Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Allison McKenzie White Date

Relationships among Microbial Indicators of Fecal Contamination on Produce Farms

By

Allison McKenzie White
Master of Public Health

Department of Epidemiology

_____ [Chair's signature]
[Juan Leon]
Committee Chair

_____ [Member's signature]
[Faith Bartz]
Committee Member

Relationships Among Microbial Indicators of Fecal Contamination on Produce Farms

By

Allison McKenzie White

B.A., The Johns Hopkins University, 2007

Thesis Committee Chair: Juan Leon, PhD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in the Department of Epidemiology
2014

Abstract

Relationships Among Microbial Indicators of Fecal Contamination on Produce Farms
By Allison McKenzie White

Background: Produce is the leading single ingredient cause of foodborne outbreaks and illnesses in the US, and detecting and preventing contamination before it reaches consumers is essential. Microbial indicators, which are related biologically and ecologically to foodborne pathogens, are used to test for potential routes of pathogen contamination along the fecal-oral pathway. **Goal**: This study examines the relationships between the presence and concentrations of four microbial indicators of fecal contamination in agricultural samples and how they differ based on sample type, produce type, and stage in the production process. **Methods**: Hand-rinse, soil, water, and produce, samples collected from 11 tomato, jalapeño, and cantaloupe farms in Mexico were tested for the presence and concentration of pathogens and four microbial indicators of fecal contamination (fecal coliforms, *Enterococcus*, generic *E. coli*, and somatic coliphage). Linear and logistic regression, as well as chi-squared and Spearman's correlation were used to assess the relationships between the microbial indicators on each sample type, controlling for potential effect modifiers (produce type, time of sample collection, and stage in the production process). **Results**: Significant relationships between concentrations of *Enterococcus* and fecal coliforms were observed across all sample types Several other significant relationships between microbial indicators were observed with no apparent trends by either sample type or indicator-pair. **Conclusions**: These results provide evidence to support the idea that, while some indicators may be related in certain settings, unknown factors influence the presence of indicators in the agricultural environment and using multiple indicators may be the best way to test for the presence of fecal contamination in produce farms. **Implications**: These results should be considered as safety guidelines and standards for produce are being developed, since the choice of indicator can have implications on whether fecal contamination is being identified, and whether the potential presence of various types of pathogens is being accurately predicted.

Relationships Among Microbial Indicators of Fecal Contamination on Produce Farms

By

Allison McKenzie White

B.A., The Johns Hopkins University, 2007

Thesis Committee Chair: Juan Leon, PhD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in the Department of Epidemiology
2014

Contribution of student

This work herein is the product of a secondary data analysis performed by the student. The student did not have a role in the collection of data. However, the student did perform all work after data collection independently, including the analysis of the data, the construction of regression models, summation of results in tables, and all writing. Advisement throughout this process was provided by the student's thesis advisors.

Table of Contents

Literature Review

Each year, one in six Americans, or almost 50 million people, become ill from a foodborne related illness (1). According to the US Centers for Disease Control and Prevention (CDC) estimates, three thousand people are killed and 128,000 more are hospitalized from foodborne illnesses each year (2). Produce is the leading single ingredient cause of foodborne outbreaks and of illnesses in the US, causing nearly 25% (over 25,000) of all foodborne illnesses (1). Since the year 2000, multiple produce items have been involved in outbreaks throughout the US —including tomatoes, sprouts, jalapeño and serrano peppers, lettuces, cantaloupes, papayas, mangoes, spinach, and cucumbers (3).

Produce related illnesses are caused by an equally wide mix of bacterial, viral, and parasitic pathogens—including multiple strains of *Salmonella,* several *Escherichia coli* serogroups, norovirus, Hepatitis A, and *Cyclospora* (3). In a CDC study of outbreak-related foodborne illnesses from 1998 to 2008, 60% of viral, 30% of parasitic and 27% of bacterial illnesses were attributed to produce (4). Between 2005 and 2006, four separate outbreaks spanning multiple states were caused by three different strains of *Salmonella*, infecting at least 429 people in 21 states (5). Many cases of *Salmonella*, particularly in adults, do not require hospitalization, and even fewer have a specimen tested, thus, the majority of *Salmonella* cases are not reported and lab confirmed (6-8). Because of this, the overall reach of this outbreak, and others involving *Salmonella* may have actually been much

higher (8). Epidemiological analyses and lab testing linked these specific outbreaks to tomatoes consumed in restaurants. Fields, farms, and/or packing facilities located across the US (Virginia, Florida, and Ohio) were implicated in three of the four outbreaks (5). Another, more recent, example of produce-related outbreaks involves contamination by *Listeria monocytogenes* in cantaloupes. This 2011 outbreak infected 147 people in 28 different states (9). Cantaloupes had previously been associated with outbreaks of *Salmonella* in 2008, 2011, and in 2012 (3). Produce-related outbreaks lead to a significant public health burden nationwide, and also have substantial negative economic impacts.

A 2008 multistate outbreak of *Salmonella* on jalapeño and serrano peppers-and potentially tomatoes-was traced to farms and a packing facility in Mexico (8). This lead to temporary FDA recommendations to avoid peppers grown, harvested, or packaged in Mexico (8). During the initial stages of the outbreak investigation, before the source had been identified, the FDA recommended that consumers in some areas of the US avoid raw tomatoes grown domestically (10). An analysis of the economic impact of the regulations regarding tomatoes alone estimates total losses of hundreds of millions of dollars—figures which do not include the economic and trade impact of the later recommendations regarding peppers grown in Mexico (11). These figures highlight the impact that produce related outbreaks can have, not only on human health, but on the agricultural community as well as both domestic and international economies. These issues emphasize the need for effective ways to prevent produce outbreaks by detecting and preventing contamination before it reaches consumers.

Contamination that starts at the farm (pre- or post-harvest) or processing levels has the potential to spread over a wide geographic area (i.e. lead to multi-state outbreaks) and must be traced back from cases to the ultimate source(12). The trace back investigation process can take time and often involves laboratory testing to identify pathogen subtypes such as pulse-field gel electrophoresis (PFGE), which can be timely and costly (12-14). The increasing intricacy of the global food chain makes outbreaks more difficult to trace and contain, which can result in more illnesses (15). Thus, prevention of contamination is essential in reducing outbreaks.

How does produce become contaminated?

Foodborne pathogens are typically transmitted to produce through various routes in the fecal-oral pathway, whereby infection of a new host occurs as a result of consumption of contaminated feces from an infected warm blooded animal (including humans) (7, 16). Produce can become contaminated at any point in the food production process, from before the produce is harvested to steps involving food preparation and serving. Looking specifically at the farm and production stages, contamination can occur at almost any stage: before produce is harvested, at the time of harvest, while produce is being transported, or during cleaning, packaging, and preparation (16-18).

Furthermore, fecal contamination can be spread to produce through a variety of sources. These sources include agricultural water (used for irrigation, cleaning, and fertilizer mixing), soil amendments (such as manure), domestic and wild animals, human

contact, and equipment, tools, and buildings (17, 19). Interrupting these pathways of fecal contamination is crucial in preventing produce contamination and reducing foodborne illnesses. Although detecting the specific sources and most commonly implicated pathways can be difficult, given the multitude of steps and potential points for contamination, source detection is crucial as it allows for the development of targeted interventions and preventative measures (12). One common way of detecting contamination is through testing for pathogens or related microorganisms.

What causes these produce-related foodborne diseases?

The organisms commonly linked to produce-related outbreaks vary in their biological and ecological characteristics. Bacterial, viral, and parasitic organisms have all been linked to produce-related outbreaks. *Salmonella* is the most prevalent bacterial source of produce related outbreaks, followed by *E. coli* (3, 20). Norovirus is the leading viral cause of produce-related outbreaks, and is attributed to the highest number of illnesses overall (20).

Salmonella, which was attributed to 19% of all produce-related outbreaks between 1990 and 2004, is a genus of bacteria of the family Enterobacteriaceae (16, 20). Over 2,500 serotypes of *Salmonella,* not all of which are pathogenic, have been identified. Pathogenic strains can cause headache, fever, and diarrheal illness (7). Its reservoirs include domestic and wild animals and humans, where it lives in the gastrointestinal tract. Animals can often be chronic carriers (7). Food products are contaminated with *Salmonella* by coming in

contact with contaminated fecal matter (16). A wide variety of factors, including pH, acidity, moisture, competitive microflora, and nutrient content can impact the growth of *Salmonella* in food(16). Ideal growing temperatures for *Salmonella* are from 35-40°C; however, *Salmonella* can grow in temperatures from 2 to 54°C (16). Furthermore, *Salmonella* has been shown to be more resistant to extreme temperatures (both hot and cold) when previously exposed to similar sub-lethal temperatures (16). This ability to adapt quickly has wide ramifications for implementing control measures to prevent *Salmonella* outbreaks and monitor its potential spread in produce.

Escherichia coli is a species of bacteria which is also a member of the Enterobacteriaceae family. Six major strains of *E. coli* cause gastroenteritis in humans (7). *E. coli* O157:H7, which has been linked to a variety of foodborne outbreaks, is of the enterohemorrhagic strain(3, 16). Bacteria in this strain produce Shiga toxins and are often referred to as Shiga toxin-producing *E. coli* (STEC)(7, 16). STECs can cause severe diarrhea in humans and, in severe cases lead to hemolytic uremic syndrome (HUS) and death(7). Children are particularly vulnerable to developing HUS (7, 16). Like *Salmonella*, *E. coli* is found in the gastrointestinal tract of animals. Ruminant animals are the most common reservoirs of STEC; cattle are particularly important reservoirs(7). Humans may also serve as a reservoir. Most *E. coli* strains are nonpathogenic and "account for the majority of facultative flora found in the gastro-intestinal tracts of most vertebrates, including humans," a fact which is relevant in pathogen detection (16). Pathogenic *E. coli* is of particular danger as a produce contaminant, due to its ability to withstand harsh conditions including low pH, temperature stresses, and nutrient depletion, and its ability to persist in

the environment for long periods of time (16). It appears that pathogenic *E. coli* is most prevalent in fall and summer months(16).

