those from domestic livestock and poultry.

Materials and Methods

Study design

A repeated cross- sectional study design approach was used to collect animal and human fecal, and water samples and all the data related to antibiotic use in the farm during the four seasons (Spring, Summer, Fall, and Winter), starting in the winter of 2002 and ending in the winter of 2003.

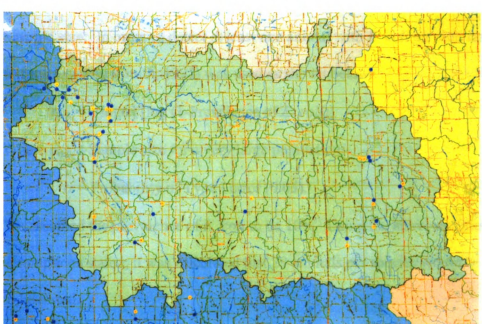
Study area

The Red Cedar Watershed was chosen as the study area, encompassing an area of 293,000 acres and takes in portions of 16 townships in Ingham and Livingston counties of the lower part of Michigan (figure 2-1). (Please notice that images in this thesis are presented in color). The Red Cedar River arises in Cedar Lake in the south central portion of the lower peninsula of Michigan and flows about 45 miles to its confluence with the Grand River in the city of Lansing. The Grand River empties into lake Michigan, which connects to the other great lakes flowing into the Atlantic Ocean. The Red Cedar River has 12 tributaries and drains, and provides mid-Michigan residents with numerous recreational opportunities. The river also serves as a source of water for the irrigation of crops throughout the watershed. Swine and dairy are the predominant livestock in this watershed (figure 2-2).

The study was conducted into two parts. Part one consisted of determining antimicrobial resistance profiles of *E. coli* isolated from fecal samples of animals and humans. The second part consisted of determining antimicrobial resistance in *E. coli* isolated from surface water.

Figure 2-1: Michigan State map showing the location of Red Cedar Watershed

Figure 2-2: The Red Cedar Watershed, Michigan, USA (green area) map showing the farms location (yellow dots) and the water sampling sites (blue dots)



*Arrows in figure 2-1 and figure 2-2 indicate north

Part I: Antimicrobial resistance profiling of *E. coli* from fecal specimens

Study population:

Enrollment of participating farmers: Farmers in the Red Cedar Watershed were sent a letter through Ingham and Livingston Counties extension agents, asking whether they would like to participate in the study, and asking them to indicate their interest in participating by returning a pre-stamped post card to the Population Medicine Center at Michigan State University.

A total of 31 farmers were contacted and farm visits were arranged through the winter, spring, summer, and fall starting in the winter of 2002 until the winter of 2003.

Sample size: In order to detect at least one animal with *E. coli* on each farm, the general formula used by (Smith, 1995) was used as shown below.

$$n_{inf} = [\log\alpha]/[\log (1\text{-Prevalence of } E.\ coli)]$$

Where n_{inf} = Sample size for a very large population.

α = Probability of Type I error (0.05).

We assumed the prevalence of *E. coli* to be 10%. Using the above equation and assumptions stated; it was calculated that 29 animals per species would be the minimum to be tested.

Data collection

Data relating to antimicrobial use and number of animal species on the farm were collected during the time of collection of fecal samples. The data relating to

antimicrobial use were collected via a question asking about any antimicrobial agents used for therapy, prevention or growth promotion during the previous 60 days.

Sample collection:

Animal fecal samples: Fecal samples were obtained from dairy and beef cattle, swine, horses, sheep, goats, chicken, cats, dogs, farmed deer, ducks and wild geese. Fecal samples from livestock (dairy cattle, beef cattle, swine, sheep and goats) were collected directly from the rectum using the culturette swab system. However, due to difficulty in capturing feedlot cattle on some farms, the fecal samples were collected from the fresh drops using the culturette swab system. Fecal samples from horses, dogs and cats were collected by rectal swabbing, using the culturette swab system. Chicken fecal samples were collected by cloacal swab. Farmed deer and some wild geese fecal samples were collected from the fresh drops on the ground. Also, some samples were collected directly from the wild birds (ducks and some Canadian geese) during the geese-banding season conducted by Department of Natural Resources (DNR) in the study area. All deer fecal samples were collected from farmed deer, and not a free-range deer. Samples from the manure storage facilities (i.e. lagoon, slurry pit, and manure pile) on the farms were collected using the culturette swab system. Every swab was labeled with the farm number and given a serial number from the farm it was collected and shipped overnight to the lab at the University of Maryland for isolation and identification of *E. coli*. A total number of 2,292 fecal samples were collected.

Human septage samples: Samples representative of human fecal material were collected with the help of the local septic pumping companies in the study area. These companies were asked to provide samples of human septage material pumped from the septic tanks

from homes in the Red Cedar River. It was determined that septic samples would provide material most likely to affect water quality via leakage from septic tanks or improper disposal of pumped septic contents in the study area. A total of 34 human fecal samples were collected from the study area.

Procedure for shipment of fecal samples to the laboratory

It was important to make sure that all culturette tubes were labeled properly and legibly. Each individual culturette tube was sealed using clear packaging tape. Every 10 to 15 culturette tubes were bundle together, wrapped with absorbent cotton and placed in a screw-top cardboard tube mailer. The cardboard tube was taped using clear packaging tape. The Maryland laboratory address was placed on the screw-top cardboard tube mailer and covered with clear tape. The cardboard tube mailer was placed in a small cardboard box with Styrofoam packing peanuts to cushion it. A copy of the sample information was placed in a Ziploc bag on top of the peanuts in the box. The box was taped and another address label was placed on the outside of the cardboard box and shipped overnight by FedEx.

Isolation and identification of *E. coli* from fecal samples

Standard methods were used for the enrichment, isolation, identification, and biochemical confirmation of *E. coli* isolates (Ameican Public Health Association, 1998). The culturettes containing the sample were placed in tubes of Tryptic Soy Broth (TSB) and incubated at 35°C for 24 hours. This is an enrichment step to stimulate bacterial growth. A loop (10μls) of the turbid broth was transferred to a plate of Violet Red Bile

(VRB) media and streaked across the plates. The plates were incubated for 18 to 20 hours at 35°C. The VRB plates were examined for reddish purple colonies that fluoresce under a black light. Five or six colonies were then selected from each plate and streaked on a plate containing MacConkey's media. This plate was incubated at 35°C for 18 to 20 hours. The MacConkey plate was examined for pink colonies that precipitate bile and have a dark red center. One or two colonies were selected and streaked on a plate containing Tryptic Soy Agar (TSA). The purpose of this plate was to make sure that there were no transparent colonies being transferred from the MacConkey plate. This plate was streaked and incubated for 18 hours. The TSA plate was then examined for single colonies that are round, milk-colored, and slightly convex. One single colony was selected and placed in a tube containing TSB and incubated for approximately 3-4 hours until turbid (Difco, 1998). The broth was transferred into tubes for the biochemical confirmation: indole, methyl red, Voges Proskauer, and Simmons Citrate (IMVIC) test (American Public Health Association, 1998). Also, a test for sugar metabolism and gas production was performed using a tube of Triple Sugar Iron (TSI) (Difco, 1998). The results of the biochemical testing determine if the colony in the TSB tube is *E. coli.* The broth must pass all the biochemical tests to be used for the antimicrobial susceptibility testing. Once the colony of *E. coli* was confirmed then another tube of TSB was inoculated with the broth and incubated until it reached a turbidity of 0.5 MacFarland Standard (usually about 2-3 hours). That broth was used for the Kirby-Bauer procedure.

Part II: Antimicrobial resistance profiling of *E. coli* from water samples.

Sample collection: Water-sampling sites were determined with the help of the Ingham county drain commissioner based on the direction of the rain flow from every single farm enrolled in the study in order to get an accurate reading results (figure 2-2). The sampler (figure 2-3) that holds a 100ml sterile plastic bottle was used to collect surface water samples. These bottles contain 10mg sodium thiosulfate to neutralize any residual chlorine in water. The water bottles were labeled with the sampling site, date, time of collection and placed in a cooler and sent to the lab within 6 hours of collection for isolation of *E. coli*. A total of 37 water samples were collected during the study period.

Figure 2-3: The sampler used to collect water samples

Isolation and identification of *E. coli* from water samples

The Membrane filtration method was used to isolate *E. coli* from water samples. The original mTEC agar enumeration method (Dufour et al., 1981) for *E. coli* was used to provide a direct count of *E. coli* in water based on the development of colonies that grow

on the surface of a membrane filter. In this procedure, water samples were filtered through the membrane, sterile, white, grid marked 47-mm diameter, with 0.45±0.02μm pore size, which retains the bacteria. After filtration, the membrane containing the bacteria was placed on a selective and differential medium, in this case the mTEC (Dufour et al., 1981; United State Environmental Protection Agency, 1985), and incubated at 35±0.5°C for 2 hrs to resuscitate the injured or stressed bacteria, and then incubated again at 44.5±0.2°C for 22 hrs. The filter was transferred from mTEC agar to a filter pad saturated with Urea Substrate Medium. After 15 to 20 minutes, yellow, yellow-green, or yellow-brown colonies on mTEC were counted with the aid of a fluorescent lamp and a glass lens (2–5x magnification) or stereoscopic microscope and used for antimicrobial susceptibility testing. Non-*E. coli* colonies turned pink or purple on the Urea Substrate Medium. The number of *E. coli* per 100 mL was calculated according to the following general formula:

$$\textit{E. coli}/100\text{ mL} = \frac{100\text{ (number of }\textit{E. coli}\text{ colonies counted)}}{\text{(Volume of sample filtered, in mL)}}$$

The results were reported as *E. coli* / 100ml colony forming units (CFUs).

Antimicrobial susceptibility testing

Antimicrobial agents used for testing E. coli isolates

The following twelve antimicrobial agents were included in the tests: neomycin, gentamicin, Streptomycin, chloramphenicol, ofloxacin, trimethoprim/sulfamethoxazole, tetracycline, ampicillin, nalidixic acid, nitrofurantoin, cephalothin, and sulfisoxazole. These antimicrobials were chosen on the basis of their importance in treating human or

animal *E. coli* infections, or used as feed additive to promote growth in animals, and to provide diverse representation of antimicrobial classes. Additionally, two or more antimicrobial agents from the same class were used to see if there is any cross-resistance between them. We were also interested in seeing whether there would be any resistance to the banned antimicrobial agents such as Chloramphenicol, which are restricted from use in food animals (Krumperman, 1983)

Kirby-Bauer Method

The standard Kirby-Bauer disc diffusion method was used to develop the antimicrobial sensitivity profile of *E. coli* isoltes (National Committee for Clinical Laboratory Standards (NCCLS), 1997, 1999) for 12 antimicrobial agants (Table 2-1).

A 150 mm plate containing Mueller-Hinton media was swabbed with tryptic soy broth that reached a turbidity of 0.5 MacFarland Standard and allowed to dry for 1 minute. The Mueller-Hinton agar was formulated to have a pH between 7.2 and 7.4. The 12 commercially prepared, and standardized antibiotics discs, were dispensed onto the inoculated plate and tapped lightly to adhere to the media surface. The plate was incubated at 35°C for 18 to 20 hours. The diameter of the clear zones of growth inhibition around the antibiotic discs including the 6-mm disc diameter was measured in mm using precision calipers (NCCLS, 1997; 1999). The principle of this test is that when an antibiotic impregnated disc is placed on an *E. coli* growth, it will create a clearing zone around the disc where the bacteria can't grow. The size of this zone of inhibition depends on the sensitivity of the *E. coli* to the specific antibiotic (figure 2-4). This test was used for its flexibility in type and number of drugs that can be tested, and because of its low cost.

The breakpoints. Breakpoints are the zones of inhibition at which an organism is considered to be susceptible or resistant based on obtainable serum concentrations of the drug and clinical trail. In other words, when a lab reports that an organism is susceptible it implies the recommended dosage of the antimicrobial agent that will reach serum or tissue concentration and sufficient to inhibit the bacterium's growth in vivo (Prescott, et al., 2000). The breakpoints used to categorize isolates as resistant or not resistant for each antimicrobial agent for *E. coli* were those recommended by the National Antimicrobial Resistance Monitoring System (NARMS) (Table 2-1). The American Type Culture Collection (ATCC) 25922 *E. coli* was used to evaluate the performance of in vitro susceptibility tests.

Figure 2-4: Disc diffusion test plate showing the zones of growth inhibition of *E. coli* around an antimicrobial agent.

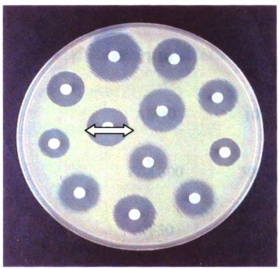


Table 2-1: Concentrations and diffusion zone breakpoints for resistance for antimicrobial agents tested in this study, by class of antimicrobial agent

Class of Agent Antimicrobial Agent	Drug Code	Disk Drug Concentration (μg)	Diffusion Zone Breakpoint* (mm)
Aminoglycosides			
Neomycin	N30	30 μg	≤ 12 mm
Gentamicin	GM10	10 μg	≤ 12 mm
Streptomycin	S10	10 μg	≤ 11 mm
Phinicols			
Chloramphenicol	C30	30 μg	≤ 12 mm
Quinolones and Fluoroquinolones			
Ofloxacin	OFX5	5 μg	≤ 12 mm
Nalidixic Acid	NA30	30 μg	≤ 13 mm
Sulfonamides and potentiated sulfuonamides			
Trimethoprim/Sulfamethoxazole	STX	23.75 μg / 1.25 μg	≤ 10 mm
Sulfisoxazole	G.25	250 μg	≤ 12 mm
Tetracyclines			
Tetracyline	TE30	30 μg	≤ 14 mm
Penicilllins			
Ampicillin	AM10	10 μg	≤ 13 mm
Nitrofurans			
Nitrofurantoin	F/M 300	300 μg	≤ 14 mm
Cephalosporins			
Cephalothin	CF30	30 μg	≤ 14 mm

* The zones of bacterial growth inhibition less than or equal the break point value indicate the bacteria was resistant, and the zone value of 6mm indicates no zone of resistance

Data Analysis

All the data were entered into computer spreadsheets (Access and Excel; Microsoft), which were used to generate descriptive statistics including graphs and charts of antimicrobial resistance. The prevalence of antimicrobial resistance was calculated as the number of samples yielding *E. coli* with resistance to a given antimicrobial agent divided by total number of samples tested.

To test the stated hypothesis, Fisher's exact test was used for data analysis (SAS ver 8.2. SAS Inc. Cary, NC).

Results

A total of 31 farms agreed to participate in the study and included 7 dairy farms, 7 beef farms, 6 sheep farms, 5 pigs farm, 2 horse farms, 2 chicken farms, and 2 deer farms. However, some farms contained more than one species, for example, sheep and horses, sheep and chicken, pigs and chicken, sheep and beef cattle, beef cattle and horses. The antimicrobial agents mentioned as being used in the different species on farms are shown in Table 2-2. No data were collected regarding antimicrobial used by humans from whom septic tank samples were collected.

