Food Microbiology 77 (2019) 137–145

Contents lists available at ScienceDirect

Food Microbiology

journal homepage: www.elsevier.com/locate/fm

Evaluation of sampling methods for the detection of pathogenic bacteria on pre-harvest leafy greens

Aixia Xu[a], Robert L. Buchanan[a,b,*]

[a] Department of Nutrition and Food Science, University of Maryland, College Park, MD, 20742, USA
[b] Center for Food Safety and Security Systems, University of Maryland, College Park, MD, 20742, USA

ARTICLE INFO

Keywords:
Random sampling
Stratified-random sampling
Z-pattern sampling

ABSTRACT

Recent outbreaks of foodborne disease associated with leafy greens have led to increased pre-harvest testing for pathogens and indicator microorganisms. However, the scientific and statistical rationale and the performance attributes for pre-harvest sampling methods are not well understood. The performance of three pre-harvest sampling methods, random, stratified random, and Z-pattern sampling, was evaluated by consideration of their mathematical derivations, computer simulations and field validation. Consideration of the probabilistic basis of the sampling methods indicated that the mean detection rates were similar. However, use of simulation modeling to assess the uncertainty associated with the three sampling methods indicated that the inherent variability of the Z-pattern sampling method was substantially greater than the other two sampling methods. A simulation tool was developed in Matlab that allowed the evaluation of the effectiveness of the three sampling methods. A limited validation study also observed that Z-pattern sampling had higher variability than the other two sampling methods. This study indicates that while the mean detection probabilities for the three sampling methods are similar, the random or stratified random sampling are less variable, particularly when the number of contamination sites or number of samples analyzed are small.

1. Introduction

Leafy greens typically grow near the soil's surface and have large surface-to-volume ratios, making them particularly susceptible to contamination (Macarisin et al., 2013). This potential for contamination combined with their consumption without cooking make leafy greens at increased risk for foodborne disease from various foodborne pathogens (Park et al., 2013). Leafy green producers and their clients are increasingly requiring pre-harvest microbiological testing of leafy greens just prior to harvest as part of their risk management programs. Such testing may target indicator microorganisms (e.g., coliforms, generic *Escherichia coli*) or specific pathogens (e.g., *Salmonella enterica*) as a means of verifying that good agricultural practices and other pre-harvest mitigation steps have been effectively implemented.

While there has been a substantial amount of research on detection methods for various pathogenic microorganisms on leafy greens and other produce, there has been little evaluation of the performance of sampling methods, and that has been largely focused on food manufacturing facilities (Jongenburger et al., 2011, 2015). Comparision of different sampling methods for the detection of foodborne pathogens on leafy greens appear not to have been systematically evaluated, but have been derived from historical approaches or adapted from other applications (e.g., pesticide residue evaluations). However, establishment of effective microbiological testing programs is dependent on an understanding of the sampling methods. This requires knowledge of both the statistical principles that are the foundation of microbiological sampling and methods for measuring the uncertainty associated with selected sampling schemes. Considering the large batch sizes associated with pre-harvest leafy greens, the low frequency and levels of pathogens or indicator microorganisms, and the heterogeneity of pathogen distribution in leafy greens fields, it is critical that appropriate sampling methods are selected that optimize the statistical power of the microbiological testing and allow appropriate interpretation of the results (Zwietering et al., 2014). However, despite the fact that pre-harvest testing of leafy greens is a common practice, there are few systematic comparisons of the sampling methods. Thus, the objectives of this study were to (i) review the mathematical principles underlying three commonly used sampling methods, (ii) use simulation modeling to assess the relative uncertainty associated with the three sampling methods, and (iii) perform limited field trials as an initial means of validating the findings.