Produce-related outbreaks can also be caused by viruses. Of all potential viral pathogens, norovirus is of greatest concern; between 1990 and 2004 norovirus alone was the cause of 39% of all produce-related outbreaks and over 9,000 illnesses, higher than any other organism (20). Norovirus, also called a Norwalk-like virus, is from the family Caliciviridae and is an RNA virus with size ranging from 27-40 nm in diameter (much smaller than a bacterial agent)(16). Norovirus can cause severe diarrhea sometimes accompanied by nausea, vomiting, and other symptoms (7). The virus lives in the human gastrointestinal tract (7). As with the organisms above, norovirus is transmitted through the fecal-oral pathway; however, unlike the other pathogens discussed, human fecal matter is the primary source of the pathogen (since humans are the only know reservoir of human norovirus) (7). Norovirus can also be transmitted person to person through contamination from fecal matter or vomitus, which can make detecting the precise source of the virus difficult (16). Norovirus is a nonenveloped virus, meaning it lacks a lipid envelope. Nonenveloped viruses are hardier than enveloped viruses, allowing them to persist for longer periods of time in the environment (16). In one study, norovirus remained infectious in groundwater for over 60 days (21).

To ensure produce safety, a number of methods have been developed to test for the presence of these disease-causing pathogens on produce (22, 23). There are three primary types of tests used to search for pathogens: cell culture methods, microscopy, and nucleic acid based detection, all of which have specific advantages and disadvantages (22). Cell-

culture methods, for example, cannot be performed on norovirus as it cannot be cultured (22). Microscopy is appropriate mainly for cysts of parasites, which are larger and easier to view than bacteria or viruses. Nucleic acid based detection is generally accomplished by polymerase chain reaction (PCR) and is often used for virus detection (22). However, PCR methods can detect non-infectious viruses and the reaction may be inhibited by certain compounds found in produce and the environment (22). While detection methods are improving, detecting specific pathogens can be difficult, time-consuming, and expensive (24, 25). Furthermore, detection of pathogens in water and agricultural samples is quite rare, and the lack of a pathogen in one specific sample does not exclude the possibility of contamination (24, 25). Since there are a number of pathogens which can contaminate produce it is unfeasible to test for every pathogen (26), testing for specific pathogens is generally considered nether the best nor the most effective way to check environmental samples for fecal contamination.

Microbial Indicators

Because of the above mentioned constraints in pathogen testing, other fecal organisms are often used to indicate risk of potential pathogen contamination. These organisms, called microbial indicators, are related to disease-causing pathogens both biologically and ecologically (26). For example, fecal coliforms, which are one of the most commonly used microbial indicators, are from the same family of bacteria as *E. coli*, which is also a coliform (27). Initially, the choice of which indicator to use was guided by the available technology (28). Total coliforms were one of the first indicators used, and

represent the total number of coliform-type bacteria present in a sample (28). However, this was later refined to test for fecal coliforms specifically, as it was felt that this would give a better suggestion of possible pathogen contamination (28). Fecal coliforms are a subset of coliforms that are thermotolerant—i.e. they produce gas from lactose at 45.5°C within 2 days—and, as such, are able to live and reproduce in feces, making them more precise indicators of potential fecal contamination than total coliforms (28, 29).

Currently, fecal coliforms, *Enterococcus*, and generic *E. coli* are commonly used fecal indicator bacteria (30). More recently, bacteriophages (viruses which infect bacteria), have been used as microbial indicators (26, 27, 31, 32). Somatic coliphage, a viral indicator which infects *E. coli*, has similar survival and reproductive patterns as several enteric viruses and, as such, has been suggested as a marker of potential pathogenic virus contamination (25, 32).

Microbial indicators have been used since the late 19th century, when water in London was first tested for bacteria (28). Since then, technologies for enumerating indicators have made many improvements. The multiple-tube fermentation procedure, also called the most probable number (MPN) procedure was one of the first tests developed, and involves estimation of bacterial counts based on production of gas and statistical procedures (28). Since then, membrane filtration and culture methods, processes in which bacteria are concentrated on membranes and grown on culture media, have become the more common method (28). There have been numerous advances in the membranes and media used for enumeration of specific bacteria (22). For viral indicators, a number of tests have been developed, including modified membrane filtration methods, PCR techniques,

and counting plaque-forming units; these tests continue to be refined for additional virus types (22, 28, 33).

Indicators and Pathogen Relationships

Despite the widespread use of microbial indicators as measures of fecal contamination and indicators of potential pathogen contamination, there is no clear scientific consensus on how microbial indicators relate to pathogens in the environment, particularly in agricultural settings (24, 30, 34). Ideal indicators are easily detected, present in fecal matter, and have environmental survival and reproductive patterns similar to pathogens (30). However, no indicator meets all of these criteria in relationship to every pathogen, therefore no single ideal exists. Thus, there is a need to identify the 'best' indicator for predicting risk of pathogen contamination (22, 35).

Several studies have sought to elucidate the relationships between specific pathogens and their possible indicators; few have found these relationships to be clear. A study in French rivers and wastewater treatment plants found that there was no correlation between the presence of the indicators *Enterococcus* or fecal coliforms and *Salmonella*, although they did find a relationship between *Cryptosporidium* and both indicators in riverine water samples only (36). A similar study looked at relationships among several pathogens and the indicators total coliforms, fecal coliforms, and *Enterococcus* in river and coastal waters in Greece. Although indicators were present in all samples where pathogens were found, the study found no significant statistical relationships between any indicators and any pathogens (23). A Canadian study, again

using water samples, found that overall relationships between indicators (total coliforms, fecal coliforms, *E. coli, Entercococcus*, and *Clostridium perfringens*) and pathogens (*Listeria monocytogenes, E. coli* O157:H7, *Salmonella,* and *Campylobacter*) were weak. The study also found positive significant (yet weak) relationships between *Salmonella* and the indicators. Less than 3% of samples with a pathogen did not have any indicator and classification and regression trees (CART) analysis found that fecal coliforms and *E. coli* were the most appropriate predictors for the pathogens studied (30). However, in a separate study in marine, estuary and river waters, significant relationships of varying strength were observed between *Salmonella* and total coliforms, fecal coliforms, fecal streptococci, and coliphage (32). Another study showed no evidence of correlation between somatic coliphage and enteroviruses in farm pond water and sewage, and enteroviruses were present in effluents absent of coliphage (25). A review of indicator and pathogen relationships in biosolids found a lack of correlation between pathogens and their biological indicators, leading the authors to conclude that in many settings indicators may be better utilized in testing for the efficacy of a treatment process, for example, than in testing for a specific pathogen (22).

A comprehensive review of research that modeled the relationships between indicators and pathogens, controlling for intervening variables, found no single indicator that correlated most often with pathogens (37). However, several were more likely to be correlated than other commonly used indicators, including coliforms and coliphage. For all significantly linearly correlated cases used in the review, across all indicator types, the average overall correlation was 0.554 (37). Coliphage correlated better with viral

pathogens (adenoviruses, astroviruses, hepatitis A virus, noroviruses, sapoviruses, enteroviruses, human enteric viruses, rotavirus) than did the bacterial indicators (total coliforms, fecal coliforms, fecal streptococci, Enterococci, *E. coli, Clostridium perfringens,* heterotrophic bacteria,), while also correlating with bacterial pathogens (*Aeromonas, Campylobacter, Helicobacter pylori, Salmonella, Staphylococcus aureus, Vibrio cholerae,* total vibrio*, Pseudomonas aeruginosa)*. This study also found that the number of samples testing positive for a pathogen was the most important predictor in determining indicator-pathogen correlation (37). As discussed above, pathogens are difficult to detect, suggesting that the lack of evidence of association between indicator and pathogens may be due, in part, to insufficient sample size (37). These relationships highlight the need for a deeper understanding of how all indicators function in the environment and emphasize their strengths and weaknesses for potential use in testing contamination pathways and monitoring microbial quality during agricultural processes.

Indicator Relationships

Due in part to the varied relationships between indicators and pathogens and the environmental factors that can contribute to this variation, testing for multiple indicators has been recommended (24, 30, 31, 37). However, there is no consensus on which indicator(s) need to be considered and how these might vary based on ecological factors. Understanding the relationships between indicators is important both in guiding regulations on produce safety and as well as identifying routes of contamination. A single

indicator may not be indicative of total contamination, since microorganisms differ in their ability to persist and replicate in different environments (24, 30, 31). However, if some indicators behave similarly in certain environments or sample types, it may be necessary to test for only one of those indicators, which could simplify and reduce the expense of environmental testing.

As with the relationships between pathogens and indicators, several studies have examined the relationships indicators have to one another. A study of surface waters in an agricultural area found weak to moderate non-significant correlations between *Enterococcus* and total coliforms, fecal coliforms, and *E. coli*, yet found that the latter three were all significantly related to each other. The strongest correlation existed between *E. coli* and fecal coliforms (rho=0.79), providing some support for the idea that some indicators may be interchangeable, depending on the sample type (23). An additional study examining river and stream samples from an agricultural area in Canada, found similar results with moderate to strong correlations between total coliforms, fecal coliforms, *E. coli*, and *Enterococcus* (rho range: 0.74-0.82) (30). A separate study in groundwater aquifers found that there was a significant difference in the percentages of detection for somatic coliphage and *E. coli*, lending support to the idea that a viral indicator might be useful in addition to bacterial indicators (31).

While it appears that some indicators are related in certain environments, the results are not definitive, and the majority of evidence comes from water samples. Although studies have used indicators to assess microbiological contamination on farms, studies examining indicator relationships on farms are particularly limited (24, 38-40). A

study examining the relationship between fecal coliforms and *E. coli* on a variety of food samples (including produce) from grocery stores found that there was no significant difference in concentrations between the indicators in any sample type, with correlation coefficients ranging from 0.78 to 0.99, providing evidence that some indicators may be interchangeable in terms of the information they provide regarding fecal contamination in agricultural environments (29). Another study examined the relationships between indicators in soil, crop, and water samples on a lettuce farm in Belgium (24). Across all sample types, a low (yet significant) correlation was observed between the indicators *E. coli* and total psychrotropic aerobic plate count (TPAC). In irrigation water samples alone, however, stronger (0.437-0.918) significant correlations were observed between all pairs of the four indicator organisms studied (TPAC, *E. coli,* coliforms, Enterococci) (24).

These differences in relationships by sample type underscore the need to further understand how these relationships change in different types of agricultural samples: specifically, how viral indicators relate to bacterial indicators on produce farms. This relationship has important implications for produce safety testing and regulations.