Table 2-2: Antimicrobial agents reported as having been used on the farms in the different species of animals sampled

1. **Dairy cattle**
   - Quarter master (penicillin)
   - Cefa-lak (cephapirin sodium)
   - Excenel
   - Naxcel (ceftiofur sodium)
   - Aureo S 700 (Chlorotetracycline and sulfamethazine)
   - Tetracycline
   - Oxytetracycline
   - Trimethoprim-sulfamethoxazole
2. **Beef cattle**
   - Excenel
   - Micotil
   - Baytril (enrofloxacin)
   - Aureo S 700 (Chlorotetracycline and sulfamethazine)
   - Tetracycline
   - Oxytetracycline
3. **Sheep**
   - OR-E-O Krumbs A (chlorotetracycline)
4. **Goat**
   - Penicillin, injection
   - Tetracycline, injection
   - Oxytetracycline, injection
5. **Pigs**
   - Penicillin
   - OR-E-O Krumbs A (chlorotetracycline)
   - Bacitracin Methylene Disalicylate
   - Tylan (Tylosin)
   - Lincomix (Lincomycin)
   - Pulmotil (Tilmicosin)
   - Tetracycline
   - CTC-8 (chlorotetracycline)
6. **Horses**
   - Sulfa drug
7. **Poultry:** data not provided
8. **Companion animals (dogs and cats):** data not provided

Source of *E. coli* isolation versus antimicrobial resistance

To answer the question whether samples from farmed deer, wild geese, water, human and different domestic species of animal have the same or different antimicrobial resistance profiles, the prevalences of such profiles were compared. (Please notice that images in this thesis are presented in color). In general, the antimicrobial resistance profiles of *E. coli* isolates from cattle, sheep and swine were largely similar. Resistance to antimicrobial agents was found in *E. coli* isolates from all animal species, human, and water samples. In cattle (n=417; figure 2-5) resistance was found to all antimicrobial agents tested. The most prevalent forms of resistance seen were to cephalothin, tetracycline, sulfisoxazole and streptomycin. Minute resistance (less than 1%) to gentamicin, nitrofurantoin, nalidixic acid and ofloxacin was seen. In sheep (n=156; figure 2-7) resistance was found to ten of twelve antimicrobial agents tested. The most prevalent forms of resistance seen were to cephalothin, tetracycline, streptomycin, and sulfisoxazole. No resistance to ofloxacin and gentamicin was noted. In pigs (n=176; figure 2-9), resistance was found to ten of twelve antimicrobial agents, the most prevalent forms of resistance seen were particularly in regards to tetracycline, sulfisoxazole, streptomycin, ampicillin and cephalothin; no resistance to nalidixic acid and nitrofurantoin was seen. In horses (n=58; figure 2-11) resistance was found to seven of twelve antimicrobial agents tested. The most prevalent resistance seen was to cephalothin, sulfisoxazole, trimethoprim sulfamethoxazole, tetracycline, streptomycin, and ampicillin, while no resistance was observed to neomycin, chloramphenicol, ofloxacin, nalidixic acid and nitrofurantoin was seen. In poultry (n=87; figure 2-13)

resistance was found to eleven of twelve antimicrobial agents tested. The most prevalent forms of resistance seen were to tetracycline, cephalothin, sulfisoxazole, streptomycin, and neomycin. No resistance to ofloxacin was seen.

In companion animals n=23 (17dogs, 6 cats; figure 2-15) resistance was found to four of twelve antimicrobial agents tested; cephalothin, streptomycin, tetracycline, and ampicillin. No resistance was shown for the rest antimicrobial agents. In farmed deer (n=34; figure 2-16) and wild geese (n= 54; figure 2-16) there was resistance only to cephalothin in farmed deer and resistance to cephalothin and tetracycline in wild geese; no resistance was recorded for the other antimicrobial agents. In humans (n=3; figure 2-17) resistance was found to three of twelve antimicrobial agents tested; cephalothin, tetracycline, and streptomycin, whereas no resistance to the rest antimicrobial agents was observed. In water samples (n=26; figure 2-18) all the *E. coli* isolates were resistant only to cephalothin.

E. coli isolated from environmental samples

To answer the question whether *E. coli* isolated from the environmental samples of domestic animals have the same or different antimicrobial resistance profile than that from domestic animals, the prevalence of such profiles was compared. In general, the resistance profiles in isolates from feces collected directly from domestic animals and from the environment on the farm were similar but with slight variation among species. Most *E. coli* isolates showed resistance to cephalothin, tetracycline, streptomycin and sulfisoxazole. In *E. coli* isolated from cattle farm (n=118; figure 2-6) the most prevalent forms of resistance seen were to cephalothin, tetracycline, streptomycin and sulfisoxazole, low resistance was seen to neomycin, ampicillin, chloramphenicol,

trimethoprim-sulfamethoxazole and nitrofurantoin. No resistance to gentamicin, ofloxacin, and nalidixic acid was observed. In sheep farms (n=31; figure 2-8) most *E. coli* isolates showed resistance to cephalothin, tetracycline, sulfisoxazole, streptomycin and trimethoprim- sulfamethoxazole, low resistance was seen for gentamicin, chloramphenicol, and no resistance was seen to neomycin, ofloxacin, ampicillin, nalidixic acid, and nitrofurantoin. The *E. coli* isolated from pigs farms (n=38; figure 2-10) showed the highest resistance to tetracycline, compared to other animal farms, then to cephalothin, sulfisoxazole, streptomycin, ampicillin and neomycin. Small resistance to gentamicin, chloramphenicol was seen. No resistance to ofloxacin, trimethoprim sulfamethoxazole, nalidixic acid or nitrofurantoin was observed. In horse farms (n=16; figure 2-12) the high resistance was seen to cephalothin, tetracycline, streptomycin, sulfisoxazole, and trimethoprim sulfamethoxazole. Slight resistance was seen to gentamicin and ampicillin, no resistance to neomycin, chloramphenicol, ofloxacin, nalidixic acid or nitrofurantoin was seen. Regarding poultry farms (n=19; figure 2-14) the highest resistance to tetracycline was seen then cephalothin, and sulfisoxazole and streptomycin and neomycin. No resistance was seen to the rest of antimicrobial agents. Over all, the resistance of *E. coli* isolated from the environment was higher to cephalothin than that isolated directly from animals.

Source of *E. coli* isolation under same condition versus antimicrobial resistance

To answer the question whether animals under same condition in the same farm have the same antimicrobial profiles, the following results answer the question.

Antimicrobial resistance of *E. coli* from a farm with pigs and chickens

By comparing the antimicrobial resistance profile for pigs and chickens from the same farm we found that they have similar antimicrobial resistance profile except that of chickens, which do not show any resistance to ampicillin. Both, however, showed high resistance to tetracycline (figure 2-19).

Antimicrobial resistance of *E. coli* from a farm with cattle, horses and sheep

Under this farm condition the *E. coli* isolated from horse and sheep showed high resistance to cephalothin only but the *E. coli* isolated from cattle showed resistance not only to cephalothin but also to tetracycline, streptomycin, and sulfisoxazole (figure 2-20)

Antimicrobial resistance of *E. coli* isolates from chicken and food animals living on the same farm

Over all, the resistance patterns were similar but the resistance to tetracycline in *E. coli* isolated from food animals (n=132) was higher than that for *E. coli* isolated from chicken (n=52). On the other hand, the resistance among food animal isolates to cephalothin and streptomycin was less than that for *E. coli* isolated from chicken isolates. The resistance to sulfisoxazole among food animal isolates was higher than that for chicken isolates. Low resistance was observed to the following antimicrobial agents

among food animals and chicken isolates: neomycin, ampicillin and chloramphenicol (Figure 2-21).

Antimicrobial resistance of *E. coli* isolates from companion animals and food animals living on the same farm

The *E. coli* isolated from companion animals (n=8) and food animals (253) living on the same farm showed resistance to cephalothin, tetracycline, streptomycin and ampicillin but the level of resistance to these antimicrobial agents in food animals is less than that in companion animals (Figure 2-22).

Antimicrobial resistance of *E. coli* isolates from companion animals and horses living on the same farm

The prevalence of resistance among companion animals' isolates, dogs and cats (n=15), and horse (n=12) under the same farm was the same. Both were resistance to cephalothin only and susceptible to the rest antimicrobial agents (figure 2-23).

Antimicrobial resistance of *E. coli* isolates from horses and food animals living on the same farm

Both isolates from horses (n=21) and food animals (n=162) on same farm were resistant to cephalothin only (figure 2-24).

Multi-drug resistant *E. coli* isolates

When comparing the *E. coli* isolates from domestic animals, wildlife (farmed deer and wild geese), human, water and farms and the number of antimicrobial agents that

resistant to, we found that 25.4% of *E. coli* isolated from domestic animals was resistant to one antimicrobial agent, 10.0% resistant to two antimicrobial agents, 7.0% to three antimicrobial agents. In wildlife (farmed deer and wild geese) 13.3% resistant to one antimicrobial agent, 33.3% of human *E. coli* isolates were resistant one and two antimicrobial agents, 61.5% of *E. coli* from water samples were resistant to one antimicrobial agent. According to *E. coli* isolated from farms 29.4% resistant to one antimicrobial agent, 13.3% resistant to two antimicrobial agents, 9.7% resistant to three antimicrobial agents, and 5.5% to four antimicrobial agents (figure 2-25). Table 2-3 shows the percentage of resistant isolates for specific antimicrobial agents, by species exposure class.

Table 2-3: Percentage of resistant isolates for specific antimicrobial agents, by species exposure class

Agent	*Overall (n=1,041)*	Livestock (n=863)	Humans & Companion Animals (n=26)	Equines (n=60)	Wildlife (n=90)	Fisher's Exact P for all species within agent
Neomycin	*4.7*	5.7 ***	0	0	0 *	.0089
Gentamicin	*.8*	.7	0	3.3	0	.2021
Streptomycin	*13.1*	14.7 ***	11.5	10.0	0 ***	< .0001
Chloramphenicol	*1.1*	1.3	0	0	0	.8510
Ofloxacin	*.2*	.2	0	0	0	1.0
Trimethoprim Sulfamethoxazole	*2.2*	1.7 *	0	13.3 ***	0	< .0001
Tetracycline	*30.7*	35.7 ***	11.5 *	10.0 ***	3.3 ***	< .0001
Ampicillin	*6.5*	7.2	3.8	8.3	0 **	.0146
Nalidixic Acid	*.8*	.9	0	0	0	1.0
Nitrofurantoin	*.9*	1.0	0	0	0	1.0
Cephalothin	*21.2*	21.8	34.6	21.7	10.0 **	.0142
Sulfisoxazole	*14.0*	16.0 ***	0 *	13.3	0 ***	< .0001

* - Fisher's Exact P < .05 for individual species exposure class within agent

** - Fisher's Exact P < .01 for individual species exposure class within agent

*** - Fisher's Exact P < .001 for individual species exposure class within agent

Table 2-4 shows the percentage of multi-drug resistance isolates by species exposure class while table 2-5 shows the common combinations of antimicrobial agents in multi-drug resistant isolates.

Table 2-4: Percentage of multi-drug resistant isolates by species exposure class

# Agents Resistant	Overall (n=1,041)	Livestock (n=863)	Humans & Companion Animals (n=26)	Equines (n=60)	Wildlife (n=90)	Fisher's Exact P for all species within # agents
0	*53.2*	48.6	61.5	68.33	86.67	< .0001
1	*25.0*	26.7	23.1	16.67	13.33	.0142
2	*8.5*	9.7	11.5	1.67	0	< .0001
3 or more	*13.4*	15.1	3.9	13.33	0	< .0001
Mantel Haensel X^2 test for trend within class	-	39.58	1.40	2.74	41.43	
X^2 p	-	< .0001	.2366	.0980	< .0001	
Fisher's Exact p	-	< .0001	.4615	.0417	< .0001	

Table 2-5: Most commonly identified combinations of agents in multi-drug resistant isolates from food animals

Agents	Percent of All isolates [a]	Percent of Multi-resistant isolates [b]
Tetracycline – Sulfamethazine	14.1	53.9
Tetracycline – Streptomycin	12.7	48.4
Streptomycin – Sulfamethazine	10.0	38.1
Tetracycline – Cephalothin	7.2	27.5
Tetracycline – Ampicillin	6.5	24.9
Tetracycline – sulfamethazine – Streptomycin *	7.3	27.8

* Includes numbers from tetracycline-sulfamethazine, tetracycline-streptomycin, and streptomycin-sulfamethazine.

To illustrate the calculation of [a] and [b], tetracycline-sulfamethazine used as an example;

[a] = Number of isolates resistant to tetracycline-sulfamethazine / All isolates tested

[b] = Number of isolates resistant to tetracycline-sulfamethazine / Number of multiresistant isolates

Within the 4 population subgroups, the livestock group including cattle, sheep, goats, poultry and swine manifested the highest rate of antimicrobial resistance for all but 4 of the individual antimicrobials tested (gentamicin 0.7%, sulfamethoxazole 1.7%, ampicillin 7.2%, and cephalothin 21.8%). Additionally, the livestock subpopulation had the highest rate of resistance to 3 or more antimicrobials (13.5%). The rate of multi-drug resistance by subpopulation group was also statistically significant (Fisher's exact $p < 0.0001$).

Discussion

This study found antimicrobial resistance to a variety of agents in *E. coli* isolates from food animals, companion animals, farmed deer and wild geese, human and water samples in Michigan, United States. This study was not designed to determine the relationship between antimicrobial use in animals on farms and the prevalence of patterns of antimicrobial resistance. It is, however, noteworthy to mention that the majority of antimicrobials for which resistance was demonstrated, were reported as being used in different animal species on farms. Additionally, resistance in *E. coli* isolates from cattle, pigs and sheep was reported to antimicrobial agents such as chloramphenicol that was banned from use in these animals. Such resistance may be due to persistence of resistant *E. coli* on farms for a long time. The differences in the prevalence of resistance between species may be due to different exposures to antimicrobial growth promoters and the time that the animals were exposed.

The biological plausibility that antimicrobial use in animal feed and the selection pressure for bacteria with resistance to the exposed agents was observed in many bacteria (Witte et al., 2000). The resistance to tetracyclines, especially chlorotetracycline in swine and poultry industries, may be due to the widespread use of this antimicrobial agent in animals as feed additive, prophylactic or used clinically to treat diseases. Smith, (1975) showed that feeding pigs rations contain tetracyclines resulted in the recovery of large number of tetracycline- resistant *E. coli* from their feces. Enterotoxigenic *E. coli* usually is resistant to tetracycline in swine, poultry and cattle since this drug and chlorotetracycline has been used widely in medicated feed as growth promoters (Prescott et al., 2000). Tetracyclines especially oxytetracycline were frequently used in animal

ration as feed additives in cattle, and sheep yet a substantial number of cattle and sheep isolates were tetracycline resistant. The relationship between the degree of antimicrobial use and the extent of resistance was explained by Linton (1977). Acquired resistance is wide spread in tetracyclines and it is usually plasmid mediated that interfere with the active transport of tetracycline into the cell and increased efflux from the cell (Prescott et al., 2000).

Antimicrobial use in human facilitates resistance development in many pathogenic bacteria (Neu, 1992). The *E. coli* isolated from human septic tank showed resistance to tetracycline as a results of therapeutic or prophylactic application.

We found low resistance to tetracycline in *E. coli* isolated from horses' fecal samples compared to other animals in our study. It could be due to the little use of tetracycline in horses, since it has high potential for toxic effects, including local necrosis after intramuscular injection, or its broad spectrum suppression of intestinal microflora that allow for super infection with resistant salmonella and clostridia that result in severe and lethal diarrheic illness (Prescott et al. 2000).

High percentages of *E. coli* isolates were resistant to cephalothin, which is a first generation cephalosporin beta-lactam antibiotic. The clinical applications of first generation cephalosporins have decreased with the development of beta lactamase stable cephalosporins. Oral cephalosporin has been used extensively in small animal medicine for treating skin and urinary tract infections especially cephalexin the drug of choice for *Klebsiella pneumoniae* infections or as prophylaxis of surgical wound infections. It is used for the same reasons in human. This explains the high resistance to this drug among companion animals and humans in our study (Figure 2-15 and Figure 2-17).

During our visits to farms, we collected information about the antimicrobial agents that the farmers used to treat their animals or given to their animals as a feed additive. We found that most of the dairy farmers used Cefa-Lak that contains cephapirin sodium, which is a first generation-group one cephalosporin as an intramammary infusion in dairy cows to treat bovine mastitis. The other drug that has been used widely by veterinarians is Naxcel. Naxcel which contains ceftiofur sodium, third generation-group four cephalosporin, to treat systemic infections especially respiratory disease caused by gram-negative aerobes such as *E. coli*, *Salmonella*, and *Pasteurella* in cattle, swine, sheep and horse. It is also been used to treat urinary tract infections associated with *E. coli* and *Proteous mirabilis* in dogs. In poultry it has been used to control *E. coli* infections. The use of first and third generations of cephalosporin may explain the high resistance to cephalothin among these animals in our study (Figures 2-5, 2-7, 2-11, 2-13, 2-15).

