AEM Journals.ASM.org

Multifactorial Effects of Ambient Temperature, Precipitation, Farm Management, and Environmental Factors Determine the Level of Generic *Escherichia coli* Contamination on Preharvested Spinach

Sangshin Park,[a,b] Sarah Navratil,[c,d] Ashley Gregory,[e] Arin Bauer,[e] Indumathi Srinath,[a,f] Barbara Szonyi,[a] Kendra Nightingale,[c,d] Juan Anciso,[e] Mikyoung Jun,[g] Daikwon Han,[h] Sara Lawhon,[i] Renata Ivanek[a]

Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA[a]; Center for International Health Research, Rhode Island Hospital, The Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA[b]; Department of Animal Sciences, Colorado State University, Fort Collins, Colorado, USA[c]; Department of Animal and Food Sciences, Texas Tech University, Lubbock, Texas, USA[d]; Department of Horticultural Sciences, Texas A&M AgriLife Extension Service, Weslaco, Texas, USA[e]; Tarleton State University, Stephenville, Texas, USA[f]; Department of Statistics, Texas A&M University, College Station, Texas, USA[g]; Department of Epidemiology and Biostatistics, School of Public Health, Texas A&M Health Science Center, College Station, Texas, USA[h]; Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA[i]

A repeated cross-sectional study was conducted to identify farm management, environment, weather, and landscape factors that predict the count of generic *Escherichia coli* on spinach at the preharvest level. *E. coli* was enumerated for 955 spinach samples collected on 12 farms in Texas and Colorado between 2010 and 2012. Farm management and environmental characteristics were surveyed using a questionnaire. Weather and landscape data were obtained from National Resources Information databases. A two-part mixed-effect negative binomial hurdle model, consisting of a logistic and zero-truncated negative binomial part with farm and date as random effects, was used to identify factors affecting *E. coli* counts on spinach. Results indicated that the odds of a contamination event (non-zero versus zero counts) vary by state (odds ratio [OR] = 108.1). Odds of contamination decreased with implementation of hygiene practices (OR = 0.06) and increased with an increasing average precipitation amount (mm) in the past 29 days (OR = 3.5) and the application of manure (OR = 52.2). On contaminated spinach, *E. coli* counts increased with the average precipitation amount over the past 29 days. The relationship between *E. coli* count and the average maximum daily temperature over the 9 days prior to sampling followed a quadratic function with the highest bacterial count at around 24°C. These findings indicate that the odds of a contamination event in spinach are determined by farm management, environment, and weather factors. However, once the contamination event has occurred, the count of *E. coli* on spinach is determined by weather only.

Foodborne disease outbreaks associated with produce impose a considerable public health burden (1). Among different produce commodities, leafy green vegetables have been identified as a group of high concern from a microbiological safety perspective due to their being implicated in multiple outbreaks of foodborne disease with high numbers of illnesses worldwide (2). Indeed, leafy green vegetables are commonly grown in open farm fields, where they may be exposed to microbial contamination from soil, manure fertilizer, irrigation water, and intrusions of wild or domestic animals, and they are likely to be consumed fresh or minimally processed. Enteric foodborne pathogens, such as *Salmonella* and *Escherichia coli* O157:H7, have been the main causative agents responsible for foodborne outbreaks associated with leafy green vegetables in the United States (3). These pathogens are spread in the environment through feces of infected animals and humans (4). While contamination of leafy greens with these foodborne pathogens has major consequences (2), fortunately it occurs at a low frequency (5, 6). The low frequency and heterogeneous distribution of these pathogens in the produce field make their detection difficult, costly, and time-consuming. Instead indicator organisms are routinely used by the industry, environmental agencies, and public health organizations to verify effective implementation of good agricultural practices (GAPs) and good manufacturing practices (GMPs) (7, 8). Generic *E. coli* has been used as an indicator of fecal contamination on produce and to study risk factors for such contamination. Previous studies have documented the usefulness of generic *E. coli* in predicting *Salmonella* and *E. coli* O157:H7 persistence after manure and slurry application in research settings (6, 9, 10). Recently, the European Food Safety Authority (EFSA) took a new direction to define a criterion at primary production of leafy greens that would be designated the "hygiene criterion" (8). Generic *E. coli* was identified as suitable for a hygiene criterion that could be considered for validation and verification of GAPs and good hygiene practices (GHPs) and on the basis of which growers could take appropriate corrective actions. The rationale was that because *E. coli* is not often detected on leafy greens, is present in high numbers in fecal material (e.g., fresh manure), and declines in the soil or on leafy greens during

Received 18 November 2014 **Accepted** 27 January 2015

Accepted manuscript posted online 30 January 2015

Citation Park S, Navratil S, Gregory A, Bauer A, Srinath I, Szonyi B, Nightingale K, Anciso J, Jun M, Han D, Lawhon S, Ivanek R. 2015. Multifactorial effects of ambient temperature, precipitation, farm management, and environmental factors determine the level of generic *Escherichia coli* contamination on preharvested spinach. Appl Environ Microbiol 81:2635–2650. doi:10.1128/AEM.03793-14.

Editor: M. W. Griffiths

Address correspondence to Sangshin Park, dvm.spark@gmail.com.

doi:10.1128/AEM.03793-14

primary production, it can be considered an indicator of a recent exposure to risk factors for fecal contamination in which *Salmonella* could be present (8). Thus, it is of interest to identify risk factors for leafy green contamination with generic *E. coli* in order to reduce the produce contamination with fecal material and the associated foodborne pathogens and thereby prevent produce-related foodborne illnesses.

Several studies (5, 6, 11–13) have been conducted to identify the farm-related management practices, environmental, landscape, and weather factors affecting generic *E. coli* contamination of produce at the preharvest level. According to these studies, produce contamination with *E. coli* may occur from various sources (e.g., soil amendments such as raw or improperly composted manure [5, 6, 11, 13], contaminated irrigation water [6], and fecal deposition through domestic animal and wildlife [6, 13]) and may be aggravated due to weather conditions that aid transport of *E. coli* to produce (e.g., rain events) or increase the survivability and growth rate of *E. coli* (e.g., high humidity). They have also shown that the probability of contamination decreased with implementation of GAPs, such as providing toilets and handwashing facilities to field workers (6, 13) and avoiding application of animal manure (5, 6, 11, 13). Our systematic review study (14) indicated that the majority of these risk factors for contamination with generic *E. coli* have been confirmed to play a role in produce contamination with foodborne pathogens, such as *Salmonella* and *E. coli* O157:H7.

Although numerous studies have evaluated farm-related factors influencing the microbial contamination of produce (5, 6, 11–13, 15–22), only a limited number of studies (16, 17, 23) examined the bacterial count on produce as affected by risk factors other than the farming type (e.g., organic or conventional farm) (5, 12, 18–20). Surprisingly, to our knowledge, no study has systematically assessed the impact of farm-related factors, including management practices, environment, landscape, and weather factors, on the bacterial count on produce. Yet, bacterial count data can provide valuable information for (i) evaluation of the efficacy of decontamination practices, (ii) cross-contamination potential throughout the food production and distribution chain, and (iii) dose of human foodborne pathogen exposure. As a result, GAPs should be designed not only to reduce the probability of produce contamination but also to reduce the extent of contamination on produce in terms of the bacterial count.

In practice the bacterial count ranges from zero, in a product that is noncontaminated or contaminated under the lower detection limit, to some positive integer that reflects the upper limit of detection for the employed enumeration method. Because count data usually do not follow a normal distribution due to a low frequency of high counts and excess of zero counts, a $\log_{10}$ transformation is often used to approximate data normality that is often necessary for parametric data analysis (24). The long-known log-normal distribution of epiphytic bacterial populations on leaf surfaces (25) provides a biological support for this practice. Poisson models are often considered the method of choice for analysis of count data, but they are rarely adequate because they assume the equality of mean and variance (26). Recently, alternative statistical methods have been proposed for the microbial count data in the area of food safety (23, 24). That is because in practice, most $\log_{10}$-transformed bacterial count data show relatively high heterogeneity in the form of an excess number of zeros ("zero-inflation"), excess in variance above that expected by Poisson distribution ("overdispersion"), or both. In analysis of bacterial count data, researchers may therefore need to consider more flexible models that are able to handle zero-inflation (zero-inflated Poisson [ZIP] and Poisson hurdle [PH] models), overdispersion (negative binomial [NB] model), or both (zero-inflated negative binomial [ZINB] and negative binomial hurdle [NBH] models) (27), depending on the characteristics of data. The above-mentioned ZIP, PH, ZINB, and NBH models are all two-part models, albeit with important differences. The hurdle (PH and NBH) models address zero and non-zero counts separately, which would in the context of bacterial counts on produce mean that all observed zero counts are interpreted as noncontaminated samples, while non-zero counts denote contaminated samples (discussed further in Discussion). The zero-inflated (ZIP and ZINB) models also model non-zero counts as contaminated samples, but they partition zeros into two types, structural and ineligibility zeros, with zeros obtained for samples contaminated under the detection limit as structural zeros, while the true zeros in case of complete absence of contamination are modeled as ineligibility zeros.

The objective of this study was to identify farm management, environment, landscape, and weather factors that determine the count of generic *E. coli* on spinach at the preharvest level. To address this objective, we conducted a repeated cross-sectional study on 12 spinach farms in Texas and Colorado, which involved enumeration of generic *E. coli* on spinach and consideration of the following four groups of risk factors: (i) management and (ii) environmental characteristics of farms obtained through a survey of farmers and local (iii) landscape and (iv) weather characteristics obtained through spatial modeling from publicly available National Resources Information (NRI) databases. Our previously published studies (6, 13) described data on generic *E. coli* presence/absence on spinach in relation to the above-considered groups of risk factors: specifically groups i and ii were considered in reference 6, while all of groups i to iv were considered in reference 13. The present study introduces never before published data on the count of generic *E. coli* on contaminated spinach and combines the new data with previously described presence/absence (i.e., non-zero and zero counts) data to evaluate in one step the role of the above four groups of risk factors in determining the count of *E. coli* on spinach.

MATERIALS AND METHODS

Sample collection. Sample collection was described in detail in our previous study (6). Briefly, over a 2-year period from 7 June 2010 to 10 February 2012, a total of 955 spinach samples were collected on 12 spinach farms (8 in Texas and 4 in Colorado). The States of Texas and Colorado are representative of the southwestern and western United States, respectively, and are important vegetable production areas of the United States (28) albeit with different spinach growing seasons. Spinach is grown in Texas between November and March and in Colorado between April and September. Each farm was visited on 1 to 5 sampling dates per growing season for a total of 2 to 8 sampling dates over the study period. At each visit, 5 spinach samples (4 from the field's corners and an additional sample from the field center) were collected from each of 1 to 6 fields per farm. For each sample, we randomly collected more than 10 individual leaves of multiple spinach plants in an area with a 5-meter radius. After sample collection, individual samples were placed into individual Whirl-Pak bags (Nasco, Fort Atkinson, WI). The location of each sampling site was recorded using a hand-held global positioning system (GPS) device (Garmin 12XL; Garmin, Ltd., Olathe, KS). All samples

were stored in a cooler with ice packs, transported to a laboratory, and processed within 48 h.

Microbiological analyses. In the laboratory, 25 g of a spinach sample was transferred into 75 ml phosphate-buffered saline (PBS) in a stomacher bag and then mixed with PBS for 2 min using a blender (Smasher Lab Blender; AES-Chemunex, France). An aliquot of 1 ml of sample dilution and each four successive 1:10 serial dilutions was pipetted onto Petrifilm *E. coli*/coliform count plates (3M Microbiology, St. Paul, MN). Each count plate was then incubated at 37°C ± 2°C for 48 h. Blue colonies with gas bubbles were counted to enumerate generic *E. coli* colonies according to the standard Petrifilm enumeration method (http://tmacog.org/Environment/SWW_07/PetrifilmInterpretation.pdf), and the count of generic *E. coli* for an individual sample was expressed as CFU/g. The detection limit was 4 CFU/g of spinach. If no colonies were detected, the count was recorded as zero. If the colony number was "too numerous to count" (TNTC), it was recorded as 6×10^7 CFU/g, which was greater by a first order of magnitude than the upper enumeration limit of 6×10^6 CFU/g (calculated based on the upper countable range for Petrifilm of 150 CFU and 1:10,000 serial dilution of the 1:4 diluted sample as $150 \times 10{,}000 \times 4 = 6 \times 10^6$).