Regulations

One of the reasons microbial indicators are of such interest is to ensure safety and quality standards in a variety of environmental samples. The most common examples of this come from water standards. Most of these standards set acceptable concentration levels of the chosen indicators—above which a sample is considered contaminated. The levels generally have a maximum value allowed in any one sample and a (lower) max value

allowed for the geometric mean concentration across multiple samples. Examples of government quality standards which have been established using indicators include the US Environmental Protection Agency's (EPA) recreational water quality (e.g. beach/swimming water) standards which recommend testing for *E. coli* and *Enterococcus* with a geometric mean acceptable level of 126 and 35 colony forming units (cfu) per 100 mL of water over any 30 day period (41). These standards, which have been in place and undergone revisions for several decades, were developed and refined based on epidemiological studies of gastrointestinal illness, as well as through extensive lab testing to identify the concentrations that present an unacceptable level of risk (41).

The US Food and Drug Administration (FDA) also has regulations on water quality with standards based on EPA research (42). The FDA requires testing of bottled water for total coliforms and, if coliforms are detected, a follow-up test for *E. coli* is performed, since *E. coli* is not permitted in bottled water (42).

Until recently, no similar standards existed for the agricultural industry. In 2011, the Food Safety Modernization Act (FSMA) was created by the FDA (17). A part of the FSMA is the Produce Safety Rule, which focuses on 5 main areas of microbial contamination of produce (agricultural water, biological soil amendments of animal origin, health and hygiene, domesticated and wild animals, and equipment, tools and buildings) and creates science-based standards to reduce the risk of produce contamination (17). The current Produce Safety Rule sets specific indicator-based standards for agricultural water and for preparation methods for manure (non-indicator-based standards exist in the other areas)(43). These standards are based on the existing evidence, much of which comes from

studies of water. New evidence from agricultural environments is needed to ensure that present and future standards are based on the most relevant evidence.

Produce-related foodborne outbreaks contribute to a large burden of disease. Preventing these illnesses by interrupting the fecal-oral pathway and preventing produce contamination is a complex process. A key step in this process is using appropriate methods to detect fecal contamination. Due to the aforementioned low prevalence of and challenges in detection of pathogens, alternative methods, such as the use of microbial indicators for determining fecal contamination must be used. While there is a large body of epidemiological data and research studies on water quality, insufficient evidence exists to determine how microbial indicators of fecal contamination relate to one another in farm environments.

Many current standards have been adapted from water quality regulations; these are insufficient, however, since the behavior and relationships of both pathogens and microbial indicators may differ on produce, in soil or manure, on hands, or on tools and equipment. Understanding how commonly used indicators behave on produce and in the agricultural environment is essential in the development of produce specific standards and regulations to prevent illness. While the primary goal of determining appropriate and necessary indicators is to reduce produce-related illnesses, understanding clearly which indicators are related can also minimize the labor and expense of testing for multiple highly correlated pathogens. Thus, clear knowledge of the relationships between microbial indicators benefits regulators, consumers, and commercial farmers.

This study aims to characterize the relationships between the presence and concentration of microbial indicators of fecal contamination (fecal coliforms, *Enterococcus*, generic *E. coli*, and somatic coliphage) on 11 produce farms in Mexico, and determine if and how those relationships differ based on potential effect modifiers including sample type, time, and point in the production process. Understanding what, if any, relationships exist between indicators can advance food safety research, help inform regulations on produce safety, and ultimately prevent illness by making our food supply safer through reducing pathogen contamination.

Introduction

Each year, one in six Americans, almost 50 million, people become ill from a foodborne related illness (1). According to the US Centers for Disease Control and Prevention (CDC) estimates, 3,000 people are killed and 128,000 more are hospitalized from foodborne illnesses each year (2).

Produce is the leading single ingredient cause of outbreaks and of illnesses in the US, causing nearly 25% (over 25,000) of all foodborne related illnesses(1). Produce has been the recent source of several high profile outbreaks in the US. Between 2005 and 2006, four multistate outbreaks causing illness in over four hundred people were linked to *Salmonella* contamination of tomatoes supplied to restaurants from multiple farms in the US (5). In 2008, an outbreak of *Salmonella* infected 1,442 people across 43 states. The outbreak was linked to jalapeño and serrano peppers grown in Mexico (6, 8). A 2011 outbreak that infected 147 people across 28 states was traced to the bacteria *Listeria monocytogenes* on cantaloupes grown on a farm in Colorado (9).

While some of these produce-related outbreaks can be linked to specific farms, the mechanisms of contamination of produce with disease-causing agents are often unclear. Because of the high burden of disease from produce-related foodborne illness, there is a need to understand how the produce is becoming contaminated and how to prevent this contamination (1).

Produce can become contaminated by a variety of bacteria, viral, and parasitic organisms, although bacteria and viruses collectively cause the vast majority of the

outbreaks (20). *Salmonella, E. coli* O157:H7, *Shigella,* and norovirus (Norwalk-like viruses) are the pathogens which pose the greatest risk for foodborne produce outbreaks, based on historical data (20). Since 2010, there have been several high profile *Listeria* outbreaks, making it another pathogen of particular interest (3). These human disease-causing pathogens are found in human and animal fecal matter and are transmitted to produce through the fecal oral pathway (i.e. by ingesting produce which has been contaminated with pathogen-containing fecal matter). Soil, water, manure, animal contact, and human contact are the main modes of fecal contamination of produce (16).

To ensure produce safety, a number of methods have been developed to test for the presence of these disease-causing pathogens on produce (22, 23). Detecting specific pathogens is difficult, however, and their prevalence on the produce itself as well as in potential sources of contamination (e.g. irrigation water) is quite rare (32, 35). Furthermore, it is logistically infeasible to test for every potential pathogen since a veritable multitude of pathogens can cause produce-related outbreaks (22, 44). Finally, the absence of one specific disease-causing pathogen does not indicate the lack of potential for other pathogen contamination and transmission of foodborne illness.

Because of these pragmatic and logistical constraints, microbial indicators of fecal contamination are frequently used to test for evidence of fecal contamination, and risk of potential pathogen contamination. Fecal coliforms, *Enterococcus,* and generic *E. coli* are commonly used fecal indicator bacteria (30). These microbial indicators are typically organisms (bacteria or viruses) that are ecologically and biologically related to the disease-causing pathogens (35, 44). For example, fecal coliforms, which are one of the most

commonly used microbial indicators, are from the same family of bacteria as *E. coli*, which is also a coliform.

Initially, the choice of which indicator to use was guided by the available technology. Total coliforms were one of the first indicators employed, and represent the total number of coliform-type bacteria present in a sample (28). However, this was later refined to test for fecal coliforms specifically, (i.e. coliforms that live in the gut and generally live and reproduce in feces) as it was felt that this would give a better suggestion of possible pathogen contamination (30). More recently, bacteriophages (viruses which infect bacteria), have been used as microbial indicators (22, 26, 27, 32). Somatic coliphage, a viral indicator which infects *E. coli,* has similar survival and reproductive patterns as several enteric viruses and, as such, has been suggested as a marker of potential virus contamination (22, 28, 32).

For decades, microbial indicators have been used in setting quality and safety standards. Since these indicators (bacterial and viral alike) have reproductive and survival patterns similar to the pathogens for which they are being used as proxies, the presence of a microbial indicator in a sample therefore suggests contamination with fecal matter of that sample (25, 27) . However, as some indicators (as well as certain pathogens) can replicate outside of the animal gut (their typical environment), and have different susceptibility to decontamination methods, the validity of certain indicators has been questioned (e.g. a microbe which can readily multiply outside of the animal gut may not be the best indicator of fecal contamination) (45, 46). Because indicators often represent larger categories of which these pathogens are a subset (e.g. pathogenic *E. coli* represents only a few subtypes

of generic *E. coli*, which in turn is a small portion of fecal coliforms) and eliminating all traces of microorganisms from a sample is unrealistic, small concentrations of microbial indicators are often deemed acceptable in quality and safety standards (27, 41).

Several government organizations use microbial indicators in setting health standards, primarily for water quality. The US Environmental Protection Agency (EPA) has been regulating recreational water quality (e.g. beach/swimming water) through the use of microbial indicators for several decades (41). The 2012 Recreational Water Quality Criteria recommends testing for *E. coli* and *Enterococcus* with a geometric mean acceptable level of 126 and 35 colony forming units (cfu) per 100 mL of water (41).

These standards were developed and refined based on epidemiological studies of gastrointestinal illness as well as through extensive lab testing (41). Similarly, the US Food and Drug Administration regulates the quality of bottled water, and bases those standards on EPA research and recommendations (42). The FDA requires testing of bottled water for total coliforms and, if coliforms are detected, a follow-up test for *E. coli* is performed. *E. coli* is not permitted in bottled water (42).

In 2011 the FDA signed into law The Food Safety Modernization Act (FSMA) (47). A portion of this Act is the Produce Safety Rule, which calls for the creation of science-based standards to reduce the risk of produce contamination (17). The Produce Safety Rule focuses on 5 main areas of microbial contamination of produce: agricultural water, biological soil amendments of animal origin, health and hygiene, domesticated and wild animals, and equipment, tools and buildings (43). The FSMA sets standards for agricultural

water and manure; these standards are based on existing evidence and may be modified as the body of evidence for contamination on farms grows (17).

Testing for multiple indicators has been recommended in the literature, recognizing that a single type of indicator may not be indicative of total contamination, since pathogens (and their related microbial indicators) vary in their ability to persist outside of their host environment, and behave differently based on a multitude of environmental factors (24, 30, 31, 37). However, as suggested by the differing governmental regulations cited above, there is not a consensus on which microbial indicators need to be considered, and how that might vary based on ecological factors. Studies of indicator-pathogen relationships have found varying relationships of differing strength, with no clear overarching relationships (22, 23, 30, 32, 37). Furthermore, insufficient evidence exists to determine how microbial indicators relate to one another and how those relationships are impacted by time, potential contamination source (e.g. water v. soil), and/or other environmental conditions, particularly in the field of agricultural and produce safety (23, 24, 29-31).

Many current standards have been adapted from water quality regulations;, the behavior and relationships of both pathogens and microbial indicators may differ, however, for contamination on produce, in soil or manure, or through packing equipment (22, 24). Thus, there is a need to better understand how these microbial indicators behave on produce farms. Understanding what, if any, relationships exist between indicators can help inform regulations on produce safety.