Thirty three percent of *E. coli* isolated from human septic tank was resistant to cephalothin. The use of first, third and fourth generation cephalosporins to treat skin and urinary tract infections, in addition to respiratory tract infections and meningitis in humans, may explain this resistance (Figure 2-17).

For susceptibility testing, cephalothin is the class drug to use. There are three mechanisms of resistance to cephalosporins: first, reduced permeability, second, enzymatic inactivation and third, absence of specific penicillin binding proteins. In our study we noticed the high resistance to cephalothin, tetracycline, and streptomycin and this could be due to the cross-resistance between these antimicrobial agents. We noticed that some cephalosporin mutanted with altered outer membrane permeability that involves the extra cellular expression of beta lactamases through efflux pumps, may show

cross resistance with aminoglycosides, chloramphenicol, fluoroquinolones, tetracyclines, and trimethoprim (Prescott et al., 2000). Cross resistance means when one organism becoming resistant to one antibiotic thereby becomes resistant to another.

Transferable, broad-spectrum plasmid mediated resistance to beta-lactamase stable cephalosporin has increasingly been described; and consider a threat to the continued use of these cephalosporins (Prescott et al., 2000).

The persistence of resistant bacteria is related to the persistent use of the antimicrobial agent. The prolonged use of the antimicrobial agent is more likely associated with the persistence of resistant *E. coli*, but it may not be readily reversed by withdrawal (Langlois et al., 1983) this could explain the presence of small percentage of *E. coli* isolates that resist chloramphenicol, even after the a banned use of this drug in food animal in the USA. Chloramphenicol was banned because of its toxic effect in humans; it causes suppression to the bone marrow leading to aplastic anemia.

Interestingly, a very small percentage of *E. coli* isolates in our study were resistant to nalidixic acid, nitrofurantoin, and non resistant to ofloxacin. Because none of these drugs are approved for use in food animals in the United States, the small resistance observed may be due to the use of related approved bovine antimicrobial agents such as florfenicol and enrofloxcacin (White et al., 2000).

Recent studies have suggested the use of cephalosporins, tetracyclines, sulfa drugs, to be the major factor in the emergence and dissemination of antimicrobial resistance *E. coli* in animals (Meng et al., 1998; Galland et al., 2001; Schroeder et al., 2002a; Schroeder et al., 2002b). These studies were consistent with our findings that there was high

prevalence of resistance to cephalothin, tetracycline, and sulfisoxazole among *E. coli* isolates from different species of animals.

Streptomycin is part of the aminoglycosides class that selectively binds to kidney tissue. For that reason, the U.S. Food and Drug Administration do not approve it for use in food animals. Although, for a long time it was used extensively in combination with penicillin in food animals, to treat *Corynebacterium renale* and *Arcanobacterium pyogenes* and actinobacillosis. Its combination with tetracycline is effective against *Brucella abortus* in cattle and *Brucella melitensis* in sheep. Streptomycin used to control leptospirosis in swine, but in horses it has little implication because of wide spread of resistance (Prescott et al., 2000). In our study we found that *E. coli* isolated from different animal species was resistant to streptomycin. Acquired resistance to streptomycin is widespread in veterinary pathogens, and plasmid-mediated resistance is commonly linked with sulfonamide, ampicillin, and tetracycline resistance genes. We noticed that *E. coli* isolates that resistant to streptomycin were resistant also to tetracycline and sulfonamide (Prescott et al., 2000).

The high resistance of *E. coli* isolated from horses fecal samples to trimethoprim-sulfamethoxazole and sulfisoxazole due to the wide use of sulfa drugs to treat acute respiratory infections, acute urinary tract infections, wounds, and abscesses. Trimethoprim-sulfadiazine and sulfisoxazole are the drugs of choice to treat Salmonellosis, Actinobacillus in foals, and coliform meningitis. We also found resistance to trimethoprime sulfamethoxazole and sulfisoxazole in *E. coli* isolated from different animals species and this may due to the wide spread use of the sulfonamides in these species. For example, sulfadimethoxine is the only sulfonamide approved for use in

dairy cows over 20 months of age in the United States. Sulfonamides have been used with chlortetracycline in feedlot lambs to improve performance and prevent clostredial enterotoxemias. In addition, sulfonamide mixed with chlortetracycline to promote growth and control group E streptococcal infections, and atrophic rhinitis in pig. Because of the problem of residues in carcasses in excess of legally permitted amount, there have been moves to ban the use of sulfonamides in swine. In poultry sulfonamide has been used to treat coccidiosis, and infectious coryza.

Multiple antimicrobial resistance in *E. coli* isolates may results from the spread of genetic elements including plasmids, tranposons, and integrons (Jones et al., 1997), which may also confer resistance to numerous antimicrobials. The R factor or R resistance plasmid may code for resistance to between one and ten different antibiotics. The linkage of resistance genes on the same plasmid means that the use of any one antibiotic for which resistance was determined by the plasmid promotes resistance to all the antibiotics. Withdrawing the use of all antibiotics in a herd may not results in the loss of resistance by *E. coli* because such genes may be incorporated into bacterial chromosome (Prescott et al., 2000) and because possession of R plasmids may not be deleterious to bacterial survival (Smith et al., 1974).

Nonspecific resistance to a wide range of unrelated antibiotics associated with mutations, leads to over expression of multi-antibiotic resistance locus, which controls multidrug efflux pumps in bacteria that can be selected by low concentrations of an antimicrobial drug (Prescott et al., 2000)

It is plausible that bacteria with resistance to antimicrobial agents could be developed on farm and transmitted to human (Levy, 1987). The use of antimicrobial agents in food

animals may contribute to the antimicrobial resistance in human. For example, cattle are considered a symptomatic carrier for *E. coli*, when exposed to antimicrobial agents may serve as a reservoir of antimicrobial resistant *E. coli* (Barton, 1998; Witte, 1998). Since first and third generations cephalosporins were used in cattle in our study, the observation that human *E. coli* isolates resistant to cephalothin suggests the transfer of resistant *E. coli* from food animals to humans (Zhao et al., 2001).

Our study was not designed to determine the contribution of antimicrobial use in animals to the antibiotic resistance in humans. Rather, patterns of antimicrobial resistance were observed, which may form a data –based basis for hypothesis formulation.

Resistant bacteria may colonize the human population via food chain through many pathways: first, the resistant pathogenic bacteria could be selected in the animal's gut; it could contaminate the meat during animal slaughter or meat preparation. If a human ingested the contaminated meat, it would cause infection and the treatment will be compromised. Second, the non-pathogenic antibiotic resistant bacteria will be selected in the animal's gut, leading to contamination of food that, if ingested, the bacteria could transfer the resistance to other bacteria in the human gut. Third, residual remains of antibiotic in animal products, may allow the selection of antibiotic resistant bacteria in the consumer of the food (Piddock, 1996).

Resistant bacteria could also enter humans through occupational exposure; for example, farmers, slaughterhouse workers and food handlers are more likely to have resistant *E. coli* than the general population. Drinking or swimming in water contaminated with animal fecal materials that contain resistant bacteria especially *E. coli*

and the direct contact with animals, especially companion animals, are another ways for humans to be exposed to resistant bacteria.

Water becomes contaminated with resistant bacteria when waste run off from animal production facilities, leakage from human septic tank or sewage dumped in the river (Witte, 1998; Van den Bogaard and Stobberingh, 1999; Van den Bogaard et al., 2001). The *E. coli* strain responsible for 1989 waterborne outbreak in Missouri was resistant to streptomycin, sulfisoxazole, and tetracycline (Swerdlow et al., 1992).

Our results showed that *E. coli* isolates from domesticated animals showed resistance to more antimicrobials tested than those isolates from human septic tanks, wild geese, farmed deer or surface water. On the other hand, the level of resistance to cephalothin in *E. coli* isolated from water was closer to that observed in *E. coli* isolated from human septic tanks. These results suggested that surface water contamination by antimicrobial resistant *E. coli* is a complex issue in terms of determining the source of the problem. Even though *E. coli* from farm animals showed resistance to more antimicrobial agents (including cephalothin), the percentage of isolates that were resistant to cephalothin was much lower than that in isolates from human septic tanks or companion animals. However, a previous study with sludge and septic tank wastes showed relatively high levels of antimicrobial resistance in *E. coli* (Pillai et al., 1997).

We expected that *E. coli* isolated from wild geese would not show resistance to any of the antimicrobial agents tested, because wild birds have never been treated with antimicrobial agents. However, we found that 11% and 4% of *E. coli* isolates were resistant to cephalothin and tetracycline respectively. The access of wild birds to the medicated animal feed, and then developing the resistance in their gut, or the easy access

of wild birds to river water that may be contaminated with the resistant bacteria, may explain these results. Resistant *E. coli* could reach the water from a failed septic tank system of a nearby residence, or from runoff water over a wide field area that may have been spread with manure of animals harboring resistant bacteria, or the water could have the resistance factor from other bacteria due to the relative ease with which resistance factors are exchanged among promiscuous bacteria (LeClerc et al., 1996).

Surprisingly, the *E. coli* isolates from water were 81% resistant to cephalothin only (the highest in all *E. coli* isolates). The induction of cross-resistance by structurally related compounds might explain the presence of the corresponding resistance in the water system. Our results agree with the results of Mulamattathil et al., (2000) who found the high level of resistance to cephalothin in closed water reticulation system. He also found a correlation between the specific use of certain antimicrobial agents and the prevalence of the corresponding resistant bacterial isolates. Our results are similar to the results from a recent survey of U.S. rivers found that cefotaxime a third generation cephalosporin resistant gram negative bacterium to range from 16%-96% across 22 rivers (Ash et al., 2002).

Grabow et al., 1973; Grabow et al., (1975) suggested that R-factor mediated antimicrobial resistance increased the survival ability of coliform bacteria; on the other hand, (Anderson, 1974) suggested that R-factor mediated may reduce survival ability in *E. coli* but Smith et al., (1974) indicated that R-factor antimicrobial resistance had no effect on *E. coli* survival. Once multiresistance bacteria develop, they may persist in the host or the environment in the absence of antibiotic selection and may act as reservoirs of resistance genes that may spread to other bacteria (Prescott et al., 2000).

There were trends in patterns of antimicrobial resistance indicating common sources of resistance factors for different types of samples and for different species. First, similar patterns of resistance from fecal and environmental samples classified by animal species were similar, indicated common sources of resistant bacteria. It is possible that livestock function as a reservoir of resistant bacteria for environmental contamination, particularly in cases where higher levels of resistance were seen in fecal isolates compared to environmental isolates (e.g. tetracycline, sulfisoxazole). However, in cases where higher levels of resistance were seen in environmental samples (e.g. cephalothin, trimethoprim-sulfamethoxazole) it is also possible that contaminated environments serve as the major reservoirs for resistant bacteria for livestock. This does not support the once widely held theory that the presence and expression of resistance genes impairs viability and survival outside the host. It also suggests that, while collection of environmental samples from a farm may not be valid means of assessing the prevalence and distribution of antimicrobial resistance patterns in animals residing on the farm, it is a more accurate measure of exposure to resistance from farm runoff into watershed.

It was interesting and noteworthy to observe in our study that animal species that lived on the same farm had similar antimicrobial resistance profiles. Thus, food animal and chickens, living on the same farm as well as different combinations such as horses and companion animals, food animals and horses had similar antimicrobial resistance patterns, these observations suggest that bacteria which share a common environment also share a common mode for developing antimicrobial resistance (Kelch and Lee, 1978).

Because of logistical reasons, we were unable to collect samples in a manner that allowed us to evaluate whether there was seasonal variation in the frequency of *E. coli* in samples collected and hence antimicrobial resistance profile. This issue needs to be addressed in future research.