Questionnaire data. Information on farm management and the environment was obtained using a detailed questionnaire administered at each farm visit and described in detail in reference 6. Briefly, the considered farm management information included 50 variables under the categories of human factors (7 variables), farm and field conditions (13 variables), pesticide (8 variables), chemical fertilizer (2 variables), manure fertilizer (4 variables), compost fertilizer (2 variables), irrigation (10 variables), equipment (2 variables), routine microbial test (1 variable), and time since planting of spinach (1 variable). Farm environment information included 16 variables under the categories of terrain, buffer zone, and proximity (11 variables), domestic/wild animals (4 variables), and farm location (1 variable). Of those, five management and environmental factors (the use of portable toilets, the presence of training to use portable toilets, the use of hand-washing stations, and the absence of grazing and hay production in the field before planting of the spinach during the current growing season) always co-occurred and were evaluated by using a composite variable, "hygiene-field status." Thus, 62 variables describing farm management and environment were considered in the statistical analyses here. Similar to our approach in reference 6, all continuous farm management and environmental factors were median dichotomized for regression models, because the linearity assumption was not met between explanatory and outcome variables. For brevity, in Table 1, we list only variables (previously defined in references 6 and 13) that were in univariable analyses significant at a P value of 0.2.

NRI databases. Collection and manipulation of weather and local landscape variables were performed following the general procedure suggested by Ivanek et al. (29) and described in detail in our previous study (13). Briefly, weather and landscape information of interest were obtained from the National Resource Information (NRI) databases based on the GPS coordinates recorded for the individual spinach sampling locations. From the National Oceanic and Atmospheric Administration-National Climatic Data Center (http://www.ncdc.noaa.gov/), we obtained weather information for 90 variables under the categories of ambient temperature (54 variables), precipitation (18 variables), and wind speed (18 variables). We used weather data recorded at one of the nearby 22 land-based weather stations that was closest to the individual sampling location and had records for the dates of interest. The average distance between the individual farms and weather stations was 11.9 km (range, 1.5 to 34.7 km). We considered average, minimum, and maximum daily ambient temperature, because it is unclear which of these characteristics has the strongest effect on *E. coli* count on spinach. The amount of precipitation recorded as "trace" was assigned 0.0001 mm. A total of 18 variables were created for each weather factor (average, minimum, and maximum temperatures, precipitation, and wind speed) and considered in the study. Of the 18 variables, 4 were created to describe the weather characteristics on the day of sample collection and on days 1, 2, and 3 prior to sample collection. Additionally, we created 14 period variables describing the mean levels of weather characteristics between the day of sample collection (day 0) and days 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, and 29 prior to sample collection. The reason for these point and period measures was to assess if weather influences *E. coli* count on spinach instantly or cumulatively (29). For brevity, in Table 1, we define only variables significant at a P value of 0.2 in univariable analyses.

We also considered local landscape characteristics in terms of soil properties and distance factors, which are described in detail in reference 13. Briefly, data on the soil property factors (4 variables: soil acidity, soil texture, slope, and organic matter) were obtained from the Soil Survey Geographic (SSURGO) database (http://websoilsurvey.nrcs.usda.gov/) for each sampling location. The distance factors (2 variables) of interest were the distances to the nearest water body and road from the individual sampling location. Data on the locations of water bodies and roads in Texas were, respectively, obtained from the National Hydrography and TxDOT Roadways data sets of the Texas Natural Resources Information System (http://www.tnris.org/). In Colorado, data on water bodies and roads were, respectively, obtained from the Hydrography-1M and Transportation-1M data sets of the Colorado Department of Natural Resources (http://data.geocomm.com/catalog/US/61076/datalist.html). The data on soil properties, hydrology, and roads were imported into ArcGIS 10 (ESRI, Redland, CA), reprojected into the Universal Transverse Mercator, North American Datum of 1983, and overlaid with GPS coordinates of spinach sampling locations. Then, for each sampling location we extracted soil property information and estimated the distances to the nearest water body and road. Among the above landscape factors, a single variable (soil acidity) was significant at a P value of 0.2 in univariable analyses, and so it is the only landscape variable defined in Table 1.

Statistical analyses. All statistical analyses were performed using R version 2.15.1 (R Foundation for Statistical Computing; http://www.r-project.org/). The positive counts of generic *E. coli* on spinach samples (including for two samples with *E. coli* TNTC and recorded as 6×10^7 CFU/g) were $\log_{10}$ transformed (because their distribution was highly skewed) and rounded off to the nearest integer for regression modeling. A single exception to this approach was a raw count of 2 CFU/g (obtained by taking an average of two replicate plates with 4 CFU and 0 CFU). For analysis, this count (2 CFU/g) was recorded as 4 CFU/g before $\log_{10}$ transformation. Thus, the $\log_{10}$-transformed generic *E. coli* counts on contaminated spinach ranged from 1 to 8 $\log_{10}$ CFU/g, while noncontaminated spinach was assigned a value of 0 $\log_{10}$ CFU/g. The excess of zeroes (892/955 observations) in the data caused overdispersion: the arithmetic mean and variance of generic *E. coli* counts among 955 observations were 0.213 and 0.899 $\log_{10}$ CFU/g, respectively. The classical Poisson regression model, which assumes that variance is equal to the mean, or its extensions (such as quasi-Poisson and NB regression models) are not useful for such data, because those models may not yield a good fit for the distribution of counts showing overdispersion due to zero inflation and extra variation in the positive count data (27, 30). Indeed, the estimated dispersion parameter using the "dispersiontest" function in the AER package was 3.0 (alternative hypothesis: true dispersion is greater than 1; $P = 0.002$), which indicated overdispersion (26). Thus, for analysis of these data we considered the following two-part models: ZINB, ZIP, PH, and NBH, and we compared their fit using the Akaike information criterion (AIC), which was computed according to reference 30 as $\text{AIC} = [\text{AIC}_{\text{P\_count}} \times (1 - n_{>0}/n)] + \text{AIC}_{\text{binary}}$, where $\text{AIC}_{\text{P\_count}}$ and $\text{AIC}_{\text{binary}}$, respectively, correspond to the values of AIC for the P_count and binary model parts defined in the next paragraph and n and $n_{>0}$, respectively, correspond to the total number of samples and the subset with positive counts only. For the variables included in the final mixed-effect model, AIC values were as follows: ZINB, 641.6; ZIP, 640.9; PH, 543.7; and NBH, 521. Based on these AIC values and the ability to handle both zero-inflation and overdispersion (27, 30), a mixed-effect NBH model with farm (12 farms) and date (37 dates) as random effects was considered the most appropriate model

TABLE 1 Definitions of explanatory variables that were significant at the 20% level in the univariable mixed-effect negative binomial hurdle models (with the random effects farm and date) in Table 4[a]

Factor and variable	Description and level(s)[b]	Unit
Farm management factors		
Workers_time	Time since last workers' visit during CGS	Days
Hygiene-field status	Composite variable coded with 1 indicating use of portable toilets and washing stations in the field, training to use portable toilets provided to staff/temporary workers, and absence of grazing and hay production in the field before spinach planting and 0 indicating otherwise (1/0)	
Organic	Organic farming practices currently applied on farm (yes/no)	
Organic_certified	Organic farming certified by National Organic Program (yes/no)	
Before_fallow	Field condition before planting of spinach during CGS: fallow (yes/no)	
Tillage_time	Time since last tilling, rotavating, or aerating of soil for CGS (con)	Days
Pesticide_time	Time since last pesticide application during CGS (con)	Days
Manure_application	Manure spread on field for CGS (yes/no)	
Manure_age	Age of manure spread onto field for CGS (con)	Wk
Irrigation_time	Time since last irrigation during CGS (con)	Days
Planting_time	Time since planting spinach (con)	Days
Environmental factors		
Proximity_dairy	Proximity within 10-mile radius of dairy farm (yes/no)	
Proximity_beef	Proximity within 10-mile radius of beef farm (yes/no)	
Proximity_poultry	Proximity within 10-mile radius of poultry farm (yes/no)	
Proximity_residential	Proximity within 10-mile radius of residential (yes/no)	
Domestic_animal	Domestic animal intrusion of field for CGS (yes/no)	
Wildlife_control_fences	Wildlife control methods of farm: fences (yes/no)	
Wildlife_control_hunting	Wildlife control methods of farm: hunting (yes/no)	
State	Farm location (Texas/Colorado as representative states of southern U.S./southwestern U.S.)	
Weather factors		
tm2	Avg daily temp on day 2 prior to SC	°C
tmdX[c]	Mean of avg daily temperatures in period between day of SC and day *X* prior to SC (*X* is 6, 7, 8, 9, 10, 15, 20, 25, or 29)	°C
ti2	Minimum daily temp on day 2 prior to SC	°C
tidX	Mean of the minimum daily temp between day of SC and day *X* prior to SC (*X* is 10 or 15)	°C
tx3	Maximum daily temp on day 3 prior to SC	°C
txdX	Mean of maximum daily temp between day of SC and day *X* prior to SC (*X* is 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 29)	°C
ws2	Wind speed on day 2 prior to SC	°C
pdX	Mean amt of rain between day of SC and day *X* prior to SC (*X* is 1, 5, 6, 7, 8, 9, 10, 25, or 29)	mm
Landscape factor		
Soil_acidity	Relative acidity/alkalinity of soil at a sampling location (6.1–7.9/7.9–9.0)	pH

[a] These variables are a subset of the considered farm management and environmental factors defined in Table 2 in reference 6 and weather and landscape factors defined in Table 2 in reference 13.
[b] CGS, current growing season; con, continuous variable; SC, sample collection.
[c] For example, tmd7 denotes the mean of the average daily temperatures recorded for the period between the day of sample collection and day 7 prior to the day of sample collection.

to analyze associations of individual explanatory (independent) variables with the count of generic *E. coli* on spinach (dependent variable) in our data.

The NBH model assumes that a continuous outcome is generated by two different processes; one determines occurrence of non-zero (i.e., positive) versus zero counts, and the other process governs the distribution of positive counts (30). In the present study, this assumption was supported by the model fit (AIC) results, and thus we considered the detection limit (4 CFU/g, which after $\log_{10}$ transformation and rounding to the nearest integer was converted to 1 $\log_{10}$ CFU/g) as a hurdle in analyzing bacterial counts. If generic *E. coli* was found on a spinach sample (i.e., the count was $\geq$1 $\log_{10}$ CFU/g), the hurdle was crossed. Whether the count is zero or positive is generally described using a binary model ("binary" model part). On the other hand, when the hurdle is crossed, the positive count is described using a zero-truncated count model ("P_count" model part). We used logistic regression (logit link: $\ln[p'/(1 - p')]$, where p' is the expected probability of contamination) to fit the odds of spinach being contaminated (i.e., the count of $\geq$1 $\log_{10}$ CFU/g) versus not contaminated (i.e., the count is 0 $\log_{10}$ CFU/g) for the binary part of the NBH model, whereas a zero-truncated negative binomial regression [log link: $\ln(\lambda')$, where λ' is the expected bacterial count on contaminated produce] was used to fit the counts of generic *E. coli* on contaminated spinach (i.e., counts of $\geq$1 $\log_{10}$ CFU/g) for the P_count part of the NBH model. The mixed-effect NBH model was implemented through glmmADMB function in the glmmADMB package (31) in R software to identify risk factors. The multivariable binary and P_count model parts are, respectively, given by the equations $\text{logit}(p_i') = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_{b\text{-farm}(i)} + e_{b\text{-date}(i)}$, $i = 1, \ldots, n$ and $\log(\lambda_i') = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \ldots + \gamma_k X_{ki} + e_{c\text{-farm}(i)} + e_{c\text{-date}(i)} + e_i$, $i = 1, \ldots, n$, where for the binary (P_count) part, β_0 (γ_0) is the intercept, $\beta_1, \ldots,$ and β_k ($\gamma_1, \ldots,$ and γ_k) are the coefficients for the considered explanatory variables $X_1, \ldots,$ and X_k, $e_{b\text{-farm}(i)}$ and $e_{b\text{-date}(i)}$ ($e_{c\text{-farm}(i)}$ and $e_{c\text{-date}(i)}$) are the random effects of farm and

date for spinach sample *i*, respectively, drawn from normal distributions with variances $\sigma^2_{b\text{-farm}}$ and $\sigma^2_{b\text{-date}}$ ($\sigma^2_{c\text{-farm}}$ and $\sigma^2_{c\text{-date}}$), respectively; and e_i is an error term. In the binary part, $\exp(\beta_k)$ (that is, the odds ratio [OR]) indicated the increase in the odds of spinach contamination with generic *E. coli* for 1 unit increase in an explanatory variable, *k*, after adjustment for other predictors. Similarly, in the P_count part, $\exp(\gamma_k)$ (that is, relative ratio [RR]) indicated the percentage change in the $\log_{10}$ count of generic *E. coli* on a contaminated spinach sample for 1 unit increase in an explanatory variable, *k*, after adjustment for other predictors.