This study aims to characterize the relationships between the presence and concentration of microbial indicators of fecal contamination (fecal coliforms, *Enterococcus*, generic *E. coli*,

and somatic coliphage) on 11 produce farms in Mexico and determine if and how those relationships differ based on potential effect modifiers, including sample type, time, and point in the production process. Knowing how indicators relate to one another in the farm setting can guide the development of appropriate produce safety recommendations and regulations. More accurate, science-based regulations can decrease produce-related foodborne outbreaks and the resultant public health and economic impacts.

Materials and Methods

Study area

The study area comprised the Mexican states of Nuevo León and Coahuila on the United States-Mexico border. Eleven farms and three packing sheds participated in this study: five farms produced cantaloupes, five produced jalapeño peppers, and five produced tomatoes (four of which were also included as jalapeño farms). Institutional review board approval was received by the lead institution (Emory University) covering the duration of the study (approval number IRB00035460).

Sample collection

Samples were collected from May 2011 to December 2012. During each sampling event, 8 to 10 samples were obtained consisting of soil, water from the irrigation source and/or field irrigation lines, fresh produce rinses at several points in the production process (before harvest, during harvest, just prior to distribution from the field, and during packing if a packing facility was present on the farm), and of hand rinses from the pickers/packers. For each final sample, three samples from random locations were combined to form a single composite sample. The collection protocol, which differed for each sample type, is described below. All samples were placed on ice after collection, driven to the laboratory at the Universidad Autónoma de Nuevo León (UANL), and stored at 4°C until processing for microbial indicator analyses. At the time of collection, data identifying the field, farm, type of produce grown, date of sample collection, and point in chain (e.g. before harvest, during packing) were recorded for each sample.

Soil. At each random sampling location on the field, seven ~15g scoops of soil were collected to form a minimum sample of 100g. Samples were collected using the same 2 oz. scoop inserted to a depth of 5 cm below the soil surface and were transferred to the same Whirl-Pak bag (Nasco, Ft. Atkinson, WI). All three samples were combined in the laboratory using aseptic technique into a 300g composite sample. Before microbial analyses, 25 g soil was suspended in 0.15% sterile peptone water (PW) (75 ml PW for samples collected May 2011-June 2011, 225 ml for samples collected July 2011-December 2012), shaken vigorously for 60 s, and allowed to settle for 10 min. Rinse solution was used for later analyses.

Water. Water samples were collected from the source well and from the irrigation lines on the field. Well water samples were collected by first disinfecting the pump with 200 ppm hypochlorite. The pump was allowed to run for 30 s before three 1.5-liter water samples were collected in Whirl-Pak bags. Irrigation water samples were collected as close as possible to the harvest row where the drip tape deposited irrigation water or from the center of the distribution system when this was not possible and were collected in the same manner as well water. All three of the well or in-field irrigation water samples were combined to create a composite sample of ~4.5 liters, which was then re-divided into smaller subsamples for specific microbiological testing.

Produce rinses. Multiple fruits were rinsed in the same solution to create a single produce rinse sample. Specifically, 2 melons, 18 tomatoes, or 14 jalapeño peppers were rinsed to comprise one sample. For preparation of the rinses, half of each batch of produce was placed in a Whirl-Pak bag containing 500 ml PW, shaken for 30 s, massaged for 30 s,

and shaken again for 30 s. The first half of the produce batch was removed and replaced with the second half, and the process was repeated. This process was done three times with three different produce batches, and the rinses were combined to create a composite sample of 1,500 ml. The composite sample was divided into smaller subsamples for microbiological testing.

Hand rinses. Before sample collection, researchers obtained written consent from farm managers and oral consent from farm workers to collect a hand rinse sample per IRB approved protocol. The worker placed his or her hand in a Whirl-Pak bag containing 750ml PW. The worker was asked to shake the hand for 30 s, and then the hand was massaged for an additional 30 s. The first hand was removed from the PW, the second hand was placed in the same bag, and the process was repeated. Three individual hand rinse samples were combined to create a composite sample of 2,250 ml that was divided into smaller subsamples for specific microbiological testing.

Microbial indicator Analyses

All samples were tested for four microbial indicators of fecal contamination: generic *Escherichia coli, Enterococcus,* fecal coliforms, and somatic coliphage. Bacterial samples were assayed within 13 days of receipt (with the majority of samples assayed within two days of receipt). Coilphage samples were assayed within one month of receipt (<7 days was more typical). Samples were analyzed for all three bacterial indicators (generic *E. coli, Enterococcus*, and fecal coliforms) at UANL using a membrane filtration method to concentrate and incubation on defined media to enumerate. For each sample, a range of

volumes (effective original sample volume, 1 µl to 250 ml) were vacuum filtered through a 47-mm diameter, 0.45-µm-pore-size S-Pack filter (Millipore, Billerica, MA). After filtration, the filter was removed and placed onto an agar plate. *Enterococcus* was grown on KF Streptococcus agar (Oxoid Limited, Basingstoke, Hampshire, UK), fecal coliforms and generic *E. coli* were grown on RAPID'E.coli 2 (Bio-Rad Laboratories, Inc., Hercules, CA). Coliphage tests were conducted using FastPhage MPN Quanti-tray (Charm Sciences, Inc., Lawrence, MA).

The average concentration (number of CFU per volume filtered) of the bacterial indicators in each sample was determined. For statistical analyses, samples with concentrations below the assay lower limit of detection (with no growth on any plates) were reported as 0.5 CFU per greatest volume filtered (one half of the limit of detection)(48). Samples with concentrations above the upper limit of quantification (with greater than 250 CFU on all plates) were reported as 500 CFU per smallest volume tested (two times the upper limit of quantification). For coliphage samples, most probable numbers (MPN) were calculated from the Quanti-tray per manufacturer instructions, and a similar approach was followed when concentrations were outside the quantifiable range (values of 0.5 limit of detection were assigned for samples with concentrations below the quantifiable range and values of twice the limit of quantification were assigned for samples with concentrations above the quantifiable range). The limits of quantification for each assay are as follows (lower, upper): fecal coliforms: (0.104, 21500000) cfu/mL; *E. coli* (0.116, 660000); *Enterococcus* (0.112, 108000000), Coliphage (0.01, 15.85) MPN/mL.

Statistical Analyses

Statistical analyses were performed using SAS 9.3 (SAS Institute Inc., Cary, NC) at an alpha level of 0.05. Microbial indicator concentrations were transformed on a $\log_{10}$ scale; all further analyses were conducted using log transformed data. To ensure adequate sample size, imputed values described above were included in analysis.

Analyses were conducted separately for each sample type (soil, water (both source and irrigation), hand-rinse, and produce) as indicator concentrations were not in uniform units across sample types. Geometric mean and associated confidence intervals, as well as prevalence, were calculated for each of the four indicators. Spearman's rank correlation analyses were performed to compare microbial indicator concentrations to one another. Pearson's chi-squared analyses were conducted to identify any univariate associations between indicator prevalence. To further evaluate the relationship between microbial indicators, multivariate logistic and linear modeling was conducted, which allowed assessing potential interaction while controlling for the potential effect modifiers produce type (tomato, jalapeño, cantaloupe), time of sample collection (pilot, year 1, year 2) and point in chain (before harvest, after harvest, at distribution, at the packing shed).

Logistic Regression

Univariate analyses were conducted to confirm the findings of the chi-squared analysis. For each sample type, models were constructed with one indicator as the outcome

variable and a single indicator as the predictor variable (i.e. *E. coli* prevalence = Enterococcus prevalence). This was repeated for all pairwise combinations of indicators.

Potential interaction by year of sample collection was also assessed in these univariate analyses to determine if the relationships between indicators were influenced significantly by the year of sample collection—which might be expected in years with drastically different climate patterns. Interactions were assessed for each indicator pair combination across all 4 sample types (48 total models) (see Appendix A)(49). The majority of models did not show any significant effects of interaction; for those which did (two models) there were no clear patterns. Since including these effects did not impact the overall significance or interpretability of these models, they were excluded from the final models.

Collinearity was assessed for all models prior to full analysis using condition indices (CNIs) and variance decomposition properties (VDPs) to ensure that the exposure indicators were not too closely related to one another, impacting model estimates. A CNI of greater than 30 with at least two corresponding VDPs greater than 0.5 was used as the cutoff in determining if colliearity issues existed (49). No collinearity problems were found and no variables had to be removed from the model (data not shown).

The final full logistic models included all microbial indicators as well as produce type and point-in-chain, in order to examine the relationships between presence of microbial indicators after controlling for produce type, point in chain, and the other indicators. For water and soil samples, point-in-chain was excluded since all samples were collected from the same point (immediately before harvest). Additionally, Enterococcus

was uninformative and therefore excluded from hand sample models due to 100% prevalence. Similarly, fecal coliforms were excluded from water sample models due to near 100% prevalence which did not allow for appropriate separation of data.

Linear Regression

Linear regression was conducted to examine the relationship between the concentration of microbial indicators after controlling for produce type, point in chain, and the other indicators. The strategy for linear regression was the same as that for logistic regression; univariate analyses were conducted to confirm the results of the correlation analysis; and interaction was assessed for sample collection year.

As with logistic regression, interaction terms in most models were not significant and not show significant effects there were no clear trends; thus, interaction terms were not included in the full linear models (see Appendix B)(50). Collinearity and regression diagnostics were conducted using partial plots and VIFs with VIFs greater than 10 indicating collinearity issues (51). No major issues with collinearity were found (data not shown). Not all variables met the normality assumption.

Full models included all microbial indicators as well as produce type. As with logistic models, point-in-chain was included for hand and produce samples and not for water and soil samples (which were always collected before harvest).

Results

Descriptive Statistics

To gather information on the general distribution of indicators across sample types, we calculated the prevalence, geometric mean indicator concentration, and associated 95% confidence intervals (Table 1). Produce samples were also stratified by type of produce and calculations were repeated (Table 2). Concentrations and prevalence of indicators varied greatly across sample types, however there were some trends. The prevalence of fecal coliforms was greater than 90% across all sample types. Concentrations of fecal coliforms also tended to be higher than those of other indicators. *Enterococcus* was also quite prevalent across sample types, with prevalence values ranging from 69% to 100%. When different types of produce were compared, melons consistently had higher concentrations and greater prevalence than jalapeños and tomatoes for all indicators. Overall, both concentrations and prevalence of all indicators varied across sample types, with fecal coliforms and *Enterococcus* generally showing the highest prevalence and concentration compared to *E. coli* and coliphage.