Figure 2-5: Antimicrobial resistance of *E. coli* from cattle

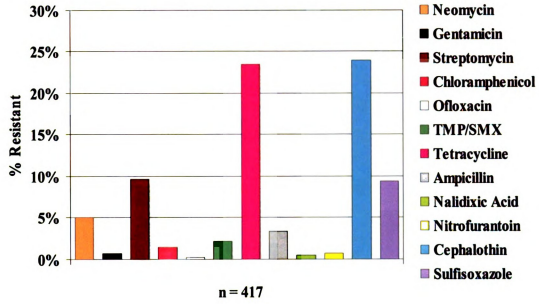


Figure 2-6: Antimicrobial resistance of *E. coli* from cattle farms

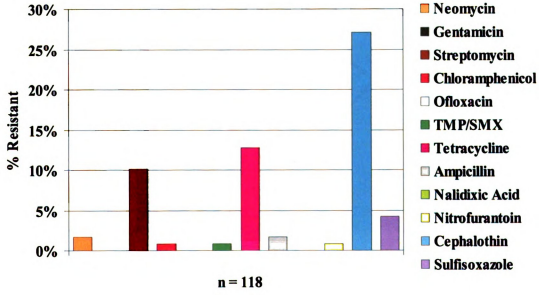


Figure 2-7: Antimicrobial resistance of *E. coli* from sheep

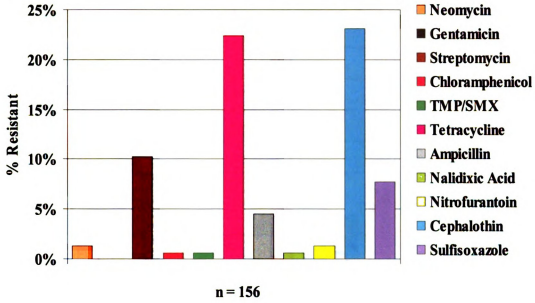


Figure 2-8: Antimicrobial resistance of *E. coli* from sheep farms

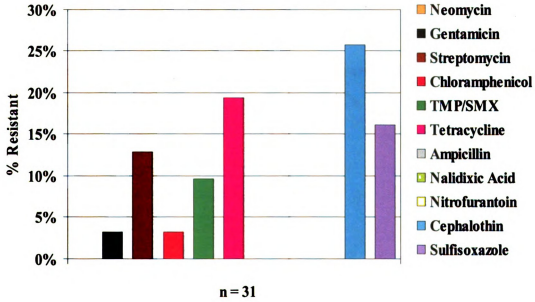


Figure 2-9: Antimicrobial resistance of *E. coli* from swine

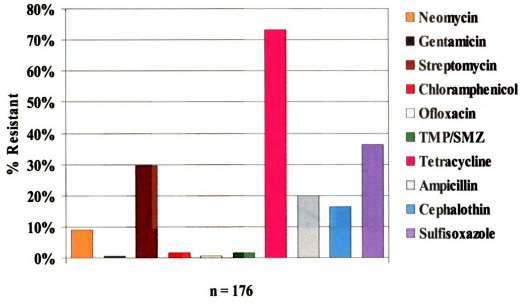


Figure 2-10: Antimicrobial resistance of *E. coli* from swine farms

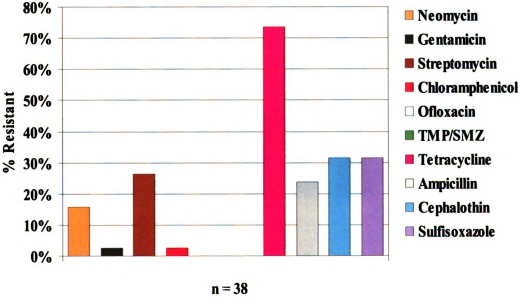


Figure 2-11: Antimicrobial resistance of *E. coli* from horses

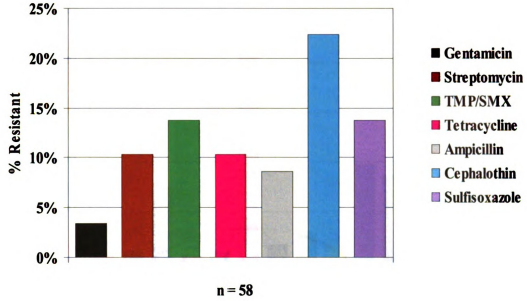


Figure 2-12: Antimicrobial resistance of *E. coli* from horse farms

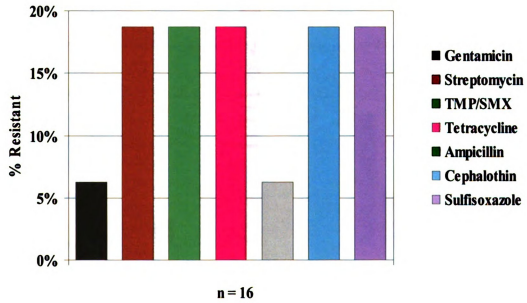


Figure 2-13: Antimicrobial resistance of *E. coli* from poultry

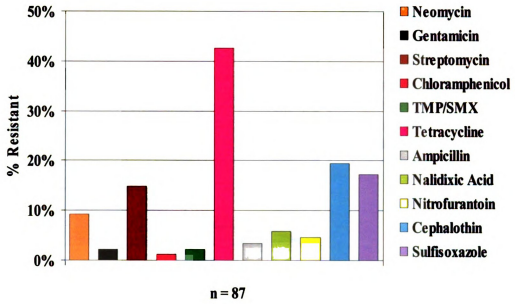


Figure 2-14: Antimicrobial resistance of *E. coli* from poultry farms

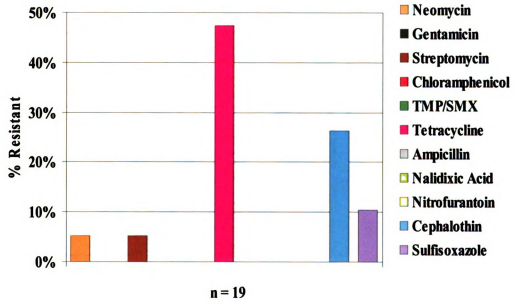


Figure 2-15: Antimicrobial resistance of *E. coli* from companion animals

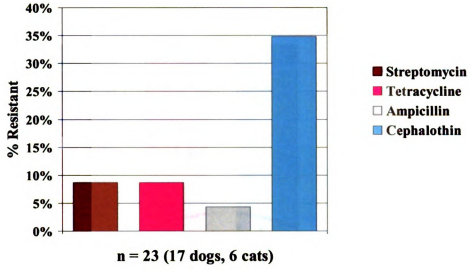


Figure 2-16: Antimicrobial resistance of *E. coli* from farmed deer and wild geese

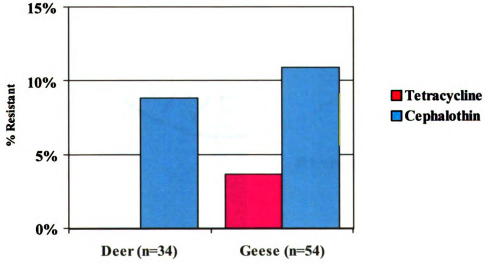


Figure 2-17: Antimicrobial resistance of *E. coli* from human septage

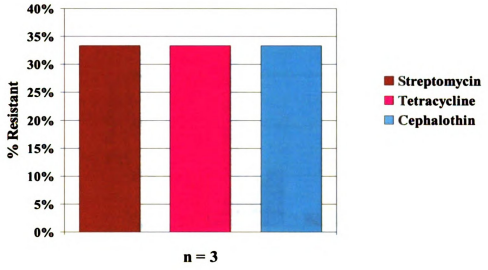


Figure 2-18: Antimicrobial resistance of *E. coli* from water

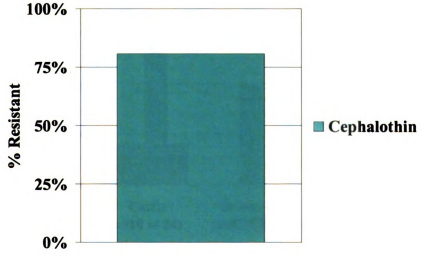


Figure 2-19: Antimicrobial resistance of *E. coli* from a farm with pigs and chickens

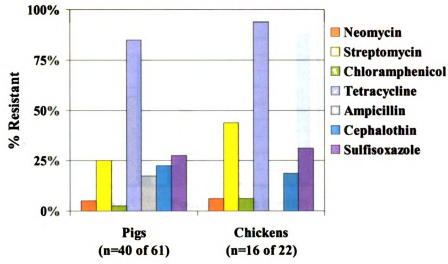


Figure 2-20: Antimicrobial resistance of *E. coli* from a farm with cattle, horses and sheep

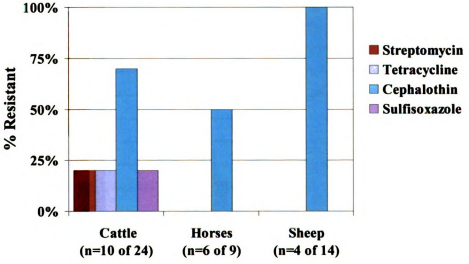


Figure 2-21: Antimicrobial resistance of *E. coli* from chicken and food animals

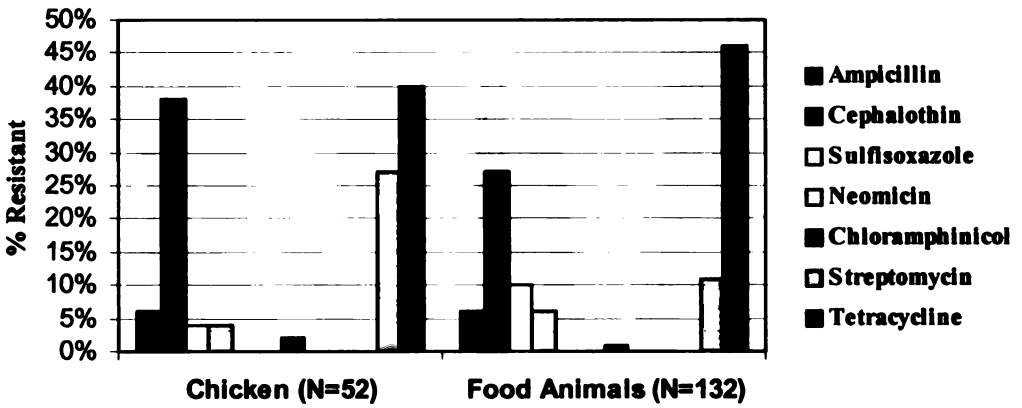


Figure 2-22: Antimicrobial resistance of *E. coli* from companion animals and food animals

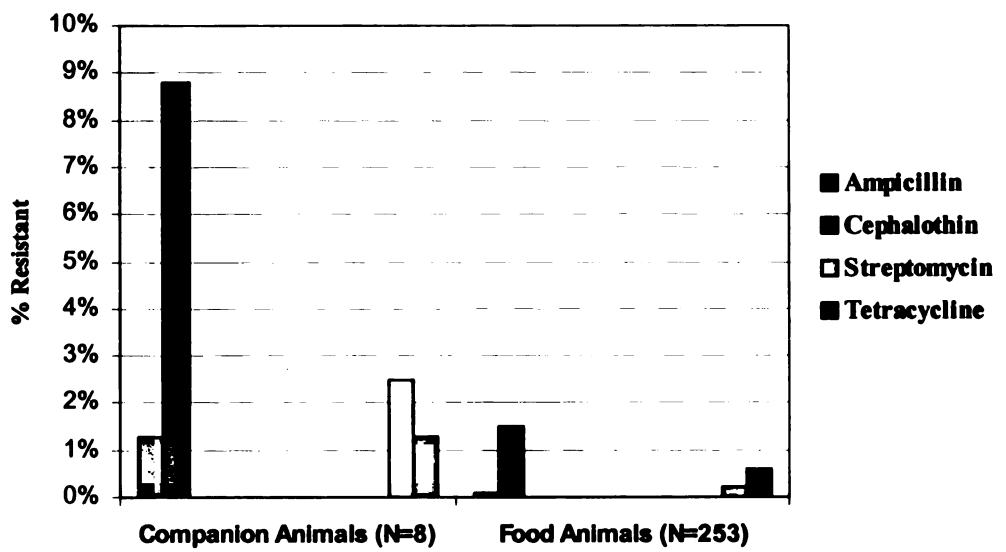


Figure 2-23: Antimicrobial resistance of *E. coli* from companion animals and horse

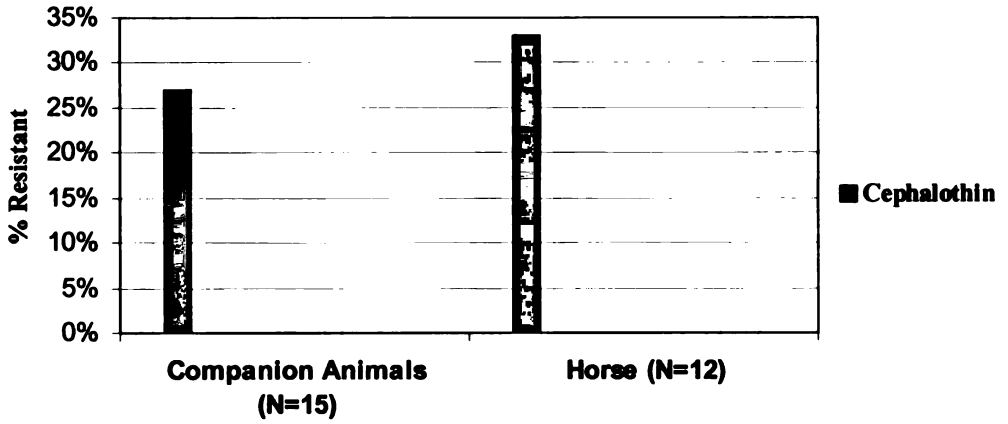


Figure 2-24: Antimicrobial resistance of *E. coli* from horse and food animals

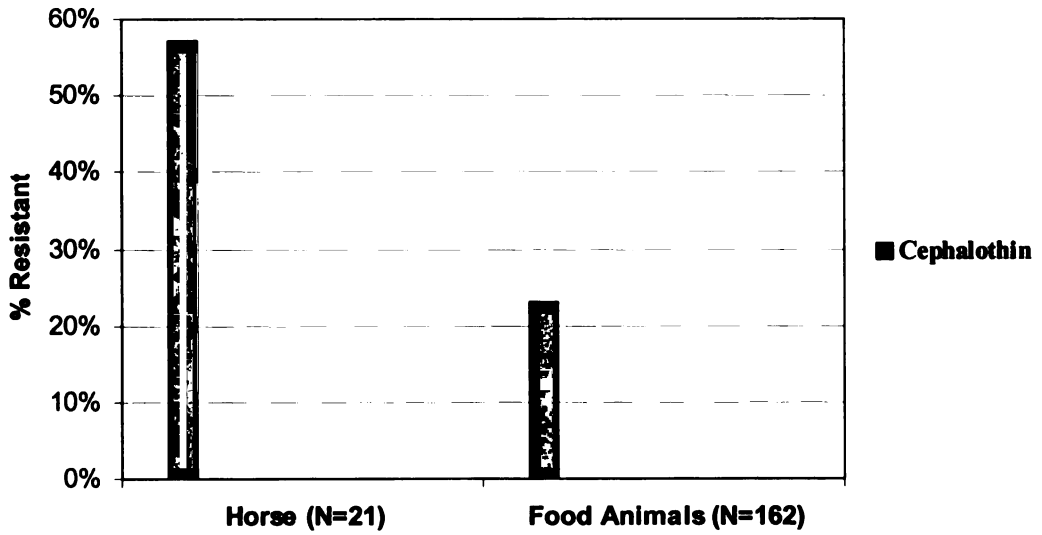
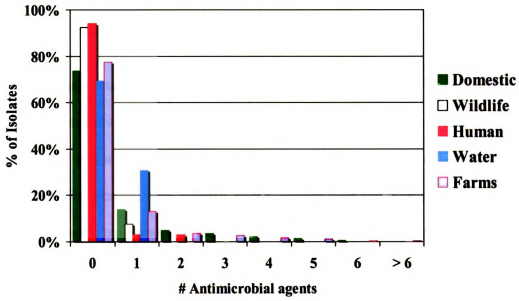


Figure 2-25: Multi-drug resistant *E. coli* isolates



CHAPTER 3

DISCRIMINANT ANALYSIS AS A TOOL FOR CLASSIFYING THE SOURCES OF *E. COLI* CONTAMINATION OF SURFACE WATER IN MICHIGAN, USA

Abstract

Discriminant analysis of patterns of antimicrobial resistance profile in *E. coli* was used to differentiate between different animals' sources of fecal pollution in water. The disc diffusion test was used to conduct in vitro antimicrobial susceptibility testing of 1169 *E. coli* isolates from cattle, pigs, poultry, sheep, farmed deer and wild geese fecal samples and human septic tank samples. Seven antimicrobial agents (neomycin, streptomycin, trimethoprim-sulfamethoxazole, tetracycline, ampicillin, cephalothin, and sulfisoxazole) were tested for susceptibility. When five source species were analyzed by discriminant analysis, the Average Rate of Correct Classification (ARCC) was 80.6% based on resubstitution method and 38.4% based on cross-validation method. The ARCC increased in both methods when less animal categories were analyzed; 96.8% and 60.8% in resubstitution and cross-validation methods, respectively, when animals pooled into domestic animals and compared with wildlife (wild geese) animals. When *E. coli* isolated from surface water receiving fecal pollution from unknown sources were analyzed, more than 19.2% of the isolates were classified as cattle, 26.9% as pig, 19.2% as sheep, 0.0% as poultry and 30.8% as wildlife, However, when animals were pooled as domestic and compared with wildlife animals 73.1% of the isolates were classified as domestic and 26.9% as wildlife animals. Based on the resubstitution discriminant analysis results, discriminant analysis should be considered further as one of the methods to determine the source of fecal pollution in water. The use of the resubstitution method

(each isolates is classified based on the patterns in the entire library, including its own patterns) as a measure of the performance of discriminant analysis gave higher ARCC than the more conservative method, the cross- validation (an individual isolate was removed from the library one at a time, then the removed isolate is classified based on the library comprised of the remaining isolates). The ARCC for two categories is higher than that of more than two in both methods.

Introduction

Fecal contamination of water is a widespread problem in the United States (United State Environmental Protection Agency, 1986). The high levels of fecal coliform bacteria in many lakes and rivers have impaired the quality of recreational and drinking water. Public exposure to pathogens in recreational and drinking water increased human health risks of acquiring pathogenic bacteria, resulting in reports of waterborne outbreaks of diseases involving fecal organisms (Swerdlow et al., 1992; Keene et al., 1994). In addition, increased levels of nutrients (phosphorus and nitrogen) in a receiving body of water can results in eutrophication that led to economic losses in shellfish industries (Crane and Moore, 1986). Knowing the source of fecal contamination is necessary to determine the degree of risk associated to human health in order to develop effective control strategies and resources management. In particular, it is important to determine whether the source of fecal contamination is of human, livestock, or wildlife origin as microorganisms of human origin are regarded as having greater potential to cause diseases in humans.

Water contamination can originate from point or non-point sources. The point source refers to single, identifiable points of origin, for example, municipal effluents or industrial discharge. On the other hand, non-point sources have diffuse origins such as surface runoff manure treated agricultural land or farm animal feedlots, failing or inadequate sewage treatment plants, leaking septic systems, sewer overflow, and wildlife waste.

In order to identify the source of water contamination whether it is from animals or humans, we need to have an indicator of fecal pollution. In our study, we used *E. coli* as an indicator bacterium of fecal contamination, because it is the type of fecal coliform bacteria that is commonly found in the intestinal tract of animals and humans. Also, it can be rapidly detected and easily numerated in the laboratory. *E. coli* is increasingly recognized as a water borne pathogen, because many outbreaks are associated with consuming drinking water (Swerdlow et al., 1992; Jones and Roworth, 1996) or coming into contact with recreational water (Keene et al., 1994; Ackman et al., 1997) had occurred. Total coliforms are used as an indicator of fecal contamination in water since they inhabit the intestine of warm-blooded animals, or are found naturally in the soil, vegetation and water. But *E. coli* is a species of total coliform that is always found in the feces and, therefore, a more direct indicator of human sewage or animal waste contamination of water.