To avoid colinearity, correlation analyses were conducted among explanatory variables based on the phi coefficient for two categorical variables and Spearman's correlation coefficient for two continuous variables or one categorical variable and one continuous variable (26, 32). If high colinearity or correlation was determined among factors (defined as $r > 0.6$), we selected the most biologically relevant factor for multivariable modeling.

The final NBH model was developed in two steps: univariable and multivariable analyses, which were conducted separately for the binary and P_count parts of the model. Variables with a *P* value of <0.2 at the univariable level were considered in multivariable modeling. Among the significant variables (*P* values of <0.2), listed in Tables 1 and 4, high correlation was determined among weather variables; therefore, we identified representative variables for multivariable regression analyses using a principal component (PC) analysis after standardizing them to unit variance. In PC analysis for the binary and P_count model parts, we used all 955 observations and 63 positive counts only, respectively. Three criteria were applied to determine how many components should be retained: (i) visual examination of a scree plot of eigenvalues, (ii) the cumulative proportion of variance accounted for by the PC analysis where a component was retained if it explained at least 80% of the total variance, and (iii) the interpretability criteria (defined as at least three variables with major loadings on each retained component, the same conceptual meaning among the variables loading on the same component, and simple structure of the rotated pattern with relatively high factor loading of a variable on only one component and relatively small loading on other components) (33). For multivariable analyses, we chose a representative weather variable for each principal component, which was the most significant in the univariable analyses and had good interpretability. We additionally used the PC scores as explanatory variables in the univariable and multivariable analyses instead of the individual weather variables.

In multivariable analyses, a manual backward stepwise selection was used for the model selection approach (a *P* value of <0.05 based on the Wald Z test). In this selection, an explanatory variable was retained if its retention significantly reduced the AIC and the residual deviance based on the likelihood ratio (LR) test (a *P* value of <0.05). Confounding was determined to exist when addition of a potential confounder changed the ln(OR) or ln(RR) of a risk factor by at least 20% (26) in the binary or P_count part of the NBH model, respectively. Two-way interaction terms were also considered in multivariable modeling. In the P_count model part, a quadratic term for temperature (i.e., $\text{txd}9^2$) was also considered because there appeared to be a curvilinear relationship between temperature and the count of generic *E. coli* on spinach. In the final model, colinearity among explanatory variables was examined by using a variance inflation factor (VIF). A locally weighted scatter plot smoothing curve was used to assess the linearity assumptions of continuous explanatory variables with the transformed outcome in terms of the link function: logit link for the binary part and log link for the P_count part (26).

We evaluated the predictive performance of each of the two parts of the final mixed-effect NBH model using a 3-fold cross-validation. The binary part was assessed by a receiver operating characteristic (ROC) curve (6). The predictive performance of the P_count part was assessed using the normalized root mean squared error (NRMSE) expressed as a percentage. NRMSE was calculated as the root mean squared error divided by the difference between the maximum and minimum observed generic *E. coli* counts. An NRMSE value close to zero indicates a good predictive performance. In cross-validation, the whole data set was randomly divided into three subsets of approximately equal sizes. Next, two subsets were used to estimate the model's coefficients by running both the binary and P_count parts of the NBH model, and the third subset was used to validate the corresponding predictive ability with the ROC curve and NRMSE, respectively. This process was repeated three times, each time with a different cross-validation subset. We also assessed an internal validity of the binary and P_count parts of the final NBH model by evaluating the ROC curve and NRMSE produced through comparison of predicted and observed values for the whole data set and the subset with positive counts, respectively.

RESULTS

Among 955 tested spinach samples, 63 (6.6%) had positive counts of generic *E. coli* (Table 2). After excluding samples with zero count, the overall geometric mean count (± geometric standard deviation) of generic *E. coli* was 1,383.7 ± 103.4 CFU/g. The highest geometric mean count of generic *E. coli* was found on a farm in Colorado (farm 2: 61,852.2 CFU/g). Only one of the enrolled 12 farms (farm 1 in Colorado) had all samples with zero counts. Most of sampled spinach was grown on the loam (34%) and silty clay loam (33%) soil: specifically, these were loam (70%) and clay loam (29%) soil in Colorado and silty clay loam (63%) and clay loam (14%) soil in Texas. The enrolled farms used either overhead (farms 1, 2, and 4 in Colorado and farms 1 and 2 in Texas) or flood (farm 3 in Colorado and farms 3 to 8 in Texas) irrigation.

Table 3 shows summary statistics for weather factors significantly associated with the count of generic *E. coli*. For the univariable mixed-effect NBH analyses, Table 4 shows the variables associated with the odds of spinach contamination with generic *E. coli* (binary part) and the count of generic *E. coli* on contaminated spinach (P_count part) at the 20% significance level. Among the farm management and environmental factors, spinach was less likely to be contaminated with generic *E. coli* in the presence of the "hygiene-field status" (defined in Table 1) group of factors on the sampled field. However, the odds of spinach contamination with generic *E. coli* were significantly increased by the fallow condition of the field before planting of spinach, the use of manure, and domestic animal intrusion. The count of generic *E. coli* on contaminated spinach was significantly increased by application of manure fertilizer and proximity (within 10 miles) of dairy, beef, and poultry farms. The count of generic *E. coli* on contaminated spinach was significantly reduced if the time since the last irrigation was >5 days, if the farm applied fences to control wildlife intrusion, or if a farm was located in Texas. Among weather factors, the odds of spinach contamination with generic *E. coli* were significantly reduced when spinach was exposed to higher maximum temperature for 25 or 29 days before sample collection but significantly increased when spinach was exposed to a larger amount of rain for 29 days before sample collection. The count of generic *E. coli* on contaminated spinach was significantly increased when spinach was exposed to a higher mean and maximum temperature for 8, 9, or 10 days before sample collection or a larger amount of rain for 29 days before sample collection.

The PC analyses results showed that two PCs accounted for 85% of the total variability for weather variables significant at the 20% level in both of the univariable binary and P_count model parts (Table 5). In both PC analyses, several variables describing mean, minimum, and maximum daily temperatures were loaded

TABLE 2 Generic *Escherichia coli* counts on spinach samples[a]

State	Farm	No. of collected samples	No. of samples with positive counts (%)[b]	Geometric mean (CFU/g)	Geometric SD	Minimum (CFU/g)	Maximum (CFU/g)
Colorado	1	115	0 (0)				
	2	120	24 (20.0)	61,852.2	11.9	264.0	1.2×10^7
	3	120	2 (1.7)	37.9	24.1	4.0	3.6×10^2
	4	120	1 (0.8)	4.0		4.0	4.0
Subtotal		475	27 (5.7)	25,018.3	35.6	4.0	2.9×10^6
Texas	1	120	11 (9.2)	619.0	82.2	8.0	6.0×10^7
	2	120	4 (3.3)	11.1	5.0	4.0	1.2×10^2
	3	10	1 (10.0)	4.0		4.0	4.0
	4	95	7 (7.4)	188.5	39.8	4.0	2.6×10^4
	5	25	7 (28.0)	168.0	315.9	6.0	6.0×10^7
	6	60	1 (1.7)	20.0		20.0	20.0
	7	10	1 (10.0)	40.0		40.0	40.0
	8	40	4 (10.0)	202.7	72.9	4.0	8.8×10^4
Subtotal		480	36 (7.5)	157.8	62.2	4.0	6.0×10^7

[a] The geometric mean, geometric standard deviation, minimum, and maximum were estimated after excluding samples with zero counts.
[b] Prevalence.

on the first PC (named "temperature PC"). Several variables describing precipitation were loaded on the second PC. However, in the P_count model part, minimum daily temperature on day 2 prior to sample collection and wind speed on day 2 prior to sample collection were additionally loaded on the second PC. In the univariable analysis of the P_count model part, the score for the temperature PC was significantly associated with the count of generic *E. coli* on contaminated spinach.

In the final mixed-effect NBH model (Table 6), the binary part included the following factors: the presence of "hygiene-field status" factors (OR = 0.06), the mean amount of precipitation (mm) between the day of sample collection and day 29 prior (OR = 3.5), the application of manure fertilizer (OR = 52.2), and farm location in Texas (OR = 108.1). According to this final model, the probability of spinach contamination with generic *E. coli* can be estimated with the equation $p' = \exp(-5.97 - 2.88 \times X_1 + 1.24 \times X_2 + 3.96 \times X_3 + 4.68 \times X_4)/[1 + \exp(-5.97 - 2.88 \times X_1 + 1.24 \times X_2 + 3.96 \times X_3 + 4.68 \times X_4)]$, where X_1 is the presence of "hygiene-field status" factors (1 versus 0), X_2 is the mean amount of precipitation (mm) between the day of sample collection and day 29 prior (continuous), X_3 is the application of manure fertilizer (where "yes" = 1 versus "no" = 0), X_4 is the state (where Texas = 1 versus Colorado = 0). For example, spinach harvested on a farm in Texas ($X_4 = 1$), that was exposed to an average of 1.5 mm of rain over the period of 29 days before harvest ($X_2 = 1.5$), and which applied manure fertilizer ($X_3 = 1$) and provided portable toilets to field workers ($X_1 = 1$), is predicted to have 83.9% of spinach contaminated with generic *E. coli*: $\exp(-5.97 - 2.88 \times 1 + 1.24 \times 1.5 + 3.96 \times 1 + 4.68 \times 1)/[1 + \exp(-5.97 - 2.88 \times 1 + 1.24 \times 1.5 + 3.96 \times 1 + 4.68 \times 1)] = 0.839$.

The P_count part of the final NBH model included the mean amount of rain (mm) for 29 days before sample collection (RR = 1.5 interpreted for txd9 = 0 [Table 6]), the mean of maximum temperatures for 9 days before sample collection (RR = 12.2, interpreted for pd29 = 0 [Table 6]), and a quadratic term for the maximum daily temperature over the past 9 days. According to this final model, the count of generic *E. coli* ($\log_{10}$ CFU/g) on contaminated spinach can be estimated with the equation $\lambda' = \exp(-29.08 + 0.38 \times X_1 + 2.50 \times X_2 - 0.0523 \times X_2^2)$, where X_1 is the mean amount of precipitation (mm) between the day of sample collection and day 29 prior (continuous) and X_2 is the mean of maximum temperatures for 9 days before sample collection (continuous). For example, if over a period of 29 days prior to harvest the mean amount of precipitation was 1 mm ($X_1 = 1$) and if over a period of 9 days prior to spinach harvest the mean of maximum daily temperatures was 25°C ($X_2 = 25$), the expected count of generic *E. coli* is approximately 3.0 $\log_{10}$ CFU/g [$\exp(-29.08 + 0.38 \times 1 + 2.50 \times 25 - 0.0523 \times 25^2)$]. These results indicate that the predicted generic *E. coli* count on contaminated spinach increases linearly between 0 and 3.2 mm of the observed precipitation over the past 29 days (Fig. 1A). However, with an increasing average maximum temperature over the past 9 days, the count increases only until around 24°C, after which it decreases (Fig. 1B). When in multivariable modeling we used the temperature PC score instead of the actual temperature variables, the PC score was not significant and so was not retained in the final model.