Indicator Concentrations

Spearman's correlation analyses were conducted as a preliminary approach to determine the presence and significance of correlations between concentrations of each of the four indicators of fecal contamination (Table 3). Concentrations of *Enterococcus* and fecal coliforms were significantly correlated across all sample types (p-value range:

<0.0001-0.0003) with rho values ranging from 0.38-0.65, indicating moderate correlations. This relationship persisted when produce samples were stratified by produce type.

Concentrations of *Enterococcus* and *E. coli* were significantly correlated for hand, water, and produce samples (p-values: all <0.0001, rho values 0.37-0.60). When produce samples were stratified by produce type, this relationship did not persist. For produce samples, all microbial indicators were significantly correlated with one another (p-values: all <0.0001, rho values 0.28-0.65). Several other significant correlations were observed and were not part of apparent trends (Table 3).

Linear regression was conducted to determine how the relationships between concentrations of the different microbial indicators differed within each sample type while controlling for other factors (point-in-chain, produce type, and the microbial indicators). Significant associations were found between concentrations of several indicators (Table 4). Concentration of *Enterococcus* was significantly associated with concentration of fecal coliforms across all sample types (p-value range: <0.0001-0.0007). Conversely, concentration of fecal coliforms also predicted concentration of *Enterococcus* regardless of sample type (p-value range: <0.0001-0.0007).

Enterococcus concentrations were significantly associated with *E. coli* concentrations in water and produce samples (p-value range: 0.0014-0.0204) and with coliphage concentrations in hand and produce samples (p-value range: 0.0008-0.0481). Concentration of fecal coliforms predicted concentration of *E. coli* in water and produce samples (p-value range: 0.0002-0.0096) and predicted concentration of coliphage in hand-rinse samples alone (0.0002-0.0242).

In summary, in produce samples, concentrations *Enterococcus* were significantly associated with concentrations of all of the three additional indicators (*E. coli,* fecal coliforms, and coliphage) (p-value range: <0.0001-0.0204) (i.e. all indicator variables included in the model were significant). This trend was not seen across any other indicator outcomes or sample types. Several models using concentrations of *E. coli* and coliphage showed significant associations; however there were no clear trends. For water and produce samples, *E. coli* concentration was a significant predictor for both *Enterococcus* and fecal coliforms samples but not for coliphage.

Indicator prevalence

Chi-squared analyses were conducted as a preliminary approach to determine the relationships between the presence of each of the four indicators of fecal contamination (Table 5). Among water samples, prevalence of *E. coli* was significantly associated with prevalence of *Enterococcus*, fecal coliforms, and coliphage (p-values: 0.0012-0.0115, OR: 4.08-8.8). Among hand samples, the prevalence of *E. coli* was significantly associated with prevalence of fecal coliforms and coliphage (p-values: 0.0027-0.0391, OR: 3.24 for coliphage sample, N/A for fecal coliforms). The relationship between *Enterococcus* and *E. coli* could not be calculated due to 100% prevalence of *Enterococcus*. No other trends were observed; other significant relationships are indicated in Table 5.

Logistic regression was conducted to determine how the relationships between the prevalence of the different microbial indicators differed within each sample type while controlling for other factors (point-in-chain and produce type and the microbial indicators)

(Table 6). There were no clear trends in significant ORs after controlling for point-in-chain and produce type. Some of the relationships observed in the chi-squared analyses persisted in the logistic regression. Due to 100% or near 100% prevalence not all estimates could be calculated.

The prevalence of *E. coli* was significantly associated with the prevalence of coliphage in hand-rinse and produce samples (OR: 3.77, p-value: 0.0025, and OR: 3.77, p-value: 0.0498, respectively). Conversely, the prevalence of coliphage was a significant predictor of the prevalence of *E. coli* in hand and produce samples (OR: 3.92, p-value: 0.0021, OR: 4.01, p-value: 0.0408, respectively). Other significant relationships are listed in Table 5.

Discussion

This analysis sought to determine the relationships between four microbial indicators of fecal contamination (fecal coliforms, *Enterococcus*, generic *E. coli*, and somatic coliphage) on produce farms and how those relationships differ based on sample type, time, and point in the production process. Across all sample types, there were significant associations between concentrations of fecal coliforms and *Enterococcus*. Results from linear and logistic regression confirmed the results seen in initial chi-squared and Spearman's correlation analyses. Although a few additional significant relationships were observed beyond the associations between fecal coliforms and *Enterococcus*, there were no additional trends either between indicator pairs or across sample types.

Key Results

Concentrations of *Enterococcus* and fecal coliforms were significantly, although modestly, associated across all sample type, after controlling for produce type, point-in-chain, and *E. coli* and coliphage concentrations (regression coefficients between 0.38 and 0.65). This association indicates that testing for both *Enterococcus* and fecal coliforms on produce farms may be redundant and that one indicator may be sufficient to indicate fecal contamination. Results from other studies do not show evidence of a clear relationship across all settings. One study which analyzed surface water contamination showed a significant relationship (r_s=0.74) between *Enterococcus* and fecal coliforms, while a separate study, also of surface water, found a non-significant correlation (r_s=0.34) between the same two indicators (23, 30). Moreover, study of recreational water found that the

correlations between concentrations of fecal coliforms and *Enterococcus* differed based on the season of sample collection, further complicating the relationship (34).

While *Enterococcus* and fecal coliforms possess different survival times outside the host and are impacted differentially by environmental factors, they remain closely related biologically and ecologically (22). Both fecal coliforms and *Enterococcus* are broader indicator terms which include numerous subgroups (i.e. fecal coliforms refers to the broad category of thermotolerant coliform-type bacteria while *Enterococcus* is an entire genus of streptococcal bacteria)(28). These indicator bacteria are subsequently detected in greater concentrations than indicators that represent only a subgroup (e.g. specific *E. coli* subtypes), which may, in part, explain this similar relationship (29). A wastewater study found similar mean concentrations of fecal coliforms and *Enterococcus* as well as the same overall frequency of detection in samples for the two indicators (46). The same study; however, found no significant correlation between the two indicators (46). Further research may need to be conducted on additional produce farms to confirm whether fecal coliforms and *Enterococcus* persist in similar concentrations across farms; if so, testing for one indicator may be sufficient to test for certain types of fecal contamination and for the potential presence of bacterial pathogens.

Both Spearman's correlation analyses and linear regression were similar in their utility to identify associations between indicators. While the correlation analyses showed more significant associations than did the linear regression analyses, both methods displayed the same overall trends in the identified relationships between concentrations of microbial indicators. That is, that fecal coliforms and *Enterococcus* concentrations were

significantly related, and that, among produce samples, microbial indicator concentrations were related across produce types. The identification of fewer significant associations by the linear regression analyses was likely due to the control variables used in the model, which helped to refine the associations observed by Spearman's correlation. Similarly, congruent patterns were observed between chi-squared analyses and logistic regression results. Again, the associations from the chi-squared analysis that did not emerge in the logistic regression were likely due to the use of control variables. This has important implications for future studies; firstly, unadjusted analyses may be sufficient to give a broad picture of how indicators and pathogens are behaving on produce farms. However, when using evidence to craft regulations and guidelines for produce studies, it is important to control for potential confounders; otherwise false relationships may be found. For example, studies have shown that levels of contamination vary by produce type (38, 52). For future research to be generalizable to a variety of produce farms and settings, it is important to control for these potential confounding variables.

Overall, few relationships and clear trends were shown across indicator-pairs or sample types, indicating that both the prevalence and concentration of microbial indicators differ within and across sample types on produce farms. Past literature has shown wide variation in the presence and concentrations of specific indicators in certain settings and that pathogen-indicator and indicator-indicator relationships differ based on the pathogen being tested, the indicator used, and various environmental factors (24, 31, 34, 37). Because of this variation, and to ensure that contamination from the full breadth of possible pathogens was being detected, testing for multiple indicators has been widely

proposed (22, 26, 27). This has particularly been the case with the use of viral indicators, such as coliphage (22, 26, 31, 32, 37). The lack of clear trends involving coliphage and any of the three bacterial indicators in our results provide further evidence that coliphage behaves differently than bacteria and that it may be not only appropriate but necessary to include a viral indicator in testing.

Our results further show that microbial indicators in samples from produce farms vary in prevalence and concentration, likely based on individual indicator characteristics and environmental factors. This contributes to the body of evidence that supports the testing for multiple indicators to prevent fecal contamination, and, in particular, the use of multiple indicator standards in produce safety guidelines. Additionally, future research to better understand the underlying relationships between indicator and pathogen behavior, particularly after controlling for site and environmental variables, should be conducted. Recognizing the environmental and farm-specific factors which influence concentrations and prevalence of indicators will provide better assistance in the development of produce safety rules and the subsequent prevention of foodborne illness.

Limitations

After all microbial indicator variables were log-transformed for analysis, some variables remained non-normally distributed. Spearman's rank correlation analyses (nonparametric tests appropriate for these data) were conducted in the preliminary analyses. However, linear regression, which allows for the introduction of control variables but requires a normality assumption, was used for subsequent analyses. The violation of

the normality assumption of linear regression may have led to less accurate results. However, as discussed above, similar patterns were observed between the Spearman's correlation analyses and the linear regression. This suggests that the violation of this assumption did not greatly impact the observed results.

The tests for microbial indicators have a lower limit of detection and an upper limit of quantification which must be accounted for in analysis. Due to sample size constraints, models could not be run using datasets restricted to contain only those indicator concentration values within the detection and quantification limits. For samples with concentrations above and below these limits in this study, values were imputed. These imputed values were used in analysis; however, they were not adjusted for in linear regression, due to the difficulties in accounting for the imputed values in both the exposure and outcome variables. These unadjusted methods may have impacted the results. Additional analyses using more complex methods for censoring of imputed values could be considered. Previous analyses using similar data found that, while the results garnered using a Tobit regression strategy to adjust linear models (an adjustment option for left-censored data) were more accurate than simple linear regression, the results did not differ in significance (53). Therefore, we hypothesize that results similar to those presented here would be obtained using a censored modeling approach, and recommend future research to test this hypothesis and confirm the findings of the uncontrolled analysis.