During rainfall or snowmelt, *E. coli* can get into rivers, streams, lakes or groundwater and contaminate them. Using untreated water or inadequately treated water as a source for drinking water has a detrimental impact on human health. The ability of *E. coli* to survive in feces, on pasture land, and water systems has implications for its spread to

crops by direct application of manure, irrigation with infected water or by direct contact with animals or contaminated soil. It appears that *E. coli* can remain viable in soil for greater than 4 months (Jones, 1999). A study by Jones (1999), found that the survival of *E. coli* was greatest in soil cores containing rooted grass and the viable number decreased after 130 days, but when the organism was inoculated into cattle feces it remained detectable at high levels for more than 50 days. The organism survived less in cattle slurry and river water in which it fell to undetectable levels in 10-27 days. The survival on stainless steel surfaces also was investigated and it was found that *E. coli* could survive for 60 days, and most stable at a temperature of 4 ^{0}C. The viability reduced at 18 ^{0}C. It also survives for extended periods on domestic (plastic) cutting boards in both temperatures (Maule, 2000).

There have been several attempts to differentiate between human and nonhuman sources of fecal pollution. Initially, the ratio of fecal coliform to fecal streptococcus was used as an indicator of the source: the ratio of 4.0 or more indicates human source pollution while a ratio less than or equal to 0.7 would suggest nonhuman source of fecal pollution (Feachem, 1975). The rational behind the use of this method was that human feces contain higher fecal coliform counts while animal feces contain higher fecal streptococcus count. However, this method is not reliable anymore due to the variability in the survival rates of fecal streptococcus and variability in detection methods. Other investigators have proposed bacteriophages as indicators of the source of fecal pollution. Furuse et al., (1981) and Osawa et al., (1981) noticed that animal and human waste contained different serotypes of RNA coliphages. Tartera et al., (1989) suggested the use of *Bacteriodes fragilis* as an indicator of human source. However, the usefulness of these

approaches is limited because only a small percentage of human fecal samples contain phages, in addition to the difficulty in performing the assay (Troy et al., 2002).

An immuonological method, which is based on the presence of somatic O antigenic determinants, has been used to differentiate *E. coli* from different sources (Parveen et al., 2001). Different serotypes of *E. coli* are associated with different animal species but many serotypes are shared among humans and animals (Bettelheim et al., 1976), even though this overlap is not significant. The need for a large bank of antisera and the use of this method in conjunction with others to get valid results are the limitations to this approach (Parveen et al., 2001).

Recently, researchers used molecular methods such as DNA fingerprinting techniques to differentiate between human and animal source of fecal contamination of water. Such methods include ribotyping (Samadpour and Chechowitz, 1995; Hartel et al., 1999; Parveen et al., 1999; Carson et al., 2001; Hartel et al., 2002 and Troy et al., 2003), PCR of repetitive DNA sequences (Dombek et al., 2000), pulsed field gel electrophoresis (Kariuki et al., 1999; Parveen, et al., 2001), and 16S ribosomal DNA length heterogeneity PCR with terminal restriction fragment length polymorphisim (Bernhard and Field, 2000). These methods are expensive, labor intensive and need special experience in molecular era. Other researchers used chemical methods, for example, looking for caffeine in water since the human is the major consumer of this product through coffee, tea and other beverages (Troy et al., 2002), or searching for coprostanol, which is the fecal stanol that formed during the catabolism of cholesterol by the gut bacteria of humans. It is considered as a chemical indicator of human fecal pollution (Leeming et

al., 1996; Chan et al., 1998). The major limitation of this approach is that it requires an expensive chemical analysis.

Multiple antimicrobial resistance (MAR) approach is another way to identify the source of fecal contamination in water. The principle behind it is that bacterial flora in the gut of humans and animals are subjected to different types, concentrations, and frequencies of antimicrobial agents. Over time, selective pressure will select resistant bacteria that posses specific fingerprints against the antibiotics that have been used (Troy et al., 2002). Several attempts have been made to compare patterns of antimicrobial resistance in fecal coliforms with the source of isolates. Krumperman (1983) showed that multiple antibiotic resistance index of *E. coli* from wild animals (low risk) was generally low while human and poultry *E. coli* isolates had higher MAR indices (high risk), suggesting that multiple antibiotic resistance *E. coli* exist in large numbers within the major reservoirs of enteric diseases for humans, while present in low number elsewhere. Kaspar and Burgess, (1990) demonstrated that urban waters have higher percentage of resistant *E. coli* strains than rural waters, and antibiotic resistance *E. coli* may offer an index of water quality related to source of fecal contamination. The MAR profile is used to distinguish between *E. coli* that comes from point sources such as industrial, municipal effluents, and meat processing plant wastes; and non-point sources such as soil erosion and runoff over a wide area of land. A study by Parveen et al., (1997) found that more than 80% of *E. coli* strains isolated from municipal waste, river and estuarine water showed antibiotic resistance. The MAR test is relatively simple, cost effective and suitable for surveillance since a simple technique (Kirby-Bauer method or disc diffusion

test) is employed. This technique is easily standardized so that results will be highly reproducible across laboratories.

While multiple antimicrobial resistance patterns in *E. coli* can differ depending on the source of *E. coli*, more research is needed in techniques that will be accurate in separating these MARs into the three major potential sources of water fecal contamination, specifically, domestic animals, wildlife and human. One method that has shown promise is the discriminant analysis.

Discriminant analysis (DA) or discriminant function analysis is a multivariate statistical method concerned with separating sets of observations and allocating new observations to previously defined groups (Tatsuoka, 1970; Morrison, 1990; Johnson and Wichern, 1992). The main purpose of discriminant analysis is to determine which variables discriminate between two or more naturally occurring groups and to classify cases into the values of categorical dependent groups (Johnson and Wichern, 1992; Tatsuoka, 1970; Morrison, 1990). In our study discriminant analysis can be used to analyze the data of the isolates from known source libraries and generate a classification rule which then can be used to classify *E. coli* from unknown sources (water) into the closest source class in the database. It has been used successfully to classify fecal streptococcus, fecal coliforms and *E. coli* isolates based on the sources. Wiggins (1996) first demonstrated the use of antimicrobial resistance patterns in fecal streptococci and discriminant analysis to differentiate between human and animal sources with more than 90% correct classification and 84% correct classification when 6 species population were being classified. Other investigators used this approach successfully to differentiate human versus animal source of fecal contamination in water, which help in the direction

of water quality improvement (Hagedorn et al., 1999; Parveen et al., 1999; Wiggins et al., 1999; Harwood 2000; Carson 2001;Grave et al., 2002; Guan et al., 2002). Discriminant function analysis of antimicrobial resistance profile can offer a very low cost and statistically valid means of identifying the most probable species as a source for fecal contamination of surface water compared with new molecular methods of source identification such as RNA ribotyping and pulsed field gel electrophoresis (PFGE). The use of individual animal sampling, including domestic and wild animals, would minimize the risk of misclassification of *E. coli* isolates from fecal samples resulting in a more accurate discriminating function.

The overall goal of our study is to apply the discriminant analysis technique to classify *E. coli* isolates from a known source (e.g. cattle, pigs, sheep, etc.) by creating a classification rule on the basis of their patterns of antimicrobial resistance, and then to classify *E. coli* isolated from surface water contaminated with unknown source on the basis of the patterns of the known isolates.

Hypothesis tested

Application of the discriminant analysis technique on antimicrobial resistance profiles of *E. coli* isolated from domestic animals, wildlife (wild geese), and human septage is an accurate method for determining the source of fecal contamination of surface water.

Objectives

The objectives of this study were: 1) To determine the accuracy of discriminant analysis as a multivariate statistical technique to patterns of antimicrobial resistance of *E. coli* isolated from different species of animal's fecal samples and human septage.

2) Use this method to identify the source of *E. coli* contamination of surface water in Red Cedar Watershed in Michigan.

Materials and Methods:

Study design

A repeated cross-sectional study design approach was used to collect the animal and human fecal, and water samples and all the data related to antibiotic use on the farm during the four seasons (Spring, Summer, Fall, and Winter), starting in the winter of 2002 and end in the winter of 2003.

Study area

The Red Cedar Watershed was chosen as the study area, encompassing an area of 293,000 acres and takes in portions of 16 townships in Ingham and Livingston counties of the lower part of Michigan (figure 2-1). (Please notice that images in this thesis are presented in color). The Red Cedar River arises in Cedar Lake in the south central portion of the lower peninsula of Michigan and flows about 45 miles to its confluence with the Grand River in the city of Lansing. The Grand River empties into Lake Michigan, which connects to the other great lakes, the water of which collects in the Atlantic Ocean. The Red Cedar River has 12 tributaries and drains and provides mid-Michigan residents with numerous recreational opportunities, which include angling, canoeing, kayaking,

photography and bird watching. The river also serves as a source of water for the irrigation of crops throughout the watershed (figure 2-2). Swine and dairy are the predominant livestock in this watershed.

The study was conducted into two parts. Part one consisted of determining antimicrobial resistance profiles of *E. coli* isolated from fecal samples of animals and humans. The second part consisted of determining antimicrobial resistance in *E. coli* isolated from surface water.

Part I: Antimicrobial resistance profiling of *E. coli* from fecal specimens

Study population, enrollment of participating farmers, sample size, collection of fecal samples and data regarding antimicrobial use, isolation and identification of *E. coli* from fecal samples and antimicrobial susceptibility testing using Kirby Bauer method as described in chapter 2. We used the following seven antimicrobial agents in this study (neomycin, streptomycin, trimethoprim-sulfamethoxazole, tetracycline, ampicillin, cephalothin, and sulfisoxazole) rater than twelve.

Part II: Antimicrobial resistance profiling of *E. coli* from water samples

Water sample collection, isolation and identification of *E. coli* from water samples and antimicrobial susceptibility testing using Kirby Bauer method as described in chapter 2, except that we used seven (neomycin, streptomycin, trimethoprim-sulfamethoxazole, tetracycline, ampicillin, cephalothin, and sulfisoxazole) antimicrobial agents in this study rater than twelve.

Data analysis

Data were entered into an Excel spreadsheet. Fisher's exact test was conducted to identify significant differences in the rate of resistance to each antimicrobial agent and in the rate of multi-drug resistance by subpopulation, distributional characteristics and tests for homogeneity of covariance matrices, and multiple analysis of variance (MANOVA) were conducted (SAS v. 8.2, SAS Institute, Cary, NC). Those antimicrobial agents with a statistically significant difference in the percentage of resistance for one or more sub-populations at $p<0.05$ level were presented to the discriminant function model for multivariate analysis. The data were analyzed by SAS v. 8.2 (SAS institute, Cary, NC) by using the procedure DISCRIM. The nonparametric option was used for the development of the discriminant function model because the data were not distributed in a multivariate normal fashion and the covariance matrices were not homogeneous.

Assessing the validity of the discriminant analysis model

The classification table produced by discriminant analysis from the library of known source *E. coli* isolates was used to calculate the percentage of misclassified isolates and determine the average rate of correct classification. The table is a source-by-source matrix, in which the number and percentage of correctly classified isolates are found on the diagonal. The ARCC was computed by averaging the percentages of correctly classified isolates for each source along the diagonal. The percentage of misclassified isolates for a given source was determined by adding the percentages of misclassified isolates in the appropriate row of the table, excluding the value in the diagonal (Wiggins,

1996, Harwood et al., 2000). Once an acceptable classification rule is developed, discriminant analysis applies that rule to a set of isolates from an unknown source species (in this case the water samples) against the database of known source isolates and then classifies each water isolate into the most probable source species population.

Two techniques were applied to the development of the discriminant function classification rule: resubstitution and cross-validation. The default method for the software is the resubstitution method. With the resubstitution method each isolate is classified based on the patterns in the entire dataset including its own pattern. As a result, the ARCC from this method may overestimate the validity of the classification rule because the exact same dataset is used for both development of the rule and evaluation of its accuracy. The most conservation approach is to select the option to use the cross-validation method. This method of developing the classification rule is called the leave-one-out method. An individual isolate is removed from the dataset one at a time. The classification rule is developed from the remainder of the dataset and then the removed isolate is classified based on the rule created by n-1 observations.

The known species source library was stratified in two ways. The first consisted of 5 separate sub-populations: cattle, pigs, poultry, sheep, and wildlife (wild geese). The second consisted of only two subpopulations domestic animals (cattle, pigs, poultry, sheep) in one population and wildlife (wild geese) in another. Isolates from pets, and humans were not used to develop the discriminant analysis rule. Low rates of isolation of biochemically confirmed *E. coli* from these samples resulted in sample sizes that were too small for inclusion in the analysis. Discriminant function analysis is very sensitive to imbalances in sample size for the population being analyzed.

Results

A total of 2,292 fecal and septage samples were collected. The patterns of antimicrobial use reported by the study participants are detailed in table 2-2 (shown in chapter 2). The results of the Fisher's exact test, for the selection of antimicrobials to include in the multivariate analysis are detailed in table 2-3. Based on these findings the antimicrobial resistance profiles for the following drugs were used to develop the discriminant function: neomycin, streptomycin, sulfamethoxazole, tetracycline, ampicillin, cephalothin, and sulfisoxazole. Data from 1169 biochemically confirmed *E. coli* were used to develop the discriminant function model. The MANOVA tests the null hypothesis that the observations are all from the same population. Failure to reject the null hypothesis of the MANOVA would indicate that there is not adequate differentiation between the two populations to conduct discriminant analysis. The results of the MANOVA found a significant difference in population distributions for both the 5-species stratification scheme and the two-species stratification scheme (Wilk's lambda <0.001). The data were then presented to the discriminant function model. Tables 3-1 and 3-2 detail the average rate of correct classification obtained using the two methods of classification rule creation. Table 3-3 shows the comparison of the average rate of correct classification (ARCC) values between re-substitution and cross-validation rule development for different animal classification systems. The average rate of correct classification for the 5-source population classification scheme was unacceptably low for application to the water samples. However, for 2 sources classification 73.1% of *E. coli* isolated from water samples were classified as domestic including livestock (cattle, pig,

sheep) and poultry, and 27.0% as wildlife as shown in table 3-4.