* Corresponding author. Department of Nutrition and Food Science, University of Maryland, College Park, MD, 20742, USA.
E-mail addresses: xuaixiaasia@gmail.com (A. Xu), rbuchana@umd.edu (R.L. Buchanan).

https://doi.org/10.1016/j.fm.2018.09.007
Received 5 April 2017; Received in revised form 19 August 2018; Accepted 5 September 2018
Available online 05 September 2018

2. Materials and methods

2.1. Simulations

Three commonly used sampling methods, random sampling, stratified random sampling, and Z-pattern sampling, were compared by first evaluating their mathematical basis and then characterizing their relative performance using simulation modeling. This combination of approaches allowed estimation of both mean detection probabilities and the relative uncertainty of detection likelihood.

A variety of sampling methods can be employed, either individually or in combination. For example, simple random sampling selects all subsets of a batch of food product with equal probability. Systematic sampling selects every k*th* elements in the batch or field. Stratified random sampling first divides a batch or field into distinct, independent strata, and then randomly selects equal numbers of samples from each stratum. Choice of pre-harvest sampling method depends on nature and quality of the leafy greens, the prevalence and likely sources of contamination, the degree of confidence required, and the cost and related operational concerns associated with microbiological testing (e.g., the time needed to complete the microbiological analyses).

2.1.1. Description of model fields used to compare sampling methods

For the purposes of this study, we considered a generic, rectangular leafy green field that was divided into $N_x \times N_y$ plots. Each plot, in turn, was subdivided into $S_x \times S_y$ subplots. The total number of subplots was $N_{subplot} = N_x N_y S_x S_y$. We assumed that the field was contaminated N_c times ($N_c > 0$,)(the number of contaminated subplots $\leq N_c$), with each contamination event randomly contaminating a subplot. For example, Fig. 1 depicts the actual field layout used in a subsequent blinded field trial, where we used a 5 × 6 plot array, with each plot having 9 subplots ($N_x = 5$, $N_y = 6$, $S_x = 3$, $S_y = 3$). Each subplot was 1.0 m^2, with the entire field being 15 × 18 m. The x's within the field diagrams (Fig. 1) represent the 30 subplots selected for sampling ($N_{sample} = 30$) for one iteration of a simulation.

2.1.2. Mathematical basis for sampling efficiency

The underlying equations for detection probabilities for the three sampling methods were derived on probability principles outlined in Grinstead and Snell (2003).

2.1.2.1. Random sampling. Simple random sampling is a strategy wherein each part of the field has an equal chance of being selected (European Commission, 2007). A random sampling method selects N_{sample} subplots randomly from the total of $N_{subplot}$ subplots. Thirty random samples were selected as shown in Fig. 1-A.

If the sampling is without replacement, meaning exactly N_{sample} unique samples were selected. The probability that none of the contaminated subplots is detected is:

$$P_{miss} = 1 - P_d = \left(1 - \frac{N_{sample}}{N_{subplot}}\right)^{N_c} \quad (1)$$

Where:

N_{sample} = number of samples taken
$N_{subplot}$ = number of subplots in the field
N_c = number of times the field is randomly contaminated
P_{miss} = probability that no positive sample is detected
P_d = detection probability = 1 − P_{miss}

If the sampling is with replacement, the probability that a subplot is contaminated (P_c) is:

$$P_C = 1 - \left(1 - \frac{1}{N_{subplot}}\right)^{N_c} \quad (2)$$

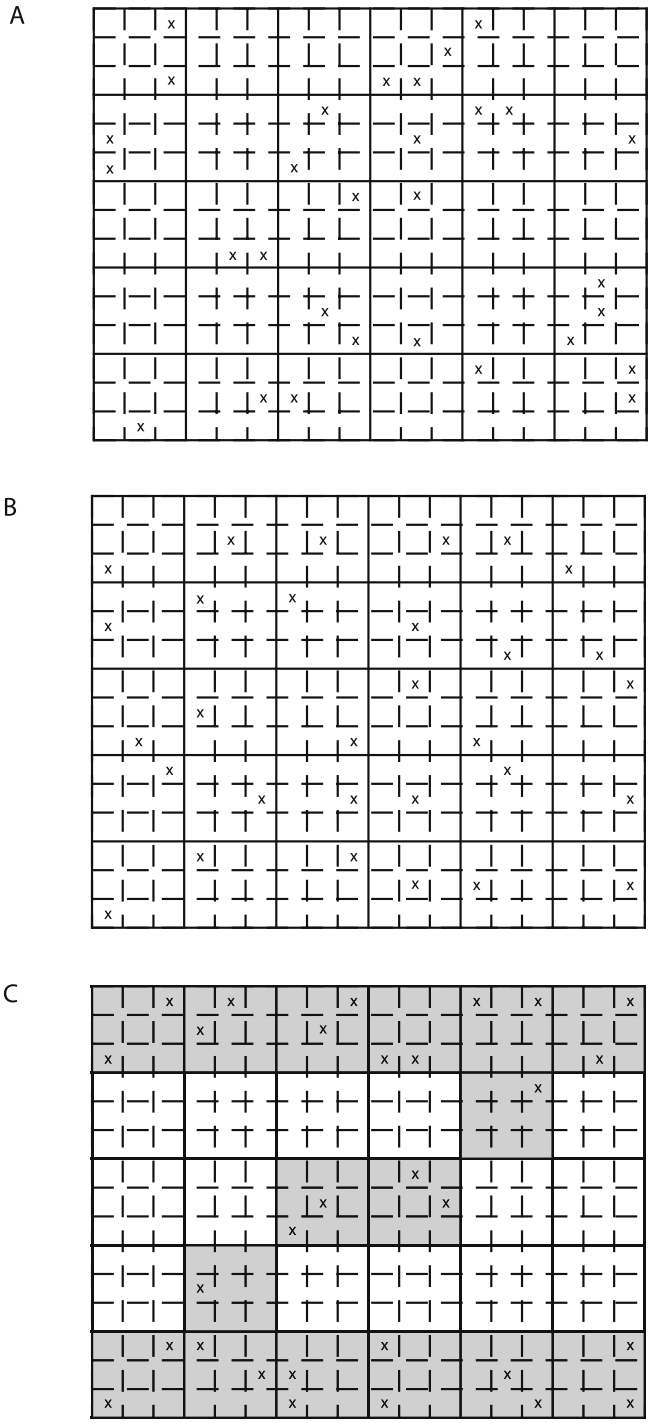

Fig. 1. Example of samples drawn according to (A) random, (B) stratified random, and (C) Z-pattern (see shaded plots) sampling (N_{sample} = 30). The solid lines represent 30 (5 × 6) plots and the dashed lines represent the subplots (9 subplots/plot) for a total of 9 × 30 = 270 subplots. Random sampling: samples were randomly selected chosen from the 270 subplots. Stratified sampling: each plot has one subplot that was randomly selected for sampling. Z-pattern sampling: each plot within "Z" area has one or two subplots randomly selected for sampling. No samples were selected in the plots outside the Z-pattern.

The probability that a subplot is selected by random sampling is:

$$P_s = 1 - \left(1 - \frac{1}{N_{subplot}}\right)^{N_{sample}} \quad (3)$$

Since the contamination is independent of the sampling procedure,

the probability that a subplot is both contaminated and selected by the random sampling is $P_C P_s$. The probability that the random sampling fails to detect any contaminated sample is:

$$P_{miss} = (1 - P_c P_s)^{N_{subplot}} \quad (4)$$

The random sampling with replacement always has a higher "miss" probability (P_{miss}) than the random sampling without replacement because the number of unique samples drawn from the population will be smaller or equal to N_{sample} for random sampling with replacement.