Strengths

To the best of our knowledge, this study is one of the first to examine the relationships among fecal indicators in four sample types and in agricultural environments. A similar study on lettuce farms considered only bacterial pathogens and indicators, whereas our study examined both bacterial and viral indicators (24). Furthermore, the sampling methods and lab tests employed in this study were standardized across sample types, thus allowing for comparability across sample types.

The sample sizes for this study were sufficient to provide adequate power for linear and logistic regression including several control variables. Many previous studies have focused primarily on Spearman's correlation analyses in reporting their results. Controlling for variables such as produce type and point-in-chain allows for more interpretable and generalizable results, which is desirable as produce quality standards and tests are not likely to be developed separately for each specific type of produce, nor for each step in the farming process. Thus, understanding the overall relationships between the indicators in the presence of these control variables allows the results to guide the development of practical policy recommendations, with the confidence that these findings are not being heavily influenced by any of the factors discussed above.

Implications

This study adds to the growing evidence that multiple indicators are necessary to predict potential pathogen contamination and also provides important context for these tests in an agricultural setting. This evidence is relevant not only for policy-makers, but for farm owners as well. As the FDA revises and updates the regulations under the FSMA,

increased evidence on relatedness of indicator and pathogen behavior on produce farms is needed. Currently, only agricultural water has specific regular microbial testing requirements and thresholds under the FSMA. Although organic soil amendments (i.e. compost) must be treated using methods shown to reduce pathogen and indicator contamination, these standards do not include regular testing requirements. Given that contamination was found across sample types, implementing standards that require testing of both soil and produce, for example, should be considered as well.

Furthermore, standards that require testing for multiple indicators should be considered, given the evidence that indicator patterns vary in the environment and that certain indicators are better predictors of pathogens than others. Whether the goal is to test for the presence of fecal contamination or for the potential presence of a disease-causing pathogen, the presence of a single indicator may not be sufficient.

These results also highlight the need to further understand the interplay between disease-causing pathogens and microbial indicators on farms. A review of randomized and control trials focused on produce contamination found that several trials have been able to characterize some of the ways in which specific pathogens behave on various produce types and how this behavior differs based on various factors, including environmental factors such as water, soil, and produce type, as well as pathogen specific factors (54). Combining these trials with indicator testing as a means to understand how well microbial indicator presence detects specific pathogens will further ensure that the best indicators are being used to predict pathogen contamination.

Conclusions

Our results support findings from research on water safety and show that, on produce farms, indicator prevalence and concentrations vary, and that there are no two indicators which are related across all characteristics. Overall high prevalence of certain indicators provides evidence of fecal contamination on produce farms. Even without detailed understanding of pathogen-indicator interplay, these results underscore the need for increased use of good agricultural practices and other methods to minimize contamination of farms, and highlight the need for further research and development of industry-specific standards of testing for produce contamination. This study is the first to characterize indicator relationships across agricultural samples of hand-rinse, soil, water, and produce, considering several steps in the pre-harvest and packaging process. Hopefully this research can be the first step in fully understanding the complex relationships and processes at play in the produce growing environment. With the development of the FSMA's Produce Safety Rule, this evidence should to contribute to scientific standards on produce safety regulations which, in turn, have the potential to improve food safety and reduce pathogen contamination and foodborne outbreaks.

References

1. DeWaal CS, Glassman M. Outbreak Alert! 2001-2010. A Review of Foodborne Illness in America. : Center for Science in the Public Interest March 2013 March 2013

2. CDC Estimates of Foodborne Illness in the US In: Centers for Disease Control and Prevention 2014

3. Centers for Disease Control (CDC). List of Selected Multistate Foodborne Outbreak Investigations. In: Foodborne Outbreaks; 2014.

4. Painter J, Hoekstra R, Ayers T, Tauxe R, Braden C, Angulo F, et al. Attribution of Foodborne Illnesses, Hospitalizations, and Deaths to Food Commodities by using Outbreak Data, United States, 1998–2008 In: Emerging Infectious Diseases: Centers for Disease Control 2013.

5. Centers for Disease Control and Prevention. Multistate Outbreaks of Salmonella Infections Associated with Raw Tomatoes Eaten in Restaurants --- United States, 2005--2006. Morbidity and Mortality Weekly Report (MMWR) 2007;56(35):909-911.

6. Multistate Outbreak of Salmonella Saintpaul Infections Linked to Raw Produce (Final Update). In: Salmonella: 2009 through 2006 Outbreaks. online: Centers for Disease Control; 2008.

7. Heymann DL, editor. Control of Communicable Diseases Manual 19 ed. Washington, DC: American Public Health Association 2008.

8. Centers for Disease Control and Prevention. Outbreak of Salmonella Serotype Saintpaul Infections Associated with Multiple Raw Produce Items --- United States, 2008. Morbidity and Mortality Weekly Report (MMWR) 2008;57(34):929-934.

9. Centers for Disease Control and Prevention. Multistate Outbreak of Listeriosis Linked to Whole Cantaloupes from Jensen Farms, Colorado. In: Centers for Disease Control and Prevention 2012.

10. Food and Drug Administration. FDA Warns Consumers in New Mexico and Texas Not to Eat Certain Types of Raw Red Tomatoes. In: Administration FaD, editor.; 2008.

11. Meyerson A. An Analysis of the First-Order Economic Costs of the 2008 FDA Tomato Warning In: The Leonard N. Stern School of Business, Glucksman Institute for Research in Securities Markets; 2009. p. 1-15

12. Finding the Point of Contamination and Source of the Food. In: Foodborne Outbreaks: Centers for Disease Control and Prevention 2013.

13. Barclay E. Disease Detectives Are Solving Fewer Foodborne Illness Cases. In. NPR.com: National Public Radio (NPR) 2014.

14. Advanced Molecular Detection & Response to Infectious Disease Outbreaks FY 2015 President's Budget Request | $30 Million. In: HHS, editor.: Centers for Disease Control and Prevention; 2014. p. 1.

15. Ercsey-Ravasz M, Toroczkai Z, Lakner Z, Baranyi J. Complexity of the international agro-food trade network and its impact on food safety. PLoS One 2012;7(5):e37810.

16. Pathogens and Toxins in Foods: Challenges and Interventions Washington, DC: ASM Press; 2010.

17. FSMA Proposed Rule for Produce Safety In: Administration UDoHHSUFaD, editor. FDA Food Safety Modernization Act: FSMA Facts US Department of Health & Human Services/ US Food and Drug Administration p. 1-5

18. Commodity-Specific Food Safety Guidelines for Cantaloupes and Netted Melons 2013 March 29, 2013.

19. California: Commodity Specific Food Safety Guidelines for the Production, Harvest, Cooling, Packing, Storage, and Transporting of Cantaloupes and Other Netted Melons 2013 April 17, 2013

20. Leon J, Jaykus L, Moe C. Microbiology of fruits and vegetables. In: J. S-G, editor. Microbiologically safe foods. New Jersey: John Wiley & Sons; 2007

21. Seitz SR, Leon JS, Schwab KJ, Lyon GM, Dowd M, McDaniels M, et al. Norovirus infectivity in humans and persistence in water. Appl Environ Microbiol 2011;77(19):6884-8.

22. Sidhu JPS, Toze SG. Human pathogens and their indicators in biosolids: A literature review. Environment International 2009;35(1):187-201.

23. Economou V, Gousia P, Kansouzidou A, Sakkas H, Karanis P, Papadopoulou C. Prevalence, antimicrobial resistance and relation to indicator and pathogenic microorganisms of Salmonella enterica isolated from surface waters within an agricultural landscape. Int J Hyg Environ Health 2013;216(4):435-44.

24. Holvoet K, Sampers I, Seynnaeve M, Uyttendaele M. Relationships among hygiene indicators and enteric pathogens in irrigation water, soil and lettuce and the impact of climatic conditions on contamination in the lettuce primary production. Int J Food Microbiol 2014;171:21-31.

25. Leclerc H, Edberg S, Pierzo V, Delattre JM. Bacteriophages as indicators of enteric viruses and public health risk in groundwaters. Journal of Applied Microbiology 2000;88(1):5-21.

26. Savichtcheva O, Okabe S. Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Res 2006;40(13):2463-76.

27. Yates MV. Classical indicators in the 21st century--far and beyond the coliform. Water Environ Res 2007;79(3):279-86.

28. Ashbolt NJ, Grabow WOK, Snozzi M. Indicators of Microbial Water Quality. In: Ferwtrell L, Bartram J, editors. Water Quality: Guidelines, Standards, and Health. London: IWA Publishing 2001.

29. Doğan-Halkman H, Çakır İ, Keven F, Worobo R, Halkman AK. Relationship among fecal coliforms and Escherichia coli in various foods. European Food Research and Technology 2003;216(4):331-334.

30. Wilkes G, Edge T, Gannon V, Jokinen C, Lyautey E, Medeiros D, et al. Seasonal relationships among indicator bacteria, pathogenic bacteria, Cryptosporidium oocysts, Giardia cysts, and hydrological indices for surface waters within an agricultural landscape. Water Res 2009;43(8):2209-23.

31. Lucena F, Ribas F, Duran AE, Skraber S, Gantzer C, Campos C, et al. Occurrence of bacterial indicators and bacteriophages infecting enteric bacteria in groundwater in different geographical areas. J Appl Microbiol 2006;101(1):96-102.

32. Borrego JJ, Moriñigo MA, de Vicente A, Córnax R, Romero P. Coliphages as an indicator of faecal pollution in water. Its relationship with indicator and pathogenic microorganisms. Water Research 1987;21(12):1473-1480.

33. BioVir Laboratories. Bacteriophage. In: Johnson C, editor.: BioVir Laboratories,.

34. Noble RT, Moore DF, Leecaster MK, McGee CD, Weisberg SB. Comparison of total coliform, fecal coliform, and enterococcus bacterial indicator response for ocean recreational water quality testing. Water Research 2003;37(7):1637-1643.

35. Savichtcheva O, Okabe S. Alternative indicators of fecal pollution: Relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Research 2006;40(13):2463-2476.

36. Lemarchand K, Lebaron P. Occurrence of Salmonella spp and Cryptosporidium spp in a French coastal watershed: relationship with fecal indicators. FEMS Microbiol Lett 2003;218(1):203-9.