We detected a confounding effect of state on the association of spinach contamination probability with the mean amount of rain over the past 29 days ("pd29") [crude OR = 2.4, adjusted OR = 2.9, and change of ln(OR) = 22.0%] and the use of manure [crude OR = 10.4, adjusted OR = 68.9, and change of ln(OR) = 81.0%]. Also, the mean amount of rain for 29 days before sample collection seemed to have confounded the association between the "hygiene-field status" and the probability of spinach contamination [crude OR = 0.14, adjusted OR = 0.05, and change of ln(OR) = 54.1%]. In cross-validation of the final binary part, the mean AUC was 81.1% (range, 79.4% to 83.6%) (Fig. 2). Internal validation of 82.0% also showed good predictability of the model. For the final P_count part, the NRMSE, averaged across cross-validation predictions, was 25% (range, 23% to 27%), with an internal validation of 22.6%.

TABLE 3 Summary statistics for weather factors significantly associated with the count of generic *Escherichia coli* in the univariable negative binomial hurdle models in Table 4[a]

Factor and variable[b]	Value for: Total spinach samples ($n = 955$)			Positive *E. coli* counts ($n = 63$)			Zero *E. coli* counts ($n = 892$)		
	Mean	Median	Range	Mean	Median	Range	Mean	Median	Range
Temp (°C)									
tm2	17.3	17.0	2.7–25.7	16.4	16.3	2.7–24.2	17.3	17.1	2.7–25.7
tmd6	17.1	16.7	3.5–22.8	16.3	15.9	3.5–20.6	17.1	16.8	3.5–22.8
tmd7	16.8	16.7	5.3–22.9	16.8	15.8	5.3–19.5	16.8	16.7	5.3–22.9
tmd8	16.6	16.7	6.7–23.2	16.9	15.7	6.7–18.4	16.6	16.7	6.7–23.2
tmd9	16.6	16.7	7.5–23.3	16.6	15.6	7.5–17.7	16.6	16.8	7.5–23.3
tmd10	16.7	16.7	7.8–23.3	16.7	15.6	7.8–17.7	16.7	16.8	7.8–23.3
tmd15	16.5	16.6	8.2–23.4	16.2	15.4	8.2–18.9	16.6	16.7	8.2–23.4
tmd20	17.1	16.7	8.6–23.4	15.5	15.3	8.6–19.3	17.5	16.8	8.6–23.4
tmd25	17.4	16.8	8.5–23.5	14.6	15.1	8.5–19.4	18.0	16.9	8.5–23.5
tmd29	17.4	16.6	8.4–23.3	13.6	14.8	8.4–19.3	17.5	16.7	8.4–23.3
ti2	10.0	9.6	−8.4–22.0	7.5	9.2	−8.4–22.0	10.0	9.7	−8.4–22.0
tid10	9.5	9.7	0.8–18.7	8.8	8.6	0.8–14.1	9.8	9.8	0.8–18.7
tid15	9.9	9.5	1.0–19.1	7.8	8.4	1.0–13.5	10.1	9.6	1.0–19.1
tx3	26.1	25.9	7.3–36.0	25.8	24.5	7.3–31.7	26.5	26.0	7.3–36.0
txd3	25.8	25.6	16.5–34.3	25.5	24.9	16.5–30.7	26.0	25.7	16.5–34.3
txd4	25.1	25.3	13.4–34.2	24.9	24.3	13.4–30.9	25.1	25.4	13.4–34.2
txd5	25.2	25.0	11.2–34.3	25.1	23.9	11.2–31.0	25.4	25.1	11.2–34.3
txd6	25.3	25.1	13.3–34.3	25.4	24.0	13.3–30.3	25.3	25.2	13.3–34.3
txd7	24.9	25.1	15.0–33.9	25.7	24.0	15.0–28.9	24.9	25.2	15.0–33.9
txd8	24.8	25.1	16.2–33.7	24.6	23.8	16.2–27.1	24.8	25.2	16.2–33.7
txd9	24.9	25.1	15.7–33.7	24.9	23.7	16.7–26.9	24.9	25.2	15.7–33.7
txd10	25.1	25.1	15.8–33.8	25.1	23.8	17.3–26.9	25.1	25.2	15.8–33.8
txd15	24.7	25.0	16.5–33.8	24.7	23.8	17.5–28.0	24.7	25.1	16.5–33.8
txd20	25.0	25.2	16.6–33.7	23.9	23.4	17.4–28.7	25.4	25.3	16.6–33.7
txd25	24.9	25.2	16.7–33.9	23.4	23.1	16.7–28.9	25.7	25.3	16.7–33.9
txd29	25.4	25.0	16.4–33.4	22.1	22.7	16.4–28.3	25.5	25.1	16.4–33.4
Wind speed (m/s)									
ws2	3.0	3.6	1.3–8.7	3.1	3.6	1.8–8.7	3.0	3.6	1.3–8.7
Precipitation (mm)									
pd1	0.0	0.6	0.0–6.0	0.0	0.1	0.0–4.4	0.0	0.6	0.0–6.0
pd5	0.0	1.6	0.0–14.4	1.9	2.8	0.0–14.4	0.0	1.5	0.0–14.4
pd6	0.1	1.5	0.0–13.2	2.0	2.5	0.0–13.2	0.1	1.4	0.0–13.2
pd7	0.2	1.4	0.0–11.5	2.2	2.3	0.0–11.5	0.2	1.3	0.0–11.5
pd8	0.1	1.3	0.0–10.2	2.0	2.1	0.0–10.2	0.1	1.2	0.0–10.2
pd9	0.1	1.3	0.0–9.2	1.8	2.0	0.0–9.2	0.1	1.2	0.0–9.2
pd10	0.3	1.3	0.0–8.4	1.6	1.9	0.0–8.4	0.3	1.2	0.0–8.4
pd25	0.6	1.0	0.0–4.3	1.4	1.5	0.0–3.7	0.6	1.0	0.0–4.3
pd29	0.6	1.0	0.0–3.8	1.7	1.7	0.0–3.2	0.6	1.0	0.0–3.8

[a] Only variables with *P* values of <0.20 in Table 4 are shown.

[b] tm2, average daily temperature on day 2 prior to sample collection (SC); tmdX, mean of the average daily temperatures (°C) in the period between the day of SC and day *X* prior to SC; ti2, minimum temperature on day 2 prior to SC; tidX, mean of the minimum daily temperatures (°C) between the day of SC and day *X* prior to SC; tx3, maximum daily temperature (°C) on day 3 prior to SC; txdX, mean of the maximum daily temperatures (°C) between the day of SC and day *X* prior to SC; ws2, wind speed on day 2 prior to SC; pdX, mean amount of rain (mm) between the day of SC and day *X* prior to SC.

DISCUSSION

Our study investigated the role of farm management, environment, landscape, and weather factors in determining the count of generic *E. coli* on spinach at the preharvest level. The results indicated that the count of generic *E. coli* on spinach at the preharvest level is jointly determined by meteorological (precipitation and temperature), environmental (state), and management (implementation of hygiene practices and application of manure) factors. These risk factors were separated into those that determine occurrence of a contamination event and those that affect the bacterial count given that the contamination event has occurred.

Both parts of the mixed-effect NBH model identified the mean amount of rain over 29 days before sample collection as a risk factor for spinach contamination. Indeed, the odds of spinach contamination with generic *E. coli* could increase by the short-distance dispersal of generic *E. coli* on rainy days due to rain splashing (34, 35), aerosols (34), or high humidity that occurs before or after rain events (14, 36). Alternatively, flooding due to heavy rain may also transmit generic *E. coli* to spinach plants and increase the probability of spinach contamination (15). These previously reported risk factors for an increase in the probability of a contamination event may also partially explain the finding of

TABLE 4 Variables associated with the count of generic *Escherichia coli* ($\log_{10}$ CFU/g) in the binary or P_count part of the univariable mixed-effect negative binomial hurdle models (with the random effects farm and date)

Variable (comparison level)[a]	Reference or unit	Binary part (binomial with logit link)					P_count part (zero-truncated negative binomial with log link)				
		Coeff[b]	SE[c]	OR[d]	95% CI[e]	*P* value[f]	Coeff	SE	RR[g]	95% CI	*P* value[f]
Farm management factors											
Workers_time (>3 days)	≤3 days	−1.15	0.66	0.32	0.09–1.15	0.081	−0.78	0.41	0.46	0.20–1.03	0.060
Hygiene-field status (level 1)[h]	0	−1.97	0.94	0.14	0.02–0.89	0.037					
Organic (yes)	No	1.30	0.81	3.7	0.8–17.9	0.109	0.57	0.30	1.77	0.99–3.16	0.055
Organic_certified (yes)	No	−4.52	2.46	0.01	0.00–1.34	0.066					
Before_fallow (yes)	No	1.76	0.83	5.8	1.2–29.4	0.033					
Tillage_time (>17 days)	≤17 days						−0.60	0.33	0.55	0.29–1.05	0.069
Pesticide_time (>10 days)	≤10 days						−0.60	0.35	0.55	0.27–1.09	0.087
Manure_application (yes)	No	2.34	1.07	10.4	1.3–84.3	0.029	0.75	0.29	2.12	1.19–3.75	0.010
Manure_age (>13 weeks)	≤13 weeks						−0.47	0.32	0.63	0.34–1.16	0.139
Irrigation_time (>5 days)	≤5 days						−0.83	0.26	0.44	0.26–0.73	0.002
Planting_time (>66 days)	≤66 days	0.78	0.55	2.2	0.7–6.5	0.158					
Farm environmental factors											
Proximity_dairy (yes)	No						0.90	0.28	2.46	1.42–4.24	0.001
Proximity_beef (yes)	No	2.20	1.44	9.0	0.5–152.0	0.126	0.90	0.28	2.46	1.42–4.24	0.001
Proximity_poultry (yes)	No	2.43	1.38	11.4	0.8–169.5	0.077	0.90	0.28	2.46	1.42–4.24	0.001
Proximity_residential (yes)	No						0.51	0.31	1.67	0.90–3.08	0.102
Domestic_animal (yes)	No	1.96	0.93	7.1	1.2–43.4	0.034					
Wildlife_control_fences (yes)	No	−1.83	1.41	0.16	0.01–2.53	0.194	−0.99	0.35	0.37	0.19–0.73	0.004
Wildlife_control_hunting (yes)	No	−2.78	1.99	0.06	0.00–3.08	0.163					
State (Texas)	Colorado	2.54	1.53	12.7	0.6–254.4	0.098	−0.62	0.31	0.54	0.29–0.99	0.045
Weather factors											
tm2							0.05	0.04	1.05	0.97–1.14	0.196
tmd6							0.07	0.05	1.07	0.97–1.19	0.172
tmd7	°C	−0.21	0.16	0.81	0.60–1.10	0.182	0.11	0.06	1.11	0.98–1.26	0.093
tmd8	°C	−0.25	0.16	0.78	0.57–1.07	0.128	0.14	0.07	1.15	1.01–1.32	0.042
tmd9	°C	−0.27	0.16	0.77	0.56–1.06	0.104	0.16	0.07	1.18	1.02–1.35	0.023
tmd10	°C	−0.27	0.16	0.77	0.56–1.05	0.100	0.15	0.07	1.16	1.01–1.33	0.033
tmd15	°C	−0.27	0.16	0.76	0.56–1.04	0.087					
tmd20	°C	−0.28	0.16	0.76	0.56–1.03	0.075					
tmd25	°C	−0.28	0.16	0.75	0.56–1.02	0.070					
tmd29	°C	−0.28	0.15	0.76	0.56–1.02	0.064					
ti2							0.04	0.03	1.04	0.98–1.11	0.174
tid10	°C	−0.20	0.15	0.82	0.61–1.10	0.178					
tid15	°C	−0.21	0.15	0.81	0.60–1.08	0.153					
tx3	°C	−0.15	0.10	0.87	0.72–1.04	0.130					
txd3	°C	−0.17	0.13	0.84	0.65–1.09	0.197					
txd4	°C	−0.18	0.12	0.84	0.66–1.06	0.147					
txd5	°C	−0.18	0.12	0.84	0.67–1.06	0.140					
txd6	°C	−0.18	0.13	0.83	0.65–1.06	0.142	0.07	0.05	1.07	0.97–1.17	0.159
txd7	°C	−0.21	0.13	0.81	0.62–1.06	0.120	0.10	0.05	1.10	1.00–1.22	0.057
txd8	°C	−0.23	0.13	0.80	0.61–1.04	0.089	0.13	0.05	1.14	1.03–1.26	0.009
txd9	°C	−0.23	0.13	0.80	0.61–1.03	0.086	0.14	0.05	1.15	1.05–1.26	0.002
txd10	°C	−0.22	0.13	0.80	0.62–1.04	0.099	0.14	0.05	1.15	1.05–1.26	0.003
txd15	°C	−0.22	0.13	0.80	0.62–1.04	0.098	0.08	0.06	1.08	0.97–1.21	0.166
txd20	°C	−0.26	0.13	0.77	0.60–1.00	0.053					
txd25	°C	−0.27	0.13	0.76	0.59–0.99	0.042					
txd29	°C	−0.28	0.13	0.76	0.59–0.98	0.034					
ws2							0.15	0.11	1.16	0.94–1.44	0.163
pd1	mm	−0.54	0.41	0.58	0.26–1.29	0.183					
pd5	mm	0.17	0.12	1.18	0.94–1.48	0.147					
pd6	mm	0.17	0.12	1.19	0.94–1.51	0.156					
pd7	mm	0.21	0.14	1.23	0.93–1.63	0.143					
pd8	mm	0.25	0.16	1.28	0.93–1.76	0.127					
pd9	mm	0.30	0.18	1.36	0.96–1.92	0.087					
pd10	mm	0.28	0.19	1.32	0.91–1.90	0.141					