Table 3-1: Comparison of classification tables for classification rules generated by non-parametric discriminant analysis of antimicrobial resistance profiles (neomycin, streptomycin, tetracycline, ampicillin, sulfamethoxazole, cephalothin, sulfisoxazole) using two different rule development methods, for five animal classes

		Re-substitution Method					Cross-validation Method				
Species	N	Cattle	Pigs	Poultry	Sheep	Wildlife[a]	Cattle	Pigs	Poultry	Sheep	Wildlife[a]
Cattle	534	63.3	12.4	4	10.7	10.7	41.0	18.4	5.6	12.9	22.1
Pigs	213	0.0	83.1	11.3	0.0	5.6	15.4	39.9	21.1	16.0	7.5
Poultry	106	0.0	0.0	95.3	0.0	4.7	6.6	28.3	43.4	14.2	7.6
Sheep	217	0.0	23.0	9.2	61.3	6.5	11.5	29.5	18.4	27.2	13.4
Wildlife	99	0.0	0.0	0.0	0.0	100.0	31.3	11.1	4.0	13.1	40.4
Total	1,169	ARCC* = 80.6					ARCC* = 38.4				

* ARCC: Average Rate of Correct Classification

[a]: Wild geese

Table 3-2: Comparison of classification tables for classification rules generated by non-parametric discriminant analysis of antimicrobial resistance profiles (neomycin, streptomycin, tetracycline, ampicillin, sulfamethoxazole, cephalothin, sulfisoxazole) using two different rule development methods, for two animal classes

		Re-substitution Method		Cross-validation Method	
Species	N	Domestic	Wildlife[a]	Domestic	Wildlife[a]
Domestic	1,070	93.5	6.5	87.8	12.2
Wildlife[a]	74	0.0	100.0	66.2	33.8
Total	1,144	ARCC* = 96.8		ARCC* = 60.8	

* ARCC = Average Rate of Correct Classification

a: Wild geese

Table 3-3: Comparison of average rate of correct classification (ARCC) values between re-substitution and cross-validation rule development for different animal classification systems

Animal classification	**Re-substitution**	**Cross-validation**
	ARCC (%)	**ARCC (%)**
Cattle, Pig, Poultry, Sheep, Wildlife[a]	80.6	38.4
Cattle, Pig, Poultry, Sheep	81.1	43.5
Cattle, pig, sheep	81.0	50.4
Domestic, Wildlife[a]	96.8	60.8

a: Wild geese

Table3-4: Classification of *E. coli* from water samples using discriminant analysis

Number of groups	Species (% classified)
5	Cattle (23.1), sheep (19.2), pig (27.0), poultry (0.0), wildlife[a](30.8)
4	Cattle (19.2) sheep (23.1), pig (27.0), wildlife[a](30.8)
2	Domestic (73.1), wildlife[a] (27.0)

[a] : Wild geese

Discriminant analysis using separate versus pooled sources

When five sources (cattle, pig, poultry, sheep, and wild geese) were analyzed by discriminant analysis using the resubstitution method, based on the resistance to seven anitmicrobial agents (neomycin, streptomycin, trimethoprim-sulfamethoxazole, tetracycline, ampicillin, cephalothin, and sulfisoxazole), the average rate of correct classification (ARCC) was 80.6%. However, when the same sources and antimicrobial agents were analyzed by discriminant analysis based on the cross-validation method the ARCC was 38.4% (Table3-1). When cattle, pig, poultry, and sheep were pooled and renamed domestic animals and compared with wild geese the ARCC increased to 96.8% and 60.8% by using the resubstitution and cross- validation methods respectively (table 3-2). The high classification rates were achieved when only two sources were compared strongly supports the conclusion that different species of animals harbor *E. coli* bacteria with different patterns of antimicrobial resistance profiles.

Analysis of unknown isolates from stream water

The antimicrobial resistance profiles of a total of 26 *E. coli* isolates from water samples were measured and classified by discriminant analysis (the known isolates were

used as reference); in other word, discriminant analysis was used to assign each isolate to a source category based on the comparison of its antimicrobial resistance patterns to those isolates from known sources. Using 5-source species (cattle, pig, poultry, sheep, and wild geese), 23.1% of *E. coli* isolates from water samples were classified as cattle, 19.2% as sheep, and 27% as pig, 0.0% as poultry and 30.8% wild geese (table 3-4). When 4-source species cattle, pig, sheep, and wild geese were analyzed, 19.2% of *E. coli* isolates from water samples were classified as cattle, 23.1% as sheep, 27.0% as pig and 30.0% as wild geese (table 3-4) suggesting that among domesticated animals, pigs contributed the greatest fecal pollution to the water stream. But when the isolates from known sources were grouped as both domestic and wild geese, 27% were classified as wild geese and 73.1% as domestic animals (table 3-4). This suggests that the water stream polluted with animal fecal material mainly from domestic animals.

Discussion

Several researchers have used discriminant analysis of antimicrobial resistance profile to identify fecal pollution sources (Wiggins, 1996; Hagedorn et al.1999; Wiggins et al., 1999; Harwood et al., 2000; Graves et al., 2002; Guan et al., 2002; Whitlock et al., 2002; Wiggins et al., 2003). The advantage of discriminant analysis method is that it generates a classification rule based on all the isolates, then uses this rule to classify each individual isolates into one of many possible sources. Wiggins (1996), who first used this method, reported that the ARCC was 84% when streptococcal isolates were pooled into four categories (cattle, human, poultry, and wild life). Moreover, the ARCC increased to 92% when all animals were pooled together and compared with human. However, Wiggins

used the resubstitution method to measure the performance of discriminant analysis in his study and five antibiotics with different concentrations. Our findings are similar to Wiggins's (1996) findings based on the resubstitution method as a measure of the performance of discriminant analysis. We found that the ARCC for 4 source species was 81.1%, for 3 source species was 81.0% and for 2 source species was 96.8% (table 3-3). There were consistently more correctly classified isolates (higher ARCC) when discriminant analysis was performed using two sources than when all four or five sources were used, because there are fewer possible categories. ARCC reflects both the false negative from that source and false positive from other sources (Wiggins, 1996); it is a better measure of the ability of a given analysis to classify the isolates. In addition, Wiggins (1996) tested a total of 1,435 isolates from 17 samples from cattle, poultry, human and wild life (multiple isolates from the same sample) leading to increase similarity of isolates within samples (isolates from same sample might have similar patterns) than similarity between samples which would make the library seem more representative than it actually is. Such an approach leads to high ARCC. In our study, we overcame this issue by using one isolate from one sample and not many isolates from the same sample that might have similar antimicrobial resistance patterns. We tested a total of 1169 *E. coli* isolates from 1169 fecal samples from different animal species. The drawback of sample level analysis is the assumption that all of the isolates in a given sample are from the same source, and this is valid for known homogeneous samples but not for a sample which was contaminated by more than one source (Wiggins et al., 1999). In a recent study, Wiggins et al., (2003) avoided this problem by using pulled-sample cross-validation analysis, where he removed all of the isolates from each sample and then

classified them against the remaining isolates. By comparing the difference between the ARCC of the resubstitution analysis, and the ARCC of the pulled-sample analysis, the representativeness of the library can be estimated. If the difference is small (less than 5%) then the library can be considered representative. In other words, new isolates are classified almost as well as the isolates in the library, but in this a case large number of isolates is needed to be in the library. Hagedorn et al. (1999) followed Wiggins's steps but he used a larger sample size of 7,058 fecal streptococcus isolates from humans, livestock, and wildlife (3 sources species) and demonstrated that the ARCC was 87%. Combining all animal isolates and comparing it with humans (2 source species) increased the correct classification rate to more than 95%. These results are similar to ours when we used the resubstitution analysis (ARCC for 3 species was 81.0% and for 2 species was 96.8%). Harwood et al. (2000) used discriminant analysis of antimicrobial resistance profiles but in his study using two separate datasets one from fecal streptococcus (4,619 isolates) and the other from fecal coliform (6,144 isolates) isolated from wild bird, cattle, chicken, dogs, pigs, and raccoons. He used nine antimicrobial agents to determine which one is the best to classify and identify the source of surface water contamination. He found that fecal streptococcus and fecal coliform databases classified isolates from known sources with similar accuracies, and the ARCC for fecal streptococcus was 62% and for fecal coliform was 63.9% by applying discriminant analysis on two categories animal and human using the cross-validation analysis. These results are similar to ours; when we applied the cross-validation discriminant analysis for two categories only (domestic animals and wild geese) and found the ARCC was 60.8%.

Guan et al., (2002) applied discriminant analysis on a total of 319 *E. coli* isolates using 14 antimicrobial agents and found that the ARCC when the data were pooled into three categories (human, livestock, and wildlife) was 64.5%, this low ARCC comparing with Wiggins and Hagedorn results may be due to several factors: first, fecal streptococcus was investigated by Wiggins and Hagedorn rather than *E. coli* so that the antimicrobial resistance profile in *E. coli* bacteria will be different from that in Streptococcus. Second, different antimicrobial agents were used, and the method of performing the antimicrobial susceptibility test was different between study by Guan et al., and studies by Wiggins (1996) and Hagedorn et al. (1999) and third, in Guan's study samples were collected from animals over a wide geographical area leading to more heterogeneous collection of bacterial isolates than other studies sampling protocol. When we applied the discriminant analysis on all *E. coli* (1169 isolates) from cattle, pigs, poultry, sheep and wild animals the ARCC was 38.4% based on the cross-validation results of discriminant analysis performance. Our results were similar to the results obtained by Guan et al., (2002) when applied discriminant analysis on multiple antimicrobial resistance (MAR) profile of *E. coli* isolated from nine host groups (human, beef, dairy, chicken, pig, turkey, deer, goose, moose); the ARCC was 33.9% when the cross-validation method was used.

Over time, as patterns of antibiotic use change, so do bacterial patterns of antibiotic resistance, and because the selective pressure of antibiotic treatment on microflora of animals and humans is an important determinant of the prevalence of antibiotic resistance in a population, the database that developed for discriminant analysis will require periodic updating (Harwood et al., 2000). The other issue is that it is not known if the

discriminant analysis of antimicrobial resistance profile from one geographical location can be used to predict the source of isolates from another since the patterns of antimicrobial agents may vary from one geographical area to another.

There are several issues that should be taken into consideration in order for the dataset or library to reliably identify fecal sources. First, the library needs to be representative of the sources that are present in the watershed. In other words, it should contain examples of all of the patterns found in the bacteria from each of the sources types that are found in the watershed with attention to the choice of antimicrobial agents used and variability of animal husbandry practices in different regions. Second, the library should be able to classify isolates from other geographical areas, and should be stable overtime so that new libraries do not need to be continually created. Wiggins et al. (2003) were able to obtain a representative, temporally stable merged multi-watershed library (6,587 enterococci isolates), from six Virginia Watersheds. When isolates from the contributed watersheds approximately one year later were analyzed with this library, they were classified as well as the isolates in the library suggesting that the resistance patterns are temporally stable for at least one year, but more samples will be needed to determine the extent of that variability (Wiggins et al., 2003).

Wiggins et al., (2003) found that the larger the multi-watershed library the lower the ARCCs and this is due to variability in the resistance patterns of the isolates within each source type in the watershed. The more isolates of each source type that are contained in the library (the more representative it is) the greater the chance they will vary in their resistance patterns, which would result in lowering the classification success. If there is variability in the patterns in a watershed, then the ARCC of a small library could be

misleading because it would be unable to classify the large number of unknown isolates that have a pattern that is not included in it. Therefore, it is unwise to relay on the ARCC of a library without also knowing its representativeness (Graves et al., 2002).

The best way to determine if the library is representative is to regularly add samples of known sources to it, if the ARCC and or the individual correct classification do not change significantly as new samples are added then the library should be representative (Hagedorn et al., 1999; Graves et al., 2002)

Discriminant analysis can assign each unknown isolate to one of the known sources according to the baseline data and the unknown organism's resistance pattern. The ability of the library to predict the source of the unknown isolates should be determined by assessing the classification accuracy of isolates from known sources that are not included in the library (Whitlock et al., 2002).

When we classified the *E. coli* isolates from the water samples (unknown) based on 5 categories (cattle, pigs, poultry, sheep and wild geese) 23.1% of *E. coli* isolates from water samples were classified as cattle, 19.2% as sheep, and 27% as pig, 0.0% as poultry and 30.8% wild geese (table 3-4). These results suggested that *E. coli* in the water samples came from animal fecal material mainly wild geese. When combining cattle, pig, sheep and poultry as domestic animals then compare the domestic animals with wild geese, 73.1% of the *E. coli* isolated from water samples were classified as domestic animals and only 27.0% classified as wild geese, this could be due to the similarity of antimicrobial resistance profile of *E. coli* isolated from water and the wild geese. The *E. coli* isolated from both the water samples and wild geese samples were resistant to cephalothin.

Although there is no established standard of accuracy that has been defined for any bacterial source tracking method, any method with a correct rate of classification of over 50% when there are five or more possible source categories, has been considered as a worthwhile tool for predicting the potential sources of fecal pollution in environmental water (Harwood et al., 2000). Based on this, and our results, the resubstitution discriminant analysis method gave higher ARCC (80.6%) than cross-validation discriminant analysis (38.4%) when 5 possible source categories were analyzed. This is not surprising since the same dataset that was used to create the classification rule was used to test it in the resubstitution analysis. Our findings support the work done by Wiggins et al. (2003) where he found that the ARCC based on the resubstitution analysis was higher than the cross-validation analysis.

The water quality authorities are interested in first, discriminating between human and animal sources of fecal pollution; second, determining the major sources of animal's contamination. It would be desirable to identify which type of animal was causing the pollution especially when the streams run through mixed agricultural areas. The results obtained by Hagedorn et al. 1999 study were very helpful for the watershed authorities in which cattle access to the river was prohibited by installation of fencing, and in- pasture watering stations, which lead to a reduction in fecal coliforms counts from 15,900 per 100ml to 960 per 100ml.

Antimicrobial resistance analysis requires several days to obtain results, because growth of bacteria is required to apply the susceptibility test. This time period may be critical for public health officials who need to make rapid decisions on closure of recreational water contaminated by fecal material. However, the antimicrobial resistance

test is relatively inexpensive, and technicians can quickly be taught how to perform the assay and apply it on several hundreds of isolates per week. Comparing with the molecular methods that is expensive, needs experience, and could only be performed on a small number of isolates.

Although we did not have enough samples from human source to be classified, *E. coli* could be classified well since the antimicrobial resistance profile of *E. coli* isolated from human septic tanks was similar to that from wild geese and water. In addition, sampling the septic tanks rather than treatment plant has an advantage because it is more likely that influent in many plants could have included overland flow from agricultural land, which should have introduced contaminating bacteria from animal sources and thus reduced the classification success.

In the future, it will be worthwhile to combine discriminant analysis of antimicrobial resistance profiles with that for molecular methods, the ribotyping, on a large dataset to cross validate both approaches and to assess wither one method might be more suitable than the other.

In conclusion, the results of our study suggest that discriminant analysis of antimicrobial resistance profile could be used to differentiate among isolates from several sources of fecal pollution. It provides a strong method for classifying and identifying fecal *E. coli* and will serve to help identify the non-point sources of fecal pollution in water, which will aid in the evaluation of risk to the public health. The underlying assumption of antimicrobial resistance analysis is that the differential use of antimicrobial agents in humans, domestic animals or in wild animals lack thereof allows for discrimination of indicator organism isolates from those hosts. The use of resubstitution

method (each isolates is classified based on the patterns in the entire library, including its own patterns) as a measure of the performance of discriminant analysis gave higher ARCC than the more conservative method, the cross-validation (an individual isolate was removed from the library one at a time, then the removed isolate classified based on the library comprised of the remaining isolates). The ARCC for 2 categories is higher than the one of more than two in both methods.

Moreover, our study not only classify and identify the sources of fecal contamination in water but also provided us with information about the antimicrobial agents that *E. coli* isolated from different animals species in the watershed showed resistance. This information would be useful for farmers and veterinarians to know for future use.
The size of the library is critical to the success of discriminant analysis in predicting the sources of fecal contamination in water. It has been demonstrated that small libraries generally give higher correct classification and this due to under sampling of the true population diversity in fecal material (Wiggins et al., 1999; Harwood et al., 2000).
The ability of the library to predict the source of indicator organisms should be determined by assessing the classification accuracy of isolates from known sources that are not included in the library (Whitlock et al., 20002). In order to classify a new isolates from another geographical area the library used should be representative, large and temporally stable for at least one year. In order for discriminant analysis to be valid it should show its applicability in the field rather than simply showing high correct classification rates.

The sampling protocol should reflect the type of the watershed under study; for example, companion animals could be added as a category for more urban watersheds,

and wildlife removed if necessary. On the other hand, sampling should be taken more from domestic animals such as cattle, pigs, poultry, sheep and wild animals in addition to human in rural watershed. By doing that, the samples will be representative to the type of source species that live in the watershed.

Knowing the sources of fecal pollution in water is a very important step in determining the degree of risk for humans exposed to contaminated water, assessing the development of best management practices to reduce the fecal loading which include stream fencing, establishing riparian-zone buffers, installing in pasture watering stations, and improving waste treatment facilities and amending leakage in a septic tank.
We concluded that antimicrobial resistance analysis of *E. coli* isolates analyzed with discriminant analysis was a suitable method to differentiate and identify sources of fecal pollution in Red Cedar River where the classified isolates came from multiple sources, if our results based on taking the resubstitution method as a measure of performance of discriminant analysis, but this method is biased and it is recommended to take cross-validation method into consideration.