37. Wu J, Long SC, Das D, Dorner SM. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. Journal of Water & Health 2011;9(2):265-278.

38. Ailes EC, Leon JS, Jaykus LA, Johnston LM, Clayton HA, Blanding S, et al. Microbial concentrations on fresh produce are affected by postharvest processing, importation, and season. J Food Prot 2008;71(12):2389-97.

39. Gagliardi JV, Millner PD, Lester G, Ingram D. On-farm and postharvest processing sources of bacterial contamination to melon rinds. J Food Prot 2003;66(1):82-7.

40. Mukherjee A, Speh D, Dyck E, Diez-Gonzalez F. Preharvest evaluation of coliforms, Escherichia coli, Salmonella, and Escherichia coli O157:H7 in organic and conventional produce grown by Minnesota farmers. J Food Prot 2004;67(5):894-900.

41. 2012 Recreational Water Quality Criteria In: (EPA) USEPA, editor.; 2012

42. Shuren J. Beverages: Bottled Water. In: Food and Drug Administration H, editor. 102; 2009. p. 25651-25665.

43. Food and Drug Adminstration. FSMA Proposed Rule for Produce Safety. In: Services UDoHaH, Adminstration UFaD, editors.; 2013.

44. Field KG, Samadpour M. Fecal source tracking, the indicator paradigm, and managing water quality. Water Research 2007;41(16):3517-3538.

45. Jofre J. Is the replication of somatic coliphages in water environments significant? Journal of Applied Microbiology 2009;106(4):1059-1069.

46. Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, et al. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Appl Environ Microbiol 2005;71(6):3163-70.

47. Food and Drug Administration. FSMA Facts: Background on the Food Safety Modernization Act In: Food and Drug Administration 2011.

48. Johnston LM, Jaykus LA, Moll D, Martinez MC, Anciso J, Mora B, et al. A Field Study of the Microbiological Quality of Fresh Produce. Journal of Food Protection 2005;68(9):1840-1847.

49. Kleinbaum DG, Klein M. Logistic Regression: A Self-Learning Text. 3rd Edition ed: Springer; 2010.

50. Kleinbaum DG, Kupper LL, Nizam A, Muller KE. Applied Regression Analysis and Other Multivariable Methods. 4 ed. Belmont: Duxbury Press; 2008.

51. Nizam A. Linear Regression: Regression Diagnostics In; 2013. p. 40-45.

52. Alam S, Khalia S, Ayub N, Bibi A, Saeed B, Khalid S, et al. Prevelance of total coliforms, faecal coliforms and E. coli in Rawalpindi vegetable markets. Natural Science 2013;5(12):1298-1304.

53. Adam E. The Effect of Produce Type, Season, and Postharvest Handling on Microbial Quality of Fresh Produce Collected Near the U.S.-Mexico Border Emory University 2012.

54. Park S, Szonyi B, Gautam R, Nightingale K, Anciso J, Ivanek R. Risk Factors for Microbial Contamination in Fruits and Vegetables at the Preharvest Level: A Systematic Review. Journal of Food Protection 2012;75(11):2055-2081.

Tables

Table 1: Mean[¥] and Prevalence of Microbial Indicators on Produce and Environmental Samples

	Sample Type														
	Produce (cfu/fruit)*			Hand-Rinse (cfu/hand)*			Soil (cfu/g)*			Source Water (cfu/100ml water)*			Irrigation Water (cfu/100ml water)*		
Indicator	Geometric Mean (CI)[Σ]	Prevalence N (%)	N	Geometric Mean (CI)[Σ]	Prevalence N (%)	N	Geometric Mean (CI)[Σ]	Prevalence N (%)	N	Geometric Mean (CI)[Σ]	Prevalence N (%)	N	Geometric Mean (CI)[Σ]	Prevalence N (%)	N
E. coli	1.31 (1.06, 1.56)	69 (27%)	254	2.40 (2.12, 2.67)	63 (37%)	171	-0.004 (-0.26, 0.25)	17 (21%)	81	0.14 (-0.15,0.42)	27 (54%)	50	-0.09 (-0.29, 0.12)	31 (41%)	76
Enterococcus	5.06 (4.76, 5.36)	215 (85%)	254	6.53 (6.29, 6.76)	171 (100%)	171	1.36 (1.15, 1.58)	56 (69%)	81	0.52 (0.26, 0.77)	41 (87%)	47	0.52 (0.25, 0.80)	63 (84%)	75
Fecal coliforms	5.22 (4.96, 5.49)	243 (97%)	250	5.72 (5.43, 6.00)	164 (96%)	171	2.50 (2.25, 2.75)	80 (100%)	80	1.59 (1.27, 1.90)	41 (95%)	43	1.65 (1.35, 1.95)	68 (93%)	73
Coliphage	2.12 (1.85, 2.38)	159 (83%)	191	2.13 (1.84, 2.43)	85 (65%)	130	-0.42 (-0.65, -0.19)	22 (36%)	61	1.15 (0.41, 1.87)	11 (44%)	25	0.99 (0.54, 1.45)	21(45%)	47

¥ Log_{10} transformed data are used in calculation of geometric mean

*For coliphage samples units are MPN rather than cfu (e.g. MPN/fruit, MPN/g); units otherwise remain unchanged

Σ 95% confidence interval (lower bound, upper bound)

Table 2: Mean¥ and Prevalence of Microbial Indicators on Produce Samples by Produce Type

	Tomato (cfu/fruit)*			Jalapeño (cfu/fruit)*			Melon (cfu/fruit)*		
Indicator	Geometric Mean (CI)Σ	Prevalence N (%)	N	Geometric Mean (CI)Σ	Prevalence N (%)	N	Geometric Mean (CI)Σ	Prevalence N (%)	N
E. coli	0.19 (-0.06, 0.44)	16 (19%)	84	0.27 (-0.02, 0.56)	10 (16%)	64	2.83 (2.44, 3.21)	43 (41%)	106
Enterococcus	3.50 (3.17, 3.83)	65 (77%)	84	3.58 (3.10, 4.05)	45 (70%)	64	7.20 (6.90, 7.49)	105 (99%)	106
Fecal coliforms	4.59 (4.17, 5.00)	81 (98%)	83	3.88 (3.24, 4.52)	56 (92%)	61	6.49 (6.27, 6.71)	106 (100%)	106
Coliphage	0.91 (0.61, 1.22)	53 (80%)	66	1.22 (0.77, 1.67)	36 (78%)	46	3.64 (3.35, 3.93)	70 (89%)	79

¥ Log_{10} transformed data are used in calculation of geometric mean

*For coliphage samples units are MPN rather than cfu (e.g. MPN/fruit, MPN/g); units otherwise remain unchanged

Σ 95% confidence interval (lower bound, upper bound)

Table 3: Spearman's Correlations Between Microbial Indicator Concentrations by Sample Type*¥€

		E. coli	*Enterococcus*	Fecal coliforms
Environmental Samples				
Hand-rinse	*Enterococcus*	0.37 (<0.0001)‡		
	Fecal coliforms	0.24(0.0017)‡	0.56(<0.0001)‡	
	Coliphage	0.15 (0.0938)	0.17 (0.0601)	0.07 (0.4634)
Soil	*Enterococcus*	-0.07 (0.5263)		
	Fecal coliforms	0.08 (0.4581)	0.38 (0.0003)‡	
	Coliphage	-0.31 (0.0111)‡	0.15 (0.2385)	0.15 (0.2371)
Water	*Enterococcus*	0.60 (<0.0001)‡		
	Fecal coliforms	0.50 (<0.0001)‡	0.54(<0.0001)‡	
	Coliphage	0.34(0.0031)‡	0.26(0.0251)‡	0.17 (0.1642)
Produce Samples				
Produce (All)	*Enterococcus*	0.47 (<0.0001)‡		
	Fecal coliform	0.28 (<0.0001)‡	0.65(<0.0001)‡	
	Coliphage	0.39 (<0.0001)‡	0.62(<0.0001)‡	0.44(<0.0001)‡
Tomato	*Enterococcus*	-0.13 (0.2372)		
	Fecal coliform	-0.17 (0.1322)	0.47 (<0.0001)‡	
	Coliphage	-0.16 (0.1958)	0.17 (0.1739)	0.17 (0.1704)
Melon	*Enterococcus*	-0.02 (0.8762)		
	Fecal coliform	0.30 (0.0018)‡	0.25(0.0094)‡	
	Coliphage	0.07 (0.5157)	0.27(0.0163)‡	-0.17 (0.1373)
Jalapeño	*Enterococcus*	-0.09 (0.4616)		
	Fecal coliform	-0.07 (0.5748)	0.58 (<0.0001)‡	
	Coliphage	-0.12 (0.4208)	0.17 (0.2476)	0.17 (0.2703)

*Rho values and p-value (in parentheses) reported for all sample types
¥Log_{10} transformed variables used for all analyses
€ Spearman's correlation test used due to non-normally distributed data
‡Indicates p-value of less than 0.05, significant difference from Rho=0 (null)

Table 4: Relationships between Microbial Indicator Concentrations, controlling for produce type, point in chain, and other microbial indicator concentrations[¤]