(Continued on following page)

TABLE 4 (Continued)

Variable (comparison level)[a]	Reference or unit	Binary part (binomial with logit link)					P_count part (zero-truncated negative binomial with log link)				
		Coeff[b]	SE[c]	OR[d]	95% CI[e]	*P* value[f]	Coeff	SE	RR[g]	95% CI	*P* value[f]
pd25	mm	0.58	0.37	1.78	0.86–3.68	0.121	0.24	0.14	1.27	0.96–1.68	0.097
pd29	mm	0.87	0.40	2.37	1.08–5.22	0.031	0.30	0.13	1.34	1.03–1.75	0.027
PC score											
pc1 ("temperature PC score")		0.26	0.14	1.29	0.98–1.7	0.069	−0.13	0.06	0.87	0.78–0.98	0.018
Landscape factor											
Soil_acidity (pH 7.9–9.0)	pH 6.1–7.9						−0.45	0.33	0.64	0.34–1.21	0.169

[a] Variables are defined in Table 1.
[b] Coeff, value of the regression coefficient.
[c] SE, standard error.
[d] OR, odds ratio.
[e] CI, confidence interval.
[f] Only variables with *P* values of <0.20 are shown.
[g] RR, relative risk.
[h] The OR and 95% CI apply to all factors within the composite variable "hygiene-field status." Here, "level 1" indicates the presence of toilet training and use of toilets and washing stations but absence of field grazing and hay production before planting of the spinach during the current growing season.

higher counts of generic *E. coli* on spinach after exposure to a larger amount of rain for 29 days before sample collection. The current study also enabled us to evaluate the changes in generic *E. coli* count as affected by the amount of rain in multiple time windows. In line with our result for the P_count part of the NBH model, a previous study (37) showed that the population of *Salmonella enterica* serovar Thompson on cilantro leaves increased when they were maintained for 2 days under wet conditions (50 or 60% relative humidity) and at 26°C.

In addition to the average precipitation during the 29-day period before sample collection, the binary part of the final NBH model identified significant associations of the presence of spinach contamination with the "hygiene-field status" group, the application of manure fertilizer, and state. The "hygiene-field status" group was reported to have a protective factor on the probability of spinach contamination with generic *E. coli*, when we considered the farm management and environmental factors only (6) or along with weather factors (13). However, interestingly, the "hygiene-field status" group of factors was not significant in univariable analyses for the P_count part of the model, despite the importance of this group in the binary part. The final model of the P_count part did not retain the use of manure fertilizer, although the univariable analysis results showed that the use of manure fertilizer was significantly associated with the count of generic *E. coli* on contaminated spinach. Regarding the farm location (state), the findings of the univariable analyses indicated the opposite effect of state on the probability of a contamination event (OR = 12.7) and on the extent (count) of microbial contamination given a contamination event has occurred (RR = 0.54). Specifically, spinach samples collected in Texas had higher odds of being contaminated but were more likely to be contaminated with low counts (Table 2). This may be explained by relatively unfavorable temperatures for generic *E. coli* counts (Fig. 1B) during the spinach growing season in Texas compared to those in Colorado and unique environments or practices on Texas farms compared to Colorado that favor a higher frequency of contamination events. For the 63 positive samples, the average of maximum temperatures over 9 days before sample collection (txd9) in Texas ranged between 16.7°C and 25.8°C (mean, 21.6°C), while in Colorado the temperature ranged between 24.9°C and 26.9°C (mean, 26.6°C). Indeed, the state factor dropped during multivariable modeling of the P_count part when temperature and state were considered simultaneously, supporting the finding that the state effect may be a proxy for the differences in maximum temperatures during the growing season.

The results of univariable analyses suggested that higher temperatures over a period of time prior to sample collection may have a protective effect on the odds of contamination but that they may increase the risk of occurrence of higher counts of generic *E. coli* on produce. However, neither of the temperature factors identified in the univariable analyses were retained in the binary part of the final model, whereas in the P_count part, temperature was identified as a risk factor, indicating that the extent of spinach contamination with generic *E. coli* is significantly associated with an average maximum daily temperature between the day of sample collection and day 9 prior to sample collection and that the relationship between the bacterial count and temperature follows a quadratic function peaking at around 24°C. It should be noted that variables for the average maximum daily temperatures between the day of sample collection and day 8, 9, or 10 prior to sample collection all produced similar results (albeit with the variable for the 9-day time window showing the best fit). This wider window of times for the influence of the maximum temperature gives more confidence that this finding is not due to the type I error and provides a more realistic biological support for a true effect of temperature on bacterial counts than for example if a single narrow window of time would be identified as a risk factor. Consistent with the results obtained for the binary part in the present study, our previous study using a mixed-effect logistic regression model identified temperature as a protective factor in the univariable analyses, but this factor was dropped from the final multivariable model, which was contributed to a distorting effect of state (13). Because of a higher frequency of contaminated samples and the lower temperatures during the growing season in Texas, higher temperatures appeared to have a protective effect on the odds of contamination (13). However, the results of the pres-

TABLE 5 Principal component analysis for weather variables in Table 3

Variable[a]	Value for:			
	Binary part (binomial with logit link)[b]		P_count part (zero-truncated negative binomial with log link)[b]	
	PC1	PC2	PC1	PC2
tmd7	**−0.21**	0.04		
tmd8	**−0.22**	0.03	**−0.32**	0.10
tmd9	**−0.22**	0.01	**−0.32**	0.07
tmd10	**−0.22**	0.01	**−0.31**	0.05
tmd15	**−0.22**	0.05		
tmd20	**−0.21**	0.06		
tmd25	**−0.21**	0.08		
tmd29	**−0.20**	0.07		
ti2			−0.14	**0.48**
tid10	**−0.18**	0.11		
tid15	**−0.18**	0.12		
tx3	**−0.19**	0.01		
txd3	**−0.20**	−0.03		
txd4	**−0.21**	−0.04		
txd5	**−0.21**	−0.05		
txd6	**−0.21**	−0.05	**−0.30**	0.11
txd7	**−0.21**	−0.04	**−0.32**	0.02
txd8	**−0.21**	−0.05	**−0.31**	−0.10
txd9	**−0.21**	−0.06	**−0.30**	−0.12
txd10	**−0.21**	−0.06	**−0.30**	−0.14
txd15	**−0.22**	0.00	**−0.30**	−0.16
txd20	**−0.22**	0.01		
txd25	**−0.21**	0.03		
txd29	**−0.21**	0.03		
ws2			−0.05	**0.50**
pd1	−0.01	**0.16**		
pd5	0.02	**0.36**		
pd6	0.02	**0.36**		
pd7	0.01	**0.36**		
pd8	0.01	**0.37**		
pd9	0.01	**0.36**		
pd10	0.01	**0.35**		
pd25	0.05	**0.26**	−0.12	**−0.40**
pd29	0.05	**0.24**	−0.13	**−0.46**
StD (proportion of variance [%]; cumulative proportion [%])	4.49 (63; 63)	2.64 (22; 85)	3.27 (67; 67)	1.73 (19; 85)

[a] tmdX, mean of the average daily temperatures (°C) in the period between the day of sampling collection (SC) and day *X* prior to SC; ti2, minimum daily temperature (°C) on day 2 prior to SC; tidX, mean of the minimum daily temperatures (°C) between the day of SC and day *X* prior to SC; tx3, maximum daily temperature (°C) on day 3 prior to SC; txdX, mean of the maximum daily temperatures (°C) between the day of SC and day *X* prior to SC; ws2, wind speed on day 2 prior to SC; pdX, mean amount of rain (mm) between the day of SC and day *X* prior to SC.

[b] Boldface indicates the highest component loading a variable has in the particular model part.

ent study revealed that temperature is a risk factor for microbial count on contaminated spinach, but it does not have any effect on the odds of a contamination event. It has been well documented that warm temperature (38) and high humidity (39) improve the survival and growth rates of the generic *E. coli* population. Interestingly, the relationship between the count of generic *E. coli* on contaminated spinach and the average maximum daily temperature over the past 9 days follows a quadratic function that forms a downward-opened parabola with maximum counts predicted for temperatures around 24°C. Considering that the final model of the P_count part did not retain any farm management, environment, and landscape factors, we can conclude that an increased count of generic *E. coli* on spinach is mainly determined by favorable weather conditions defined by temperature around 24°C and high humidity that support survival and growth of the microorganism.