OVERALL SUMMARY AND CONCLUTIONS

A repeated cross–sectional design was used to conduct studies to investigate whether combining two techniques (determining antimicrobial resistance profiles of *E. coli*, and conducting discriminant function analysis on such profiles) would be useful in identifying sources of fecal contamination of surface water in Michigan. The Red Cedar watershed was used as a case site. Fecal samples were collected from livestock and companion animals living on farms that drained into the Red Cedar watershed, human septic tanks from homes in the watershed, and wildlife living in the watershed. Water from the Red Cedar was collected from several different sites for the study. The study was conducted in two parts, presented here as chapters.

There were two objectives for the first part of the study (Chapter 2): 1) determining patterns of antimicrobial resistance in *E. coli* obtained from human, domestic animals, and wildlife living in the Red Cedar watershed in Michigan, USA; and 2) comparing patterns of antimicrobial resistance from these animals with patterns found in *E. coli* obtained from surface water samples from the same watershed. Using disc diffusion assay, the following antimicrobial agents were tested for susceptibility: neomycin, gentamicin, streptomycin, chloramphenicol, ofloxacin, trimethoprim/sulfamethoxazole, tetracycline, ampicillin, nalidixic acid, nitrofurantoin, cephalothin, and sulfisoxazole. The study found:

1. Resistance to at least one antimicrobial was demonstrated in isolates from food animals, wildlife, surface water, and human septic tanks.

2. In general, *E. coli* isolates from food animals showed resistance to the largest number of antimicrobial agents, followed by *E. coli* from horses, companion animals, human septic tanks, wildlife, and surface water.
3. The three antimicrobial agents to which resistance was demonstrated most frequently in this study were cephalothin, tetracycline, and streptomycin.
4. Further, the antimicrobial resistance profiles of *E. coli* isolated from companion animals and human septic tanks were closer to those from surface water isolates, than from food animal species.
5. Within the groups tested, food animal species (cattle, sheep, goats, poultry and swine) manifested the highest rates of antimicrobial resistance for all but 4 of the individual antimicrobials tested (gentamicin 0.7%, sulfamethoxazole 1.7%, ampicillin 7.2%, and cephalothin 21.8%). Additionally, food animals had statistically significantly (Fisher's exact $p < 0.0001$) higher rates of resistance to 3 or more antimicrobials (13.5%).

In the second part of the study (Chapter 3), the objectives were: 1) to test the validity of discriminant function analysis of antimicrobial resistance profiles to identify the source of fecal *E. coli* isolates, and 2) use this method to identify the most probable sources of fecal *E. coli* contamination of surface water in the Red Cedar watershed in Michigan. After conducting statistical analyses to determine which antimicrobial agents had the greatest probability of resistance, two methods for conducting discriminant function analysis (resubstitution and cross-validation) were applied to the data. The major findings from this part of the study include the following:

1. Our studies agree with the literature, in that the resubstitution method is biased, it produces higher, but invalid ARCC. In contrast, the cross-validation method produces lower but valid ARCC.
2. Using the cross-validation method, it was found that the ARCC increased as the number of source categories was reduced. As an example, the ARCC for five source categories was 38.4%, compared to 60% when two source categories were used.

Recommendations

The use of discriminant function analysis of antimicrobial resistance profiles to determine the source of fecal contamination of surface water has shown great promise, but some limitations still remain. For this method to serve as a very useful tool future studies should include:

1. Using a larger number of samples from human and animal sources.
2. Use fewer numbers of antimicrobial agents during the development of the classification rule.
3. Evaluate the use of MICs instead of binary outcomes (resistant or not resistant) in discriminant analysis.
4. Evaluate additional statistical methods, such as cluster analysis, prior to conducting the discriminant function analysis to reduce the number of variables being assessed.
5. Compare the performance of discriminant analysis to ribotyping and other molecular techniques.

APPENDIX

Surface Water Study
Informed Consent Form

The overall goal of this study is to identify various sources of surface water contamination. To complete this goal, the study will develop a low-cost screening tool for the identification of the source species for fecal contamination of surface water, and then identify factors associated with fecal contamination of surface water by comparing bacteria from surface water with bacteria collected from deer, waterfowl, pets, household septic systems, and livestock.

A data collector from MSU will contact you, and administer a questionnaire designed to describe some of the different livestock and manure management practices being used by Michigan livestock operators. This questionnaire will be completed through and in-person interview. The interview itself should last less than half an hour. A sample collector from MSU will schedule an appointment to come to your farm to collect manure samples from manure storage units and a small percentage of your livestock. These samples will be sent to Michigan State University for laboratory testing, to compare the bacteria from your livestock to those bacteria found in local bodies of water.

All information collected by this study will be kept confidential: no identification will be kept that will tie your identity with questionnaire and laboratory test results. Only researches immediately involved with the study will have access to these data: no outside private or governmental groups will be able to use these data. Only summaries of the data will be used to generate reports, and information will be reported so that it will not be possible to identify any specific individual participating in the study. *Your privacy will be protected to the maximum extent allowable by law.*

Again, your participation in this study is voluntary: refusal to participate will involve no penalty, and you may discontinue participation at any time.

_____

Participant Data Collector

_____

Date Date

To discuss any questions regarding the research, please contact:
Dr. John B. Kaneene
Phone- (517) 353-5941, Fax (517) 423-0976
Email-

To discuss questions about your role as a subject of research, please contact:
Dr. Ashir Kumar, Chairman of the university Committee on Research Involving Human Subjects, MSU (517) 355-2180

January 30, 2002

Dear Michigan Farmer:

The college of Veterinary Medicine at Michigan State University will be conducting a study to determine the source of bacterial contamination of surface water. Bacterial surface water contamination, such as *E. coli*, can be from a number of sources, including non-agricultural sources. Your participation will help us to identify these sources.

We would like to ask for your cooperation in part of this study. If you agree, your participation in the study will be in two steps. First, you will be asked to complete a short questionnaire. This can be done on your farm or at your local county Extension office. After the questionnaire is completed, a sample collector from MSU will schedule an appointment with you, to come to your farm to collect manure samples from your manure storage units and a small percentage of your livestock. These samples will be sent to MSU for laboratory testing, to compare the bacteria types from your livestock to those bacteria found in local bodies of water.

All information collected will be kept strictly confidential. No identifying information about your specific operation will be collected during the study, so it cannot be released to any governmental agency. All results of the study will be presented as summaries, so that no specific information about your farm will be identifiable. If you would life to see the results of bacterial testing for your operation, your specific test results will be released only to you upon your request through your county Extension agent.

Your participation is optional and voluntary. If you would like to help with our research by participating in this study, please fill out the enclosed, pre-stamped post card with your name and contact information. We will work through your county Extension agent to schedule a time for you to come to the office to ask any questions you have, sign an informed form, and complete the questionnaire.

If you have any questions or concerns about the study, or your participation in the study, please contact Dr. John B. Kaneene at (517) 353-5941.

Thank you for your time and assistance.

John B. Kaneene, DVM, MPH, Ph.D
Professor of Epidemiology and
Director, Population Medicine Center

March 12, 2002

Dear Dairy & Livestock Producer:

Recently, you received an invitation to take part in a study regarding *E. coli* and livestock. The study is being performed by the College of Veterniary Medicineat Michigan State University. The *E. coli* bacteria can come from waterfowl, septic systems, and many other types of animals. The research will attempt to link *E.coli* strains with farm animal types. In part, this research will help us protect the animal industry in Michigan from being associated with *E.coli* contamination in surface water, when it may be from non-agricultural sources.

We have heard from a number of Ingham County producers so far, and *all* of these produces who have livestock have indicated their desire to participate. However, we have not yet heard from you.

Of course, your participation is completely optional and voluntary. If you would like to participate in this study, please fill out the enclosed, pre-stamped post card with your name and contact information. If not, please indicate on the post card that you are not interested and return it. By doing so, we will remove your name from the list for any further reminders.

If you have any questions or concerns about the study, or your participation in the study, please contact Dr. John B. Kaneene at (517) 353-5941. The attached letter and form will explain the study.

Thank you for your time and assistance.

Sincerely

Mark F. Hansen
Extension Coordinator, and
Extension Agriculture & Natural Resources Agent

Cc: Dr. John B. Kaneene
Dr. Raida Sayah

March 25, 2002

Dear Agricultural Producer:

The College of Veterniary Medicineat at Michigan State University is performing a study that involves *E. coli* and livestock waste. Many people are not aware that *E. coli* bacteria can come from different sources including livestock, birds, pets, wildlife and humans. The goal of this study is to identify strains of *E. coli* to better understand the sources of *E. coli*, which are found in surface wares in the mid-Michigan area.

As your Extension Agriculture Agent, I have agreed to send out this information. In this way, we were able to help the research project without revealing your name.

You have the opportunity to participate in this research project. Your participation is completely optional and voluntary. Information colleted will be kept confidential. This study will further our knowledge of bacterial contamination of surface water, which in tur help us demonstrate agriculture's quest for environmental quality. The research team has enclosed a cover letter, consent form and return postcard.

If you have questions, please call me at (517) 546-3950. You may also call the research leader, Dr. John B. Kaneene directly at the College of Veterinary Medicine, for more study details. Dr. Kaneene can be reached at (517) 355-5941.

Sincerely,

Betsy Dierberger
Extension Agriculture and Natural Resources Agent

REFERENCES

REFERENCES

Ackman, D., S. Marks, P. Mack, M. Caldwell, T. Root and G. Birkhead. 1997. Swimming-associated haemorrhagic colitis due to *Escherichia coli* O157:H7 infection: evidence of prolonged contamination of fresh water lake. Epidemiol. Infect. 119, 1-8.

Alekshun, M. N. and S. B. Levy. 1999. The *mar* regulon: multiple resistance to antibiotics and other toxic chemicals. Trend Microbiology. 7:410-413.

American public health association. 1998. Standard Methods for Examination of Water and Wastewater, 20th ed. Clesceri, Greenberg, And Eaton (ed.), American Public Health Association; American Water Works Assn; and Water Environment Federation. Washington DC Methods, Number 9225.

Anderson, J. D. 1974. The effect of R-factor carriage on the survival of *Escherichia coli* in the human intestine. J. Med. Microbiol. 7:85-90.

Anderson, T. W. 1951b. Classification by multivariate analysis. Psychometrika. 16:31-50.Application of sedimentary fecal stanols and sterols in tracing sewage pollution in coastal water. Water Res. 32:225-235.

Ash, R. J., B. Mauck, and M. Morgan. 2002. Antibiotic resistance of gram-negative bacteria in rivers, United States. 8(7):713-716.

Barton, M. 1998. Does the use of antibiotic in animals affect human health? Aust. Vet. Journal. 76(3):177-180.

Bernhard, A. E., and K. G. Field. 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Applied and Environmental Microbiology. 66(4):1587-1594.

Bettelheim, K. A., N. Ismail, R. Shinebaum, R. A. Shooter, E. Moorhouse and S. O'Farrel. 1976. The distribution of serotypes of Escherichia coli in cow pats and other animal materials compared with serotypes of *E. coli* isolated from human sources. Journal of Hygiene. 76:403-406.

Blanco, J. E., M. A. Mora and J. Blanco. 1997. Prevalence of bacterial resistance of quinolones and other antimicrobial among avian *Escherichia coli* strain isolated from septicemic and healthy chickens in Spain. Journal of Clinical Microbiology. 35(8):2184-2185

Carson, A., B.L. Shear, M. R. Ellersieck, and A. Asfaw. 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Applied and Environmental Micobiology. 67(4):1503-1507.

Casewell, M., C. Friis, E. Marco, P. McMullin, and I. Phillips. 2003. The European ban on growth- promoting antibiotics and emerging consequences for human and animal health. Journal of Antimicrobial chemotherapy. 52:159-161.

Chan, K. H., M. H. W. Lam, K. F. Poon, H.Y. Yeung, and T. K. T.Chui. 1998.

Crane, S. R., and J.A. Moore. 1986. A management strategy to reduce bacterial pollution in shellfish areas: a case study. Environ. Manag. 10:41-51.

Difco Manual. 11th ed. 1998. Difco Laboratories a Division of Becton Dickinson and Co. Sparks, MD USA. pages 527, 556, 288, 525 for isolation of *E. coli* . Page 521 for triple sugar iron test.

Dombek, P.E., L. K. Johnson, S. T. Zimmerly, and M. J. Sadowsky. 2000.Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Applied and Environmental Microbiology. 66(6):2572-2577.

Dufour, A.P., E.R. Strickland, V.J. Cabelli. 1981. Membrane filter teqnique for enumerating *Escherichia coli* . Applied and Environmental Microbiology. 41:1152-1158.

Feachem, R. 1975. An improved role for fecal coliform to fecal streptococci ratios in the differentiation between human and nonhuman pollution sources. Water Research. 9:689-690.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problem. Annals of Eugenics. 7:179-188.

Furuse, K. A., S. Osawa, and I. Watanabe. 1981. Distribution of ribonucleic acid colipghages in raw sewage from treatment plants in Japan. Applied and Environmental Microbiology. 41:1139-1143.

Galland, J. C., D. R. Hayatt, S. S. Crupper, and D. W. Acheson. 2001. Prevalence, antibiotic susceptibility, and diversity of *Escherichia coli* O157:H7 isolates from a longitudinal study of beef cattle feedlots. Applied and Environmental Microbiology. 67(4):1619-1627.

Ginns, C. A., G. F. Browning, M. L. Benham, G.A. Anderson and K. G. Whithear. 1996. Antimicrobial resistance and epidemiology of *Escherichia coli* in broiler breeder chicken. Avain Pathology. 25:591-605.

Grabow, W. O. K., I. G. Middendroff, and O. W. Prozesky. 1973. Survival in maturation ponds of coliform bacteria with transferable drug resistance. Water Research. 7:1589-1597.

Grabow, W. O. K., M. Van Zyl, and O. W. Prozesky. 1976. Behavior in conventional sewage purification process of coliform bacteria with transferable or non-transferable drug- resistance. Water Research. 10:717-723.

Graves, A. K., C. Hagedorn, A.Teetor, M. Mahal, A. M. Booth, and R. B. Reneau, Jr. 2002. Antibiotic resistance profiles to determine sources of fecal contamination in a rural virginia watershed. Journal of Environmental Quality. 31:1300-1308.

Grigal, D. F. and R. A. Goldstein. 1971. An integrated Ordination –classification analysis of an intensively sampled oak-hickory forest. Journal of Ecology. 59:481-492.

Guan, S., R. Xu, S. Chen, J. Odumeru, and C. Gyles. 2002. Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. Applied and Environmental Microbiology. 68(6):2690-2698.

Hachler, H., S. P. Cohen, and S. B. Levy. 1991. marA, a regulated locus which controls expression of chromosomal multiple antibiotic resistance in *Escherichia coli*. Journal of Bacteriology. 173:5532-5538.

Hagedorn, C., S. L. Roberson, J.R. Filtz, S. M. Grubbs, T. A. Angier and R. B. Reneau, JR. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Applied and Environmental Microbiology. 65(12):5522-5531.

Hartel, P. G., J. D. Summer, J. L. Hill, J. V. Collins, J. A. Entry, and W. I. Segars. 2002 Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. Journal of Environmental Quality. 31:1273-1278.

Hartel, P. G., W. I. Segars, N. J. Stern, J. Steiner and A. Buchan. 1999. Ribotyping to deterime host origin of *Escherichia coli* isolates in different water samples, P.377-382. In D.S. Olsen and J.P. Potyondy (ed.), Wildland hydrology. Technical publication Series TPS-99-3. American Water Resources Asssociation, Herndon, Va.

Harwood, V. J., J. Whitlock, and V. Withington. 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. Applied and Environmental Microbiology. 66(9):3698-3704.

Hix, D. M. 1988. Multifactor classification and analysis of upland hardwood forest ecosystems of the Kickapoo River watershed, Southwestern Wisconsin. Can. J. For. Res. 18:1405-1415.