		Predictor							
		E. coli		*Enterococcus*		Fecal coliform		Coliphage	
Outcome	Sample Type	Coefficient (p-value)	Std. Error	Coefficient (p-value)	Std. Error	Coefficient (p-value)	Std. Error	Coefficient (p-value)	Std. Error
E. coli	Hand-rinse			0.22 (0.1078)	0.13	0.19 (0.1202)	0.12	0.10 (0.3089)	0.10
	Soil[£]			0.15 (0.3164)	0.15	<-0.01 (.9982)	0.15	-0.17 (0.2514)	0.15
	Water[£]			0.31 (0.0014)[‡]	0.09	0.37 (0.0002)[‡]	0.10	0.07 (0.1336)	0.05
	Produce			-0.20 (0.0204)[‡]	0.09	0.23 (0.0096)[‡]	0.09	-0.04 (0.6732)	0.10
Enterococcus	Hand-rinse	0.10 (0.1078)	0.06			0.51 (<0.0001)[‡]	0.07	0.14 (0.048)[‡]	0.07
	Soil[£]	0.11 (0.3164)	0.11			0.44 (0.0003)[‡]	0.17	0.18 (0.1593)	0.13
	Water[£]	0.48 (0.0014)[‡]	0.14			0.43 (0.0007)[‡]	0.12	-.024 (0.6935)	0.06
	Produce	-0.15 (0.0204)[‡]	0.06			0.52 (<0.0001)[‡]	0.07	0.27 (0.0008)[‡]	0.08
Fecal coliform	Hand-rinse	0.11 (0.1202)	0.07	0.65 (<0.0001)[‡]	0.08			-0.17 (0.0242)[‡]	0.08
	Soil[£]	<-0.01 (0.9982)	0.11	0.45 (0.0003)[‡]	0.12			-0.03 (0.8149)	0.13
	Water[£]	0.51 (0.0002)[‡]	0.13	0.38 (0.0007)[‡]	0.11			<-0.01 (0.8907)	0.06
	Produce	0.16 (0.0096)[‡]	0.06	0.51 (<0.0001)[‡]	0.06			.005 (0.9557)	0.08
Coliphage	Hand-rinse	0.08 (0.3089)	0.08	0.24 (0.0481)[‡]	0.12	-0.24 (0.0242)[‡]	0.10		
	Soil[£]	-0.14 (0.2514)	0.12	0.19 (0.1593)	0.13	-0.03 (0.8149)	0.13		
	Water[£]	0.48 (0.1336)	0.32	-0.10 (0.6935)	0.26	-0.03 (0.8907)	0.27		
	Produce	-0.03 (0.6732)	0.06	0.23 (0.0008)[‡]	0.07	<0.01 (0.9557)	0.07		

¤ Each row represents one full linear model (non-indicator control variables not reported)

‡Significant result at α=0.05

£ Models all collected 'Before harvest'; point in chain not included in models

Table 5: Relationships between the Prevalence of Microbial Indicators by Sample Type¥

		E. coli	*Enterococcus*	Fecal coliform
Hand-rinse	*Enterococcus*	N/A€		
	Fecal coliform	(0.0391)‡**	N/A€	
	Coliphage*	3.24 (0.0027)‡	N/A€	1.95 (0.4173)
Soil	*Enterococcus*	1.31 (0.6473)		
	Fecal coliform	(0.3019) **	0.65 (0.7171)	
	Coliphage*	0.33 (0.1005)	0.33 (0.0400)‡	0.00 (0.0471)‡
Water	*Enterococcus*	8.8 (0.0012)‡		
	Fecal coliform	(0.0115)‡ **	2.25 (0.3453)	
	Coliphage*	4.08 (0.0041)‡	3.29 (0.1390)	N/A€
Produce (All)	*Enterococcus*	1.85 (0.1596)		
	Fecal coliform	(0.1009)**	8.79 (0.0011)‡	
	Coliphage*	4.36 (0.0048)‡	0.99 (0.9876)	9.00 (0.0053)‡
Tomato	*Enterococcus*	0.85 (0.8003)		
	Fecal coliform	(0.4842)**	3.76 (0.3253)	
	Coliphage*	4.74 (0.1202)	0.39 (0.2336)	(0.0025)‡**
Melon	*Enterococcus*	(0.4065)**		
	Fecal coliform	N/A€	N/A€	
	Coliphage*	4.4 (0.0584)	N/A€	N/A€
Jalapeño	*Enterococcus*	0.98 (0.9812)		
	Fecal coliform	(0.3316) **	4.50 (0.0944)	
	Coliphage*	2.17 (0.4858)	1.18 (0.8217)	2.36 (0.4967)

¥ Mantel-Hantzel OR and Pearson's chi-squared p-value (in parentheses) reported

‡Significant at α=0.05

*Smaller sample sizes for coliphage limit accuracy of results

€ Analysis not conducted due to 100% prevalence of at least one indicator

** M-H OR could not be calculated due to at least one zero cell

Table 6: Relationships between Prevalence of Microbial Indicators, Controlling for Produce Type, Point in Chain, and Other Indicators¤

Outcome	Sample Type	*E. coli*		*Enterococcus*		Fecal coliform		Coliphage	
		OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
E. coli	Hand-rinse			N/A¥	N/A¥	9.34 (0.35, 249.34)*	0.1823	3.92 (1.64, 9.37)	0.0021‡
	Soil£			1.23 (0.32, 4.76)	0.7690	1.42 (0.02, 83.94)*	0.8666	0.80 (0.16, 4.07)	0.7925
	Water£			17.05 (0.78, 373.31)*	0.0717	N/A€	N/A€	3.01 (0.94, 9.69)	0.0643
	Produce			1.06 (0.36, 3.18)	0.9124	1.77 (0.07, 47.19)*	0.7345	4.01 (1.06, 15.18)	0.0408‡
Enterococcus	Hand-rinse	N/A¥	N/A¥			N/A¥	N/A¥	N/A¥	N/A¥
	Soil£	1.22 (0.31, 4.78)	0.7775			0.08 (0.002, 3.96)*	0.2028	0.40 (0.11, 1.41)	0.1549
	Water£	15.78 (1.11, 225.01)*	0.0419‡			N/A€	N/A€	1.10 (0.17, 7.29)	0.9237
	Produce	1.09 (0.36, 3.26)	0.8792			20.43 (1.53, 272.23)	0.0224‡	0.26 (0.05, 1.25)	0.0918
Fecal coliform	Hand-rinse	7.09 (0.61, 82.54)*	0.1181	N/A¥	N/A¥			1.37 (0.24, 7.83)	0.7202
	Soil£	2.36 (0.12, 46.56)*	0.5727	0.11 (0.01, 1.85) *	0.1241			0.22 (0.02, 3.02)*	0.2555
	Water£	N/A€	N/A€	N/A€	N/A€			N/A€	N/A€
	Produce	2.25 (0.18, 28.16)*	0.5284	12.96 (1.56, 108.05)*	0.0179‡			25.51 (2.55, 255.37)*	0.0059‡
Coliphage	Hand-rinse	3.77 (1.59, 8.93)	0.0025‡	N/A¥	N/A¥	1.19 (0.22, 6.61)	0.8410		
	Soil£	1.00 (0.19, 5.20)	0.9966	0.40 (0.11, 1.44)	0.1597	0.12 (0.002, 6.14)*	0.2894		
	Water£	2.87 (0.91, 9.04)	0.0720	1.32 (0.22, 8.14)	0.7622	N/A€	N/A€		
	Produce	3.77 (1.00, 14.16)	0.0498‡	0.29 (0.06, 1.40)	0.1226	18.86 (1.85, 192.24)	0.0132‡		

¤ Each row represents one full logistic model (results for non-indicator control variables not reported)

‡Significant result at α=0.05

£ Samples all collected 'Before harvest'; point in chain not included in models

*Indicates that Firth corrections were used to yield OR and 95% CI estimates due to sparsity of data in stratified subsets

¥ Values could not be calculated due to 100% prevalence for sample types

€ Values could not be calculated; all observations the same (near 100% prevalence)

Appendices

Appendix A

Logistic Regression Interaction Tests: Interaction Term P-Value*

Outcome	Sample Type	*E. Coli*	Fecal Coliform	*Enterococcus*	Coliphage
E. Coli	Produce		0.9891	0.9749	0.9639
	Hand		0.9937	N/A€	0.6890
	Water		N/A£	0.2468	0.0296‡
	Soil		0.9979	0.9538	0.1204
Fecal Coliform	Produce	0.9981		0.9623	0.9959
	Hand	0.9963		N/A€	0.9971
	Water	0.9715		0.9952	N/A€
	Soil	0.9987		0.9683	0.9746
Enterococcus	Produce	0.9919	0.9601		0.9999
	Hand	N/A€	N/A€		N/A€
	Water	0.2467	N/A£		0.9974
	Soil	0.9763	0.9681		0.9669
Coliphage¥	Produce	0.9881	N/A£	N/A£	
	Hand	0.6891	N/A£	N/A£	
	Water	0.0296‡	N/A£	N/A£	
	Soil	0.1204	N/A£	0.9555	

*Results presented show Pilot Year and Year 1 together compared to Year 2. Analysis for the three levels separately were conducted and yielded similar results. Results not shown.

‡Significant at α=0.05

€ Could not be calculated due to 100% prevalence of at least one variable

£Interaction term is a linear combination of other variables and could not be calculated

¥ Used Coliphage Year in interaction term. All other samples used year of bacterial sample collection.

Appendix B

Linear Regression Interaction Tests: P-value of interaction term as explanation of model variance*

Outcome	Sample Type	*E. Coli*	Fecal Coliform	*Enterococcus*	Coliphage
E. Coli	Produce		0.0589	0.6235	0.6097
	Hand		0.5320	0.5366	0.2255
	Water		0.0001‡	0.0012‡	0.8196
	Soil		0.4238	0.8165	0.6557
Fecal Coliform	Produce	0.3176		0.1385	<.0001‡
	Hand	0.6286		0.3203	0.8608
	Water	0.1067		0.9417	0.1065
	Soil	0.0021‡		0.2220	0.4681
Enterococcus	Produce	0.4619	0.1401		0.0045‡
	Hand	0.1302	0.2861		0.3049
	Water	0.0001‡	<.0001‡		0.3692
	Soil	0.4096	0.5438		0.8181
Coliphage¥	Produce	0.8842	<.0001‡	0.0009‡	
	Hand	0.3521	0.2625	0.6405	
	Water	0.4305	0.2500	0.4624	
	Soil	0.8071	0.7358	0.1694	

*Interaction terms for linear regression considered Pilot, Year 1, and Year 2 separately.
‡Significant at α=0.05
¥ Used Coliphage Year in interaction term. All other samples used year of bacterial sample collection.

Appendix C

IRB Approval was received to work on this project. An excerpt of the approval document with relevant reference numbers is included below.

Institutional Review Board

TO: Juan Leon, PhD
Principal Investigator
Global Health

DATE: October 9, 2013

RE: **Notification of Amendment Approval**
AM5_IRB00035460
IRB00035460
Identification and Control of Microbiological Hazards in Imported Fresh Fruits and Vegetables: A Field Epidemiological and Intervention Study in Northern Mexico

Thank you for submitting an amendment request. The Emory IRB reviewed and approved this amendment under the expedited review process on **10/9/2013**. This amendment includes the following:

Personnel Change only: Adding Yiru Gu, Jacqueline Lickness, Allison White, and Alexandra Stern as other Emory study staff. Remove Alexander Emmit from the study.