Univariable analysis indicated that the count of generic *E. coli* on contaminated spinach significantly increased when the farm

TABLE 6 Final mixed-effect negative binomial hurdle model of risk factors associated with the count of generic *Escherichia coli* ($\log_{10}$ CFU/g) on spinach (with the random effects farm and date)[a]

Variable (comparison level)	Reference or unit	Coeff[b]	SE[c]	OR or RR[d]	95% CI[e]	*P* value
Binary part (binomial with logit link)						
Intercept		−5.97	1.58	NA[f]	NA	<.001
Hygiene-field status (level 1)[g]	0	−2.88	0.86	0.06	0.01–0.30	0.001
pd29[h]	mm	1.24	0.38	3.5	1.7–7.2	0.001
Manure_application (yes)	No	3.96	1.37	52.2	3.6–761.4	0.004
State (Texas)	Colorado	4.68	1.52	108.1	5.5–2128.7	0.002
P_count part (zero-truncated negative binomial with log link)						
Intercept		−29.08	12.02	NA	NA	0.016
pd29[i]	mm	0.38	0.11	1.5	1.2–1.8	<.001
txd9[j]	°C	2.50	1.04	12.2	1.6–94.2	0.017
txd9^2	°C	−0.0523	0.02	NA	NA	0.021

[a] Variance component values (standard deviation) were 3.53 (1.88) for date and 0.10 (0.32) for farm in the binary part and 1.62e−7 (4.00e−4) for date and 0.77e−7 (2.77e−4) for farm in the P_count part; in the final P_count part, the random effects were dropped because their effect was negligible. For the intercept-only model, variance component values were 6.11 (2.47) for date and 0.94 (0.97) for farm in the binary part and 0.16 (0.40) for date and 0.11 (0.33) for farm in the P_count model.
[b] Coeff, value of the regression coefficient.
[c] SE, standard error.
[d] OR, odds ratio; RR, relative risk. In the column, ORs are shown for the binary part and RRs are shown for the P_count part.
[e] CI, confidence interval.
[f] NA, not applicable.
[g] The estimated OR (95% CI) applies to all factors within the composite variable "hygiene-field status" group. Here, "level 1" indicates the presence of toilet training and use of toilets and washing stations but absence of field grazing and hay production before planting of the spinach during the current growing season.
[h] pd29, mean amount of rain between the day of sample collection (SC) and day 29 prior to SC.
[i] Estimate for the variable pd29 when txd9 is 0.
[j] Mean of the maximum daily temperatures between the day of SC and day 9 prior to SC; here the estimate is for the variable txd9 when pd29 is 0.

applied manure fertilizer ($P = 0.010$). Likewise, the univariable result of the binary part indicated that the probability of a contamination event increases significantly when the farm applies manure fertilizer ($P = 0.029$). Islam et al. periodically investigated the pathogen concentration ($\log_{10}$ CFU/g) on vegetables grown in manure-amended soil inoculated with *E. coli* O157:H7 and *Salmonella* and concluded that contaminated manure could play an important role in contaminating vegetables (40, 41). Another study by Mukherjee et al. showed that manure increased the risk of *E. coli* contamination of vegetables in organic and semiorganic farms (11). Nevertheless, similar to our previous study (13), explained by the confounding effect of state on the association between manure use and spinach contamination event, manure use was not retained in the final model of the binary part. The association between manure use and microbial counts observed in the univariable analysis of the P_count part may be a side effect of high correlations between manure use and the average maximum daily temperature for 9 days before sample collection (Spearman's correlation coefficient = 0.84; $P < 0.001$) and the average precipitation amount for 29 days before sample collection (Spearman's correlation coefficient = 0.73; $P < 0.001$).

Univariable analysis for generic *E. coli* count on contaminated samples demonstrated that the proximity of dairy, beef, and poultry farms identically increased the level of produce contamination with generic *E. coli*. That is because all 63 positive samples shared the same distribution of these proximity variables due to clustering of contaminated samples: 35 out of 63 positive samples (56%) were collected on 2 out of 12 enrolled farms. Because interpreting individual proximity factors was not meaningful, we could only conclude that the location of a spinach farm near a livestock farm should be considered a risk factor. Future research should address the role of proximity of different livestock production systems in the extent (count) of produce contamination.

An epidemiological study by Strawn et al. (22) reported that irrigation of field more than 3 days prior to sample collection reduced the odds of *L. monocytogenes* isolation from soil and drag swab samples in a produce field. Similarly, our previous study (6) showed that spinach irrigated more than 5 days prior to harvest had nearly 76% lower odds of generic *E. coli* isolation compared to spinach irrigated 5 days or less prior (OR = 0.24). However, importantly, the univariable analysis in the present study indicated that spinach grown on fields irrigated more than 5 days prior to sample collection also had reduced counts of generic *E. coli* (RR = 0.44). This association between microbial count on produce and the time of irrigation was previously demonstrated in controlled trials (42, 43). In our study, irrigation lapse time was not retained in the final model of the P_count part. That may be because irrigation management is dependent on weather conditions, which are better predictors of the bacterial count than irrigation management. Indeed, the irrigation lapse time variable was negatively correlated with the average maximum daily temperature for 9 days before sample collection (Spearman's correlation coefficient = −0.69; $P < 0.001$) and the average precipitation amount for 29 days before sample collection (Spearman's correlation coefficient = −0.66; $P < 0.001$).

The present study showed borderline significant positive associations of the organic versus conventional farming variable on the odds of contamination (OR = 3.7; $P = 0.109$) and the count of generic *E. coli* on contaminated spinach (RR = 1.77; $P = 0.055$) (Table 4). Moreover, the results indicated a borderline significant protective effect of the certified organic versus noncertified organic farming on the probability of contamination (OR = 0.01; $P = 0.066$) (Table 4). These results further stimulate the contro-

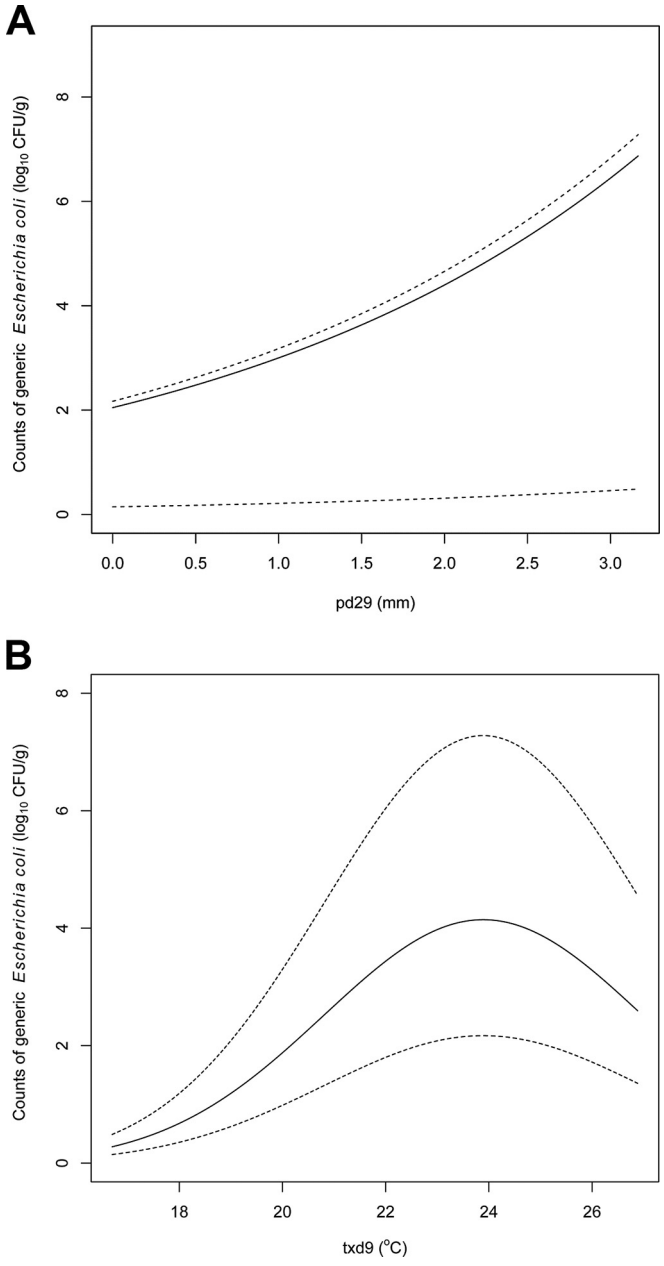

FIG 1 Predicted generic *Escherichia coli* counts ($\log_{10}$ CFU/g) (solid lines) and ranges (dashed lines) on contaminated spinach samples for different values of the mean amount of rain (mm) for 29 days before sample collection (pd29) (A) and for different values of the mean of maximum temperatures (°C) for 9 days before sample collection (txd9) (B). The median values of txd9 (24.94°C) and pd29 (1.693 mm) were held constant for the various levels of pd29 and txd9 shown in panels A and B, respectively.

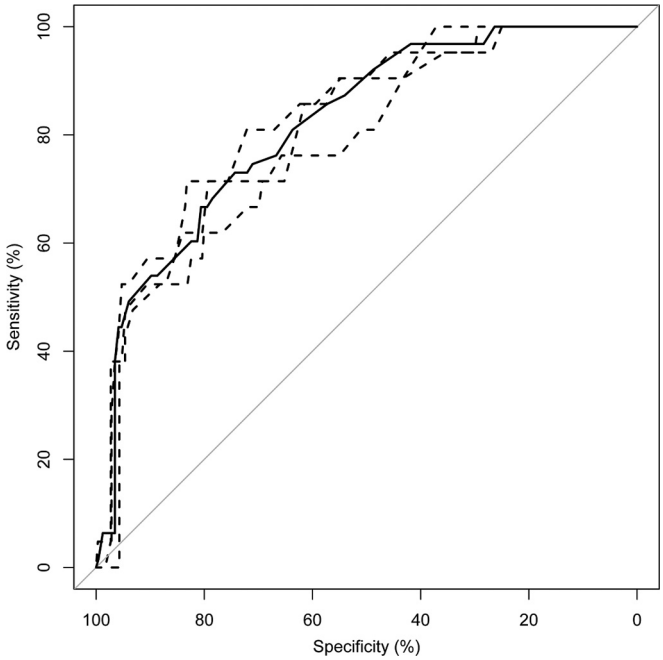

FIG 2 Receiver operating characteristic (ROC) curves for the 3-fold cross-validation (dashed line) and internal validation (solid line) of the binary part.

versy about the effect of organic farm practices on microbial contamination of produce at preharvest (5, 12, 18–20). For example, several studies showed a higher *E. coli* prevalence on organic produce compared to conventionally grown produce (5, 20), but others have not identified significant differences between the two farming types (12, 18, 19). Similarly, published studies (19, 20) have showed inconsistent results about the effect of organic practices on *E. coli* count on produce. A previous study highlighted that humans can be exposed to different bacterial species, depending on the type of produce farming (i.e., organic or conventional) (44). In our study, the organic versus conventional farming and the certified organic versus noncertified organic farming variables were not retained in either part of the final NBH model, suggesting that organic farming is not a predictive factor, at least not major, for the bacterial contamination probability or count. Nevertheless, the observed differences between the roles of organic farming in the probability and extent (count) of produce contamination could be evaluated in future research.

Our results also identified the use of fences to control wildlife intrusion as a protective factor reducing the count of generic *E. coli* on contaminated spinach. Wildlife are well known as carriers of pathogens causing foodborne illnesses. Strawn et al. (22) reported that observation of wildlife within 3 days prior to sampling increases odds of *L. monocytogenes* isolation from soil and drag swab samples in a produce field. Other studies (15, 22) have also suggested wildlife are an important source of pathogen transmission to produce. Nevertheless, to our knowledge, this is the first epidemiological study that was able to show that the extent of microbial contamination (count) on produce could be reduced by using fences to control wildlife intrusion into a produce farm. However, because of a high phi coefficient (0.71) against the proximity of dairy, beef, and poultry farms and with 25% (16/63) missing observations, this variable could not be retained in the final model.

In this study, spinach was collected at different times (between 15 and 123 days) after planting, and the length of time since planting was recorded for analysis ("Planting_time" variable, Table 1). With this approach, we captured plants at different ages (and age-related degrees of leaf rugosity) that cover the ranges of the typical ages when baby, teenage, bunched, or freezer spinach is harvested. Therefore, the "Planting_time" variable can be considered a proxy for the otherwise presumably correlated plant growth stage, the type of spinach at harvest, and rugosity. Because spinach on a

particular field may be cut 2 or 3 additional times in intervals of 20 to 30 days after the initial harvest cut (6), the harvest commodity type alone may not fully explain the cumulative exposure of the plant to the local environmental conditions since planting. In the present study, the "Planting_time" variable was not associated with the count of *E. coli* on contaminated spinach. Interestingly, this variable was significantly associated with the probability of spinach contamination with generic *E. coli* at the univariable level and was part of the final multivariable statistical model when only farm management (including the irrigation lapse time) and environment factors were considered (6). In the present study (and in reference 13), the "Planting_time" variable was significant in the univariable analysis of the probability of spinach contamination albeit at the 20% level (due to a slightly different random structure of the data compared to those in reference 13). However, this variable dropped during multivariable modeling where we also considered weather (precipitation over the past month before sampling). This is in agreement with the study by Gutiérrez-Rodríguez et al. (45), which reported water availability as the dominant factor in survival of *E. coli* in a field trial; water availability depends on irrigation, which in turn depends on the local weather (precipitation) providing the rationale for the link between spinach contamination probability and weather. Overall, these results suggest that the time since planting of spinach is a useful predictor of the probability of spinach contamination, albeit its predictive ability compared to those of other considered risk factors (e.g., precipitation) seems lower. An alternative explanation may be that the time since planting and precipitation over the past month before sampling are both important component causes but of two different sufficient causes (26) of an *E. coli* contamination event.