Johnson, R. A. and D. W. Wichern. 1992. Applied Multivariate Statistical Analysis, 3rd edition. Chapter 11, pages 493-556, New Jersey: Englewood Cliffs.

Jones, D. L. 1999. Potential health risk associated with the persistence of *Escherichia coli* O157 in agricultural environments. Soil Use and Management, 15:76-83.

Jones, I. G. and M Roworth. 1996. An outbreak of *Escherichia coli* O157 and campylobacteriosis associated with contamination of drinking water supply. Public health (London) 110:277-282.

Jones, M. E., E. peters, A. M. Weersink, A. Fluit, and J. Verhoef. 1997. Widespread occurrence of integrons causing multiple antibiotic resistance in bacteria. Lancent. 349:1742-3.

Josef, S. 2000. Computer recognition of skin structures using discriminant and cluster analysis. Skin Research and Technology. 6(2):58.

Kariuki, S., C. Gilks, J. Kimari, A. Obanda, J. Muyodi, P. Waiyaki, and C. A. Hart. 1999. Genotype analysis of *Escherichia coli* strains isolated from children and chickens living in close contact. Applied and Environmental Microbiology. 65:472-476.

Kasper, C. W., J. L. Burgess. 1990. Antibiotic resistance indexing of *Escherichia coli* to identify sources of fecal contamination in water. Can. J. Microbiol. 36: 891-894.

Keene, W. E., J. M. McAnulty, F.C. Hoesly, et al. 1994. A swimming-assiciated outbreak of hemorrhagic colitis caused by *Escherichia coli* O157:H7 and shigella Sonnei. New England Journal of Medicine. 331(9):579-584.

Kelch, W. J. and Lee J. S. 1978. Antibiotic resistance patterns of gram-negative bacteria isolated from environmental sources. Applied and Environmental Microbiology. 36(3):450-456.

Klecka, W. R. (1980). Discriminant Analysis. Quantitative Applications in the Social Sciences Series. No. 19. Thousand Oaks, CA: Sage Publications.

Kruperman, P. H.1983. Multiple antibiotic resistance indexing of *Escherichia coli* to identify high-risk sources of fecal contamination of foods. Applied and Environmental Microbiology. 46(1): 165-170.

Lachenbruch, P. A. and M. A. Mickey.1968. Estimation of error rates in discriminant analysis. Technometrics. 10:1-10.

Lachenbruch, P.A. 1975. Discriminant Analysis. NY: Hafner.

Langlois, B.E., G. L. Cromwell, T. S. Stahly, K. A. Dawson, and V. W. Hays. 1983. Antibiotic resistance of fecal coliforms after long-term withdrawal of therapeutic and subtheraputic antibiotic use in swine herd. Applied and Environmental Microbiology. 46(6): 1433-143.

Lattin, J., J. Douglas Carroll, P. E. Green. 2003. Analyzing Multivariate Data. Chapter 12, pages 426-473.

LeClerc, J. E., B. Li, W. L. Payne, and T. A. Cebula. 1996. High mutation frequencies among *Escherichia coli* and Salmonella pathogens. Science. 274:1208-1211.

Leeming, R., A. Ball, N. Ashbolt, and P. Nichols. 1996. Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters. Water Research. Pages 2893-2900.

LeMay, V. M., Tait D. E., Van Der Kamp B. J. 1994: Classification of cedar, aspen, and tru fir trees as decayed versus sound. Can. J. For. Res. 24:2068-2077.

Levy, S. B. 1987. Antibiotic use for growth promotion in animals: Ecologic and public health consequences. Journal of food protection. 50(7):616-620.

Linton, A. H. 1977. Antibiotic resistance: the present situation reviewed. Vet. Rec. 100:354-360.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. Proceeding of the national institute of sciences of India. 68:176-179.

Matthews, J. A. 1979. A study of the variability of some successional and climax plant assemblage-types using multiple discriminant analysis. Journal of Ecology. 67:255-271.

Maule, A. 2000. Survival of verocytotoxigenic *Escherichia coli* O157 in soil, water and on surfaces. Journal of Applied Microbiology. 88:71S-78S.

Mckeon, D.M., J. P. Calabrese, G. K. Bissonnette. 1995. Antibiotic resistant gram-negative bacteria in rural ground water supplies. Water Research. 29:1902-1908.

McIntosh, L. M., M. Jackson, H. H. Mantsch, M. F. Stranc, D. Pilavdzic and A. Neil Crowson. 1999. Infrared spectra of basal cell carcinoma are distinct from non-tumor bearing skin components. Journal of Investigative Dermatology. 112 (6):951.

Meng, J., S. Zhao, M. P. Doyle, S. W. Joseph. 1998. Antibiotic resistance of *Escherichia coli* O157:H7 and O157:NM isolated from animals, food and humans. Journal of Food Protection. 61(11):511-1514.

Morrison, D. F. 1990. Multivariate Statistical Methods. 3rd edition. Chapter 6, pages 269-287, McGraw-Hill Publishing Co.

Mulamattathil, S. G., H. A. Esterhuysen, P. J. Pretorius. 2000. Antibiotic resistant gram-negative bacteria in a virtually closed water reticulation system. Journal of Applied Microbiology. 88:930-937.

Munro, H. L., and R. C. Rounds. 1985. Selection of artificial nest sites by five sympatric passerines. Journal of wildlife management. 49:264-276.

National Committee for Clinical Laboratory Standards. 1997 Approved Standard M2-A6. Performance standards for antimicrobial disc susceptibility tests. 6th ed. NCCLS, Wayne, Pa.

National Committee for Clinical Laboratory Standards.1999. NCCLS document M100-S9. Performance standards for antimicrobial susceptibility testing, 9th ed. Informational supplement. NCCLS, Wayne, Pa.

Neu, H. C. 1992. The crisis in antibiotic resistance. Science. 257: 1064-73.

Norris, J. M. and J. P. Barkham. 1970. A comparison of some Costwold Beechwoods using multiple discriminant analysis. Journal of Ecology. 58:603-619.

Osawa, S., K. Furuse, and I. Watanabe. 1981. Distribution of ribonucleic acid colipghages in animals. Applied and Environmental Microbiology. 41:164-168.

Parveen, S., and M. L. Tamplin. 1998. Methods for measuring levels of human and nonhuman sources of fecal pollution in water, abstr. Q-108, P. 439. In abstracts of the 98th general meeting of the American Society for Microbiology 1998. American Society for Microbiology, Washington, D.C.

Parveen, S., K. M. Portier, K. Robinson, L. Edmiston and M. Tamplin. 1999. Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Applied and Environmental Microbiology. 65(7):3142-3147.

Parveen, S., N. C. Hodge, R. E. Stall, S. R. Farrah and M. L. Tamplin. 2001. Genotypic and phynotypic characterzation of human and non human *Escherichia coli* . Water Res. 35:379-386.

Parveen, S., R. L. Murphree, L. Edmiston, C. W. Kaspar, K. M. Portier, and M. L. Tamplin. 1997. Association of multiple antibiotic resistance profiles with point and nonpoint sources of *Escherichia coli* in Apalachicola Bay. Applied and Environmental Microbiology. 63(7):2607-2612.

Piddock, L. 1996. Does the use of antimicrobial agents in veterinanry medicine and animal husbandry select antibiotic-resistant bacteria that infect man and compromise antimicrobial chemotherapy? Journal of Antimicrobial Chemotherapy. 38: 1-3.

Pillai, S. D., K. W. Widmer, K.G. Maciorowski, S.C. Ricke. 1997. Antibiotic resistance profiles of *Escherichia coli* isolated from rural and urban environments. Journal of Environ. Science. Health. 32:1665-1675.

Pregitzer, K. S. and B. V. Barnes. 1984. Classifiacation and comparison of upland hardwood and conifer ecosystems of the Cyrus H. McCormick experimental forest, upper Michigan. Can. J. For. Res. 14:362-375.

Prescott, J. F., J. D. Baggot, and R. D. Walker. 2000. Antimicrobial Therapy in Veterinary Epidemiology. 3rd edition. Iowa State University Press, Ames, IA.

Press, S. J. and S. Wilson. 1978. Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association. 73:699-705.

Rao, C. R. 1948. The utilization of multiple measurements in problems of biological classification. Journal of the royal Statistical Society, Series B, 10:159-193.

Rao, C. R. 1973. Linear Statistical Inference and Its Applications, 2nd Edition, New York: John Wiley& Sons, Inc.

Rice, J. R., D. Ohmart, and B. W. Anderson. 1983. Habitat selection attributes of an avian community: a discriminant analysis investigation. Ecological Monographs. 53:263-290.

Richmond, M. H.1972. Some environmental consequences of the use of antibiotics: or "what goes up must come down". Journal of Applied Bacteriology. 35:155-176.

Rosenblatt, M.1956. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics. 27:832-837.

Rubens, C. E., W. F. McNeill, and W. E. Farrar, Jr. 1979. Evolution of multiple antibiotic resistance plasmids mediated by transposable plasmid deoxyribonucleic acid sequences. Journal of Bacteriology.140(2):713-719.

Samadpour, M., and N. Chechwtiz. 1995. Little Soos Creek microbial source tracking: a Survey. University of Washington, Dep. of Environ. Health, seattle, WA.

SAS institute, Inc. 1999, SAS/STAT User's Guide Version 8.2 Cary, NC: SAS institute, Inc.

Schroeder, C. M., C. Zhao, C. DebRoy, J. Torcolini, S. Zhao, D. G. White, D. D. Wagner, P. F. McDermott, R. D. Walker, and J. Meng. 2002a. Antimicrobial resistance of *Escherichia coli* O157 isolated from humans, cattle, swine, and food. Applied and Enviromental Microbiology. 68(2):576-581.

Schroeder, C. M., J. Meng, S. Zhao, C. DebRoy, J. Torcolini, C. Zhao, P. F. McDermott, D. D. Wagner, R. D. Walker, and D. G. White. 2002b. Antimicrobial resistance of *Escherichia coli* O26, O103, O111, O128, and O145 from animals and humans. Emerging Infectious Diseases. 8(12):1409-1414.

Seagle, S. W. 1985. Patterns of small mammal microhabitat utilization in cedar glade and deciduous forest habitats. Journal of Mammalogy. 66:22-35.

Smith, C., A. B. 1947. Some examples of discrimination. Annals Eugenics. 18:Page 272.

Smith, H. W. 1975. Persistence of tetracycline resistance in pig *E. coli*. Nature 258:628-639

Smith, P. R., E. Farrel, and K. Dunican. 1974. Survival of R+ *Escherichia coli* in sea water. Applied Microbiology. 27(5):983-984.

Smith, R. D.1995. Veterinary Clinical Epidemiology, 2nd Edition. pages 130-131. CRC Press, Ann Arbor.

Spies, T.A., and B. V. Barnes.1985a. A multifactor ecological classification of the northern hardwood and conifer ecosystems of Sylvania Recreation Area. Upper Peninsula, Michigan. Can. J. For. Res. 15:949-960.

Spratt, B. G. 1994. Resistance of antibiotics mediated by target alterations. Science 264:388-393.

Swann Committee. 1969. Report of Joint Committee on the Use of Antibiotic in Animal Husbandry and Veterinary Medicine. Her Majesty's Stationary Office. London, September.

Swerdlow, D.L., B. Woodruff, R. Brady, et al. 1992. A waterborne outbreak in Missouri of *Escherichia coli* O157:H7 associated with bloody diarrhea and death. Annals of Internal Medicine.117:812-819.

Tabachnick, B. G. and L. S. Fidell. 1996. Using Multivariate Statistics. Harper Collins College Publishers: New York.

Tartera, C., F. Lucena, and J. Jofre. 1989. Human origin of Bacteroides fragilis bacteriopgages present in the environment. Applied and Environmental Microbiology. 55:2696-2701.

Tatsuoka, M. M. 1970. Selected Topics in Advanced Statistics, Number 6, Illinios. Tonn, W. M., and J. J. Magnuson. 1982. Patterns in the species composition and richness of fish assemblages in northern Wisconsin lakes. Ecology. 63:1149-1166.

Troy, M. S., J. B. Rose, T. M. Jenkins, S. R. Farrah, and J. Lukasik. 2002. Microbial source tracking: current methodology and future directions. Applied and Environmental Microbiology. 68(12):5796-5803.

Troy, M. S., S. Parveen, K. M. Portier, J. B. Rose, M. L. Tamplin, S. R. Farrah, A. Koo, and J. Lukasik. 2003. Geographical variation in ribotype profiles of *Escherichia coli*

isolates from humans, swine, poultry, beef and dairy cattle in Florida. Applied and Environmental Micobiology. 69(2):1089-1092.

Tukey, J. W. 1949. anova, an analysis of variance for vectors. Huma biology. 21:65-110.

United States Environmental Protection Agency. 1985. Test methods for *Escherichia coli* and enterococci in water by the membrane filter procedure. Environmental Monitoring and Support Laboratory, Cincinnati, OH. EPA-600/4-85/076.

United States Environmental Protection Agency. 1986. Ambient water quality criteria for bacteria, 1986. Report EPA440/5-84-002. United States Environmental Protection Agency, Washington, D.C.

Van den Bogaard, A. E, N. London, C. Driessen and E. E. Stobberingh. 2001. Antibiotic resistance of faecal *Escherichia coli* in poultry, poultry farmers and poultry slaughterers. Journal of Antimicrobial Chemotherapy. 47:763-771.

Van den Bogaard, A. E., and E. E. Stobberingh. 1999. Antibiotic usage in animals: impact on bacterial resistance and public health. Drugs 58:589-607.

Verbyla, D. 1986. Potential predection bias in regression and discrimiannt analysis. Can. J. For. Res.16:1255-1257.

Verner, J., M.L. Morrison and C. J. Ralph. 1986. Wildlife 2000-modeling habitat relationships of terrestrial vertebrates. University of Wisconsin press, Madison, Wisconsin, USA.

Wald, A. 1944. On a statistical problem arising in the classification of an individual into one of two groups. Annals of Mathematical Statistics. 44:229-233.

White, D.G., C. Hudson, J. J. Maurer, S. Ayers, S. Zhao, M. D. Lee, L. Bolton, T. Foley, and J. Sherwood. 2000. Characterization of chloramphinicol and florfenicol resistance in *Escherichia coli* associated with bovine diarrhea. Journal of Clinical Microbiology. 38(12):4593-4598.

Whitlock, J. E., D. T. Jones, V. J. Harwood. 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Research. 36:4273-4282.

Wiggins, B. 1996. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Applied and Environmetal Microbiology. 62(11):3997-4002.

Wiggins, B. A., P.W. Cash, W. S. Creamer, S. E. Dart, et al., 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Applied and Environmental Microbiology. 69(6):3399-3405.

Wiggins, B. A., R. W. Andrews, R. A. Conway, C. L. Corr, E. J. Dobratz, D. P. Dougherty, J. R. Eppard, S. R. Knupp, M. C. Limjoco, J. M. Mettenburh, J. M. Rinehardt, J. Sonsino, R. L. Torrijos and M. E. Zimmerman.1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Applied and Environmental Microbiology. 65(8):3483-3486.

Williams, B. K. 1983. Some observations on the use of discriminant analysis in ecology. Ecology. 64:1283-1291.

Williams, B. K. 1988. Assessment of sampling stability in ecological applications of discriminant analysis. Ecology. 69(4):1275-1285.

Witte, W. 1998. Medical consequences of antibiotic use in agriculture. Science. 279: 996-997.

Witte, W., H. Tschape, I. Klare and G. Werner. 2000. Antibiotic in animal feed. Acta Veterinary Scandinavica. 93 (supplement):37-45.

World Health Organization. 1997. The medical impact of the use of antimicrobials in food animals. Report of a WHO meeting. WHO/EMC/ZOO/97.4.

Zhao, S., D.G. White, B. Ge, S. Ayers, S. Friedman, L. English, D. Wagner, S Gaines, J. Meng. 2001. Identification and characterization of Integron-mediated antibiotic resistance among shiga toxin producing *Escherishia coli* isolates. Applied and Environmental Microbiology. 67(4):1558-1564.


MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 02504 4599