The analysis of *E. coli* counts on 955 spinach samples described here was part of a larger study where the same 955 samples of spinach, and additionally 191 drag samples of soil and 26 samples of irrigation water, were tested for contamination with *Listeria monocytogenes*, *Salmonella* spp., *E. coli* O157:H7, and *Listeria* spp. (6). *Listeria* and *Salmonella* species were each detected in one spinach sample, and interestingly, these two samples were both from farm 2 in Colorado, which had by far the highest overall count of *E. coli*. The *Listeria*-positive sample also had *E. coli* at the level of 10^5 $\log_{10}$ CFU/g, while *E. coli* was not detected in the *Salmonella*-positive spinach sample. One soil drag sample was positive on the same farm for *Salmonella* spp. These results support that detection of generic *E. coli* in spinach at high levels could be considered an indicator of a recent exposure to risk factors for *Salmonella* and potentially other foodborne pathogens (46), as it has been envisioned by the recent initiative of the EFSA to use *E. coli* as a hygiene criterion in the production of leafy greens (8).

Collectively, risk factors identified in this study help predict the count of generic *E. coli* on spinach. We identified a group of management, environment, and weather factors that increase the odds of a contamination event. This is in agreement with our previous studies (6, 13) that identified the same risk factors as predictors of the odds of spinach contamination with generic *E. coli*. However, the present study went further and identified risk factors that determine the count of generic *E. coli* on contaminated spinach. These newly identified factors may be utilized in risk assessments of produce contamination along the production chain (47) to better predict and explain the variability in produce contamination frequency and extent. Furthermore, the identified risk factors could be used to improve control strategies for microbial count at the preharvest level. For example, farmers could adapt their farm management practices to weather (e.g., delay harvest after rainy season or rush harvest before temperature increase). However, the logistics of these changes may be difficult to implement on commercial produce farms because weather is difficult to predict, and even if confidently predicted, changing management to accommodate weather may cause a cascade of other management changes, such as changes in produce sale times, which may be predetermined by contracts. Preventing occurrence of a contamination event in the first place may be considered a more manageable and potentially far more meaningful task than reducing microbial count once the contamination has occurred. That may be the reason why control strategies depending on weather conditions have not been developed in the produce safety practices outlined in the Food Safety Modernization Act (48), the *Produce GAPs Harmonized Food Safety Standard* (49), *Guide to Minimize Microbial Food Safety Hazards for Fresh Fruits and Vegetables* (50), and California Leafy Green Marketing Agreement (51). Instead, for example, the FDA, regarding weather conditions, recommends that ready-to-eat crops that have been in contact with floodwater should be excluded from the food supply due to their potential exposure to a variety of contaminants (52).

The structure of the final mixed-effect logistic regression model in our previous study of the factors associated with the presence of generic *E. coli* on spinach (13) is identical to the structure of the binary part of the final mixed-effect NBH model in this study. Nevertheless, some of the results were slightly different between the two models. For example, the variable describing if the farm applied hunting to control wildlife intrusion (Table 4) ($P = 0.163$) was considered a potentially important factor ($P < 0.2$) and evaluated in the binary part of the multivariable model in this study. However, due to a higher P value ($P = 0.209$), that variable was not evaluated in the multivariable model in our previous study (13). In addition, 95% confidence intervals (CIs) were slightly different between the two final models: e.g., manure application had a slightly narrower 95% CI (3.6 to 761.4) in the present model than that reported in reference 13 (2.8 to 968.0). These minor differences are likely due to the different statistical packages used; our previous study (13) used the "lmer" function in the "lme4" package (31, 53), whereas this study used "glmmADMB" function in the "glmmADMB" package.

While our choice of a mixed-effect NBH model to evaluate the influence of farm management, environment, and weather factors on the count of generic *E. coli* on spinach was supported by the analysis of the model fit, the model also has good interpretability and practicability. The NBH model consisting of the binary and P_count parts is a suitable statistical method for analysis of the microbial count data on produce in the presence of zero inflation and overdispersion. Because this model simultaneously assesses two different processes: one producing zero or non-zero counts and the other producing positive counts (30), it allows analysis of the microbial count data from two different perspectives. Caution is needed in determining if the hurdle is crossed because a zero count (a negative) can be a truly negative sample or a positive sample but contaminated under the detection limit (i.e., false negative). A false negative may be a problem further down the produce production and distribution chain if conditions in the environment are suitable for bacterial growth. However, in practice, cross-contamination of produce may be considered an equally or even more important factor that determines food safety risk of a

produce commodity. Contamination under the detection limit will not contribute meaningfully to the cross-contamination process and can thus be considered to present a lower (but still present) risk than a detectable level of contamination. In preharvest practice, a hurdle model can be considered the most optimal modeling choice because modeling predictions can be directly compared and verified against observations. Nevertheless, only a small number of previous studies have applied this model in the food safety field, such as on samples from beef carcasses (24, 54). To our knowledge, there are no other studies that investigated risk factors of produce contamination with a microorganism using a hurdle model. In this study, an NBH model enabled us to approach the factors that were significantly associated with either the probability of spinach contamination with generic *E. coli* or the count of generic *E. coli* on contaminated spinach. Based on the predictability using 3-fold cross-validations, our results demonstrated that the probability of spinach contamination with generic *E. coli* and the counts of generic *E. coli* on contaminated spinach can be reliably predicted. To our knowledge, this is the only published study that has modeled, predicted, and validated bacterial counts on produce.

Our study has a few important limitations. There could be a measurement error in the weather data, which were obtained from weather stations that were up to 34.7 km away from the spinach sampling locations. Irrigation water and soil were tested for the presence of pathogens but were not tested for contamination with generic *E. coli*, which prevented us from correlating *E. coli* contamination on spinach with contamination of irrigation water and soil. However, even if such correlation results would be available, the cross-sectional nature of the study design would not provide a strong support for causal inference. The study was conducted in only two states of the United States, which limits generalization of study findings to other regions. The culture-based detection, as applied in the present study, is unable to detect viable but nonculturable (VBNC) cells, and this may vary across seasons and time since planting (55). However, Moyne et al. reported that under field conditions, the loss in culturability of *E. coli* O157:H7 was most likely due to cell death rather than an inability to form colonies on standard media (55). The resurrection of VBNC has been shown possible under optimal laboratory conditions (e.g., with temperature upshift or nutrient availability) (56). However, it is unclear if VBNC cells of foodborne pathogens pose a risk to animal or human health. For example, mice were orally or intraperitoneally inoculated with VBNC *Salmonella enterica* serovar Typhimurium at doses exceeding the 50% lethal dose (LD_{50}) values by approximately 3 orders of magnitude but failed to produce detectable infection (57, 58).

In conclusion, our study showed that the farm management, environment, and weather factors jointly influence the probability of spinach contamination with generic *E. coli*, but once a contamination event has occurred, only weather factors have an effect on the generic *E. coli* count on contaminated spinach. These findings improve our understanding of the mechanisms that determine the bacterial count of generic *E. coli* on spinach (preharvest) and may aid in the development of GAPs and risk assessments to improve produce safety. Furthermore, this study may serve as a methodological template for identification of risk factors that determine the microbial counts on spinach in other regions and on other produce commodities.

ACKNOWLEDGMENTS

This work was supported by Agriculture and Food Research Initiative grant 2009-04261, program code 93233, to R.I. from the U.S. Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA).

Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the USDA-NIFA.

We thank three anonymous reviewers for valuable comments and suggestions.

REFERENCES

1. **Gould LH, Walsh KA, Vieira AR, Herman K, Williams IT, Hall AJ, Cole D.** 2013. Surveillance for foodborne disease outbreaks—United States, 1998-2008. MMWR Surveill Summ **62:**1–34. http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6202a1.htm. Accessed 22 January 2015.
2. **Food and Agriculture Organization, World Health Organization.** 2008. Microbiological hazards in fresh fruits and vegetables: meeting report, microbiological risk assessment series. World Health Organization, Geneva, Switzerland. http://www.who.int/entity/foodsafety/publications/micro/MRA_FruitVeges.pdf. Accessed 22 January 2015.
3. **Sivapalasingam S, Friedman CR, Cohen L, Tauxe RV.** 2004. Fresh produce: a growing cause of outbreaks of foodborne illness in the United States, 1973 through 1997. J Food Prot **67:**2342–2353.
4. **Franz E, van Bruggen AH.** 2008. Ecology of *E. coli* O157:H7 and *Salmonella enterica* in the primary vegetable production chain. Crit Rev Microbiol **34:**143–161. http://dx.doi.org/10.1080/10408410802357432.
5. **Mukherjee A, Speh D, Dyck E, Diez-Gonzalez F.** 2004. Preharvest evaluation of coliforms, *Escherichia coli*, *Salmonella*, and *Escherichia coli* O157:H7 in organic and conventional produce grown by Minnesota farmers. J Food Prot **67:**894–900.
6. **Park S, Navratil S, Gregory A, Bauer A, Srinath I, Jun M, Szonyi B, Nightingale K, Anciso J, Ivanek R.** 2013. Generic *Escherichia coli* contamination of spinach at the preharvest level: the role of farm management and environmental factors. Appl Environ Microbiol **79:**4347–4358. http://dx.doi.org/10.1128/AEM.00474-13.
7. **Tortorello ML.** 2003. Indicator organisms for safety and quality—uses and methods for detection: minireview. JAOAC Int **86:**1208–1217.
8. **Hald T, Baggesen DL.** 2014. EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2014. Scientific opinion on the risk posed by pathogens in food of non-animal origin. Part 2 (*Salmonella* and *Norovirus* in leafy greens eaten raw as salads). http://orbit.dtu.dk/ws/files/104276536/FoNAO_part_2_mar14.pdf. Accessed 22 January 2015.
9. **Natvig EE, Ingham SC, Ingham BH, Cooperband LR, Roper TR.** 2002. *Salmonella enterica* serovar Typhimurium and *Escherichia coli* contamination of root and leaf vegetables grown in soils with incorporated bovine manure. Appl Environ Microbiol **68:**2737–2744. http://dx.doi.org/10.1128/AEM.68.6.2737-2744.2002.
10. **Ogden LD, Fenlon DR, Vinten AJ, Lewis D.** 2001. The fate of *Escherichia coli* O157 in soil and its potential to contaminate drinking water. Int J Food Microbiol **66:**111–117. http://dx.doi.org/10.1016/S0168-1605(00)00508-0.
11. **Mukherjee A, Speh D, Diez-Gonzalez F.** 2007. Association of farm management practices with risk of *Escherichia coli* contamination in preharvest produce grown in Minnesota and Wisconsin. Int J Food Microbiol **120:**296–302. http://dx.doi.org/10.1016/j.ijfoodmicro.2007.09.007.
12. **Mukherjee A, Speh D, Jones AT, Buesing KM, Diez-Gonzalez F.** 2006. Longitudinal microbiological survey of fresh produce grown by farmers in the Upper Midwest. J Food Prot **69:**1928–1936.
13. **Park S, Navratil S, Gregory A, Bauer A, Srinath I, Szonyi B, Nightingale K, Anciso J, Jun M, Han D, Lawhon S, Ivanek R.** 2014. Farm management, environment, and weather factors jointly affect the probability of spinach contamination with generic *Escherichia coli* at the preharvest level. Appl Environ Microbiol **80:**2504–2515. http://dx.doi.org/10.1128/AEM.03643-13.
14. **Park S, Szonyi B, Gautam R, Nightingale K, Anciso J, Ivanek R.** 2012. Risk factors for microbial contamination in fruits and vegetables at the preharvest level: a systematic review. J Food Prot **75:**2055–2081. http://dx.doi.org/10.4315/0362-028X.JFP-12-160.
15. **Orozco RL, Iturriaga MH, Tamplin ML, Fratamico PM, Call JE, Luchansky JB, Escartin EF.** 2008. Animal and environmental impact on the presence and distribution of *Salmonella* and *Escherichia coli* in hydroponic tomato greenhouses. J Food Prot **71:**676–683.

16. **Izumi H, Poubol J, Hisa K, Sera K.** 2008. Potential sources of microbial contamination of Satsuma mandarin fruit in Japan, from production through packing shed. J Food Prot **71:**530–538. http://dx.doi.org/10.1016/j.jprot.2008.08.003.
17. **Izumi H, Tsukada Y, Poubol J, Hisa K.** 2008. On-farm sources of microbial contamination of persimmon fruit in Japan. J Food Prot **71:**52–59.
18. **Allen KJ, Kovacevic J, Cancarevic A, Wood J, Xu J, Gill B, Allen JK, Mesak LR.** 2013. Microbiological survey of imported produce available at retail across Canada. Int J Food Microbiol **162:**135–142. http://dx.doi.org/10.1016/j.ijfoodmicro.2013.01.010.
19. **Bohaychuk VM, Bradbury RW, Dimock R, Fehr M, Gensler GE, King RK, Rieve R, Romero Barrios P.** 2009. A microbiological survey of selected Alberta-grown fresh produce from farmers' markets in Alberta, Canada. J Food Prot **72:**415–420.
20. **Oliveira M, Usall J, Vinas I, Anguera M, Gatius F, Abadias M.** 2010. Microbiological quality of fresh lettuce from organic and conventional production. Food Microbiol **27:**679–684. http://dx.doi.org/10.1016/j.fm.2010.03.008.
21. **Strawn LK, Fortes ED, Bihn EA, Nightingale KK, Grohn YT, Worobo RW, Wiedmann M, Bergholz PW.** 2013. Landscape and meteorological factors affecting prevalence of three food-borne pathogens in fruit and vegetable farms. Appl Environ Microbiol **79:**588–600. http://dx.doi.org/10.1128/AEM.02491-12.
22. **Strawn LK, Grohn YT, Warchocki S, Worobo RW, Bihn EA, Wiedmann M.** 2013. Risk factors associated with *Salmonella* and *Listeria monocytogenes* contamination of produce fields. Appl Environ Microbiol **79:**7618–7627. http://dx.doi.org/10.1128/AEM.02831-13.
23. **Benjamin L, Atwill ER, Jay-Russell M, Cooley M, Carychao D, Gorski L, Mandrell RE.** 2013. Occurrence of generic *Escherichia coli*, *E coli* O157 and *Salmonella* spp in water and sediment from leafy green produce farms and streams on the Central California coast. Int J Food Microbiol **165:**65–76. http://dx.doi.org/10.1016/j.ijfoodmicro.2013.04.003.
24. **Gonzales-Barron U, Kerr M, Sheridan JJ, Butler F.** 2010. Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. Int J Food Microbiol **136:**268–277. http://dx.doi.org/10.1016/j.ijfoodmicro.2009.10.016.
25. **Hirano SS, Nordheim EV, Arny DC, Upper CD.** 1982. Lognormal distribution of epiphytic bacterial populations on leaf surfaces. Appl Environ Microbiol **44:**695–700.
26. **Dohoo IR, Martin W, Stryhn H.** 2010. Veterinary epidemiologic research, 2nd ed. VER, Inc, Prince Edward Island, Canada.
27. **Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM.** 2009. Mixed effects models and extensions in ecology with R. Springer, New York, NY.
28. **US Department of Agriculture.** 2014. Vegetables 2013 summary. US Department of Agriculture, Washington, DC. http://www.clientadvisoryservices.com/Downloads/VegeSumm-03-27-2014.pdf. Accessed 22 February 2015.
29. **Ivanek R, Grohn YT, Wells MT, Lembo AJ, Jr, Sauders BD, Wiedmann M.** 2009. Modeling of spatially referenced environmental and meteorological factors influencing the probability of *Listeria* species isolation from natural environments. Appl Environ Microbiol **75:**5893–5909. http://dx.doi.org/10.1128/AEM.02757-08.
30. **Hilbe JM.** 2011. Negative binomial regression, 2nd ed. Cambridge University Press, New York, NY.
31. **Bolker B, Skaug H, Magnusson A, Nielsen A.** 2012. Getting started with the glmmADMB package. http://glmmadmb.r-forge.r-project.org/glmmADMB.pdf. Accessed 22 January 2015.
32. **Wearden S.** 2010. Phi coefficient, p 1234-1235. *In* Weiner IB, Craighead WE (ed), The Corsini encyclopedia of psychology, 4th ed, vol 3. John Wiley & Sons, Inc, New York, NY.
33. **Hatcher L.** 1994. A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. SAS Institute, Inc, Cary, NC.
34. **Cevallos-Cevallos JM, Gu G, Danyluk MD, Dufault NS, van Bruggen AH.** 2012. *Salmonella* can reach tomato fruits on plants exposed to aerosols formed by rain. Int J Food Microbiol **158:**140–146. http://dx.doi.org/10.1016/j.ijfoodmicro.2012.07.009.
35. **Madden L, Yang X, Wilson L.** 1996. Effects of rain intensity on splash dispersal of *Colletotrichum acutatum*. Phytopathology **86:**864–874. http://dx.doi.org/10.1094/Phyto-86-864.
36. **Beuchat LR.** 2006. Vectors and conditions for preharvest contamination of fruits and vegetables with pathogens capable of causing enteric diseases. Br Food J **108:**38–53. http://dx.doi.org/10.1108/00070700610637625.
37. **Brandl MT, Mandrell RE.** 2002. Fitness of *Salmonella enterica* serovar Thompson in the cilantro phyllosphere. Appl Environ Microbiol **68:**3614–3621. http://dx.doi.org/10.1128/AEM.68.7.3614-3621.2002.
38. **Cooper VS, Bennett AF, Lenski RE.** 2001. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. Evolution **55:**889–896. http://dx.doi.org/10.1554/0014-3820(2001)055[0889:EOTDOG]2.0.CO;2.
39. **Cox CS.** 1966. The survival of *Escherichia coli* sprayed into air and into nitrogen from distilled water and from solutions of protecting agents, as a function of relative humidity. J Gen Microbiol **43:**383–399. http://dx.doi.org/10.1099/00221287-43-3-383.
40. **Islam M, Doyle MP, Phatak SC, Millner P, Jiang X.** 2004. Persistence of enterohemorrhagic *Escherichia coli* O157:H7 in soil and on leaf lettuce and parsley grown in fields treated with contaminated manure composts or irrigation water. J Food Prot **67:**1365–1370.
41. **Islam M, Morgan J, Doyle MP, Phatak SC, Millner P, Jiang X.** 2004. Persistence of *Salmonella enterica* serovar Typhimurium on lettuce and parsley and in soils on which they were grown in fields treated with contaminated manure composts or irrigation water. Foodborne Pathog Dis **1:**27–35. http://dx.doi.org/10.1089/153531404772914437.
42. **Ibekwe AM, Watt PM, Shouse PJ, Grieve CM.** 2004. Fate of *Escherichia coli* O157:H7 in irrigation water on soils and plants as validated by culture method and real-time PCR. Can J Microbiol **50:**1007–1014. http://dx.doi.org/10.1139/w04-097.
43. **Solomon EB, Pang HJ, Matthews KR.** 2003. Persistence of *Escherichia coli* O157:H7 on lettuce plants following spray irrigation with contaminated water. J Food Prot **66:**2198–2202.
44. **Leff JW, Fierer N.** 2013. Bacterial communities associated with the surfaces of fresh fruits and vegetables. PLoS One **8:**e59310. http://dx.doi.org/10.1371/journal.pone.0059310.
45. **Gutierrez-Rodriguez E, Gundersen A, Sbodio AO, Suslow TV.** 2012. Variable agronomic practices, cultivar, strain source and initial contamination dose differentially affect survival of *Escherichia coli* on spinach. J Appl Microbiol **112:**109–118. http://dx.doi.org/10.1111/j.1365-2672.2011.05184.x.
46. **Jensen AN, Storm C, Forslund A, Baggesen DL, Dalsgaard A.** 2013. *Escherichia coli* contamination of lettuce grown in soils amended with animal slurry. J Food Prot **76:**1137–1144. http://dx.doi.org/10.4315/0362-028X.JFP-13-011.
47. **Danyluk MD, Schaffner DW.** 2011. Quantitative assessment of the microbial risk of leafy greens from farm to consumption: preliminary framework, data, and risk estimates. J Food Prot **74:**700–708. http://dx.doi.org/10.4315/0362-028X.JFP-10-373.
48. **US Food and Drug Administration.** 2013. Standards for the growing, harvesting, packing and holding of produce for human consumption, a proposed rule. http://www.gpo.gov/fdsys/pkg/FR-2013-01-16/pdf/2013-00123.pdf. Accessed 22 January 2015.
49. **US Department of Agriculture.** 2013. Produce GAPs harmonized food safety standards. US Department of Agriculture, Washington, DC. http://www.unitedfresh.org/assets/Harmonized%20Standard%20-%20pre-farm%20gate%20130501.pdf. Accessed 22 January 2015.
50. **US Food and Drug Administration.** 2008. Guide to minimize microbial food safety hazards for fresh fruits and vegetables. Request for comments and for scientific data and information. (Docket no. FDA–2008–N–0455.) Fed Regist **73:**51306–51309. http://www.gpo.gov/fdsys/pkg/FR-2008-09-02/pdf/E8-20187.pdf. Accessed 22 January 2015.
51. **Leafy Green Marketing Agreement.** 2013. Commodity specific food safety guidelines for the production and harvest of lettuce and leafy greens. http://www.caleafygreens.ca.gov/sites/default/files/California%20LGMA%20metrics%2008%2026%2013%20%20Final.pdf. Accessed 22 January 2015.
52. **US Food and Drug Administration.** 2005. Letter to California firms that grow, pack, process, or ship fresh and fresh-cut lettuce. US Food and Drug Administration, Silver Spring, MD. http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ProducePlantProducts/ucm118911.htm. Accessed 22 January 2015.
53. **Bates D, Maechler M, Bolker B.** 2012. lme4: linear mixed-effects models using S4 classes. http://cran.r-project.org/web/packages/lme4/lme4.pdf. Accessed 22 January 2015.
54. **Gonzales-Barron U, Cadavez V, Butler F.** 2014. Conducting inferential statistics for low microbial counts in foods using the Poisson-gamma regression. Food Control **37:**385–394. http://dx.doi.org/10.1016/j.foodcont.2013.09.032.
55. **Moyne AL, Harris LJ, Marco ML.** 2013. Assessments of total and viable

Escherichia coli O157:H7 on field and laboratory grown lettuce. PLoS One **8:**e70643. http://dx.doi.org/10.1371/journal.pone.0070643.
56. **Li L, Mendis N, Trigui H, Oliver JD, Faucher SP.** 2014. The importance of the viable but non-culturable state in human bacterial pathogens. Front Microbiol **5:**258. http://dx.doi.org/10.3389/fmicb.2014.00258.
57. **Smith RJ, Kehoe SC, McGuigan KG, Barer MR.** 2000. Effects of simulated solar disinfection of water on infectivity of *Salmonella typhimurium*. Lett Appl Microbiol **31:**284–288. http://dx.doi.org/10.1046/j.1472-765x.2000.00815.x.
58. **Smith RJ, Newton AT, Harwood CR, Barer MR.** 2002. Active but nonculturable cells of *Salmonella enterica* serovar Typhimurium do not infect or colonize mice. Microbiology **148:**2717–2726.