2.1.2.2. Stratified random sampling. Stratified random sampling randomly selects one sample from each plot. An example set of samples drawn according to the stratified random sampling plan is shown in Fig. 1-B. The number of unique samples would normally be exactly $N_{sample} = N_x N_y$ (though one could consider multiple samples per plot, e.g., 2 samples per plot for a total N_{sample} of 60). For each contamination site, the probability that the contamination site is not selected is (1-1/$S_x S_y$). The probability that no contaminated subplot is detected is:

$$P_{miss} = \left(1 - \frac{1}{S_x S_y}\right)^{N_c} \quad (5)$$

Note that because $N_{subplot} = N_x N_y S_x S_y$ and $N_{sample} = N_x N_y$, then $\frac{1}{S_x S_y} = \frac{N_{sample}}{N_{subplot}}$. The mean detection probability of the stratified random sampling method is identical to that of the random sampling (without replacement).

2.1.2.3. Z-pattern sampling. The Z-pattern sampling method only samples from plots lying at two-opposing edges and in one of the diagonal lines (Fig. 1-C) (Zehnder, 2014; Paetz and Wilke, 2005). The number of subplots that lay within the selected "Z" region is N_z. With this sampling method, N_z = 144 (Fig. 1-C). For each contamination site, the probability that a contaminated subplot lies in the "Z" region is:

$$p = \frac{N_z}{N_{subplot}} \quad (6)$$

The probability that k contamination sites lie in the "Z" region is

$$P_k = \binom{N_c}{k} p^k (1 - p)^{N_c - k} \quad (7)$$

Given that there are k contamination sites in the "Z" region, the probability that none of the contamination sites is detected is:

$$P_{miss}^k = \left(1 - \frac{N_{sample}}{N_z}\right)^k \quad (8)$$

The overall probability of non-detection is:

$$P_{miss} = \sum_{k=0}^{N_c} P_k P_{miss}^k \quad (9)$$

2.1.3. Comparison of sampling methods using simulation modeling

A simulation model was developed in Matlab (The MathWorks Inc., Natick, MA, ver. 2015b) software as a means of more fully characterizing the three sampling methods. Each field was contaminated N_c times. In each time, a plot is randomly selected and a subplot is randomly selected within the plot to be contaminated. This generated a randomly contaminated field with N_c or less than N_c subplots contaminated. The sampling of the field was then simulated using the three sampling methods, with 100 iterations per simulation. For each iteration, the sample locations were generated according to different sampling methods independently. A detection probability was estimated based on the number of iterations wherein a sampling method successfully detected at least one of the contaminated sites. The simulation was then repeated 100 times, each with a different contaminated field, to estimate the distribution of detection probabilities for each sampling method. It is important to repeat the simulation multiple times as one sampling method may be good at detecting certain contamination patterns but not others. This was particularly important with the Z-pattern sampling where specific plots are not tested if they fall outside the Z-pattern. Thus, the simulations examined a total of 10,000 combinations of random contamination site assignments and random sampling site assignments based on the number of contamination sites and the number of sampling sites specified.

2.2. Validation trials

Two limited, "blinded" field trials were conducted at the Beltsville Agricultural Research Center, Beltsville, MD in October 2015 and June 2016, using three fields of soybeans and three fields of lettuce, respectively. For each experimental trial, three fields were inoculated with a known indicator microorganism by one group, and the field subsequently sampled by another group using three sampling methods without knowledge of where the fields were inoculated.

The two validation trials employed different methods for the assignment of contamination and sampling sites. In the first trial, a GPS unit was used for site assignments. The second trial employed a more traditional "stake and string" grid for site assignments.

2.2.1. Inoculation

2.2.1.1. Preparation of dairy solids extracts. Fresh dairy cow biosolids were collected from the USDA/ARS/Beltsville Agricultural Research Center. The biosolids (100 g) were diluted 1:10 with deionized water in a 2-L sterile plastic beaker, stirred manually and with a magnetic stirring bar for 5 min before the slurry was filtered through two layers of sterile cheese cloth in a Buchner funnel and collected in 4-L flask. The extract was transferred to 9-L or 20-L carboys, where an equal volume of deionized water was added prior to autoclaving for 1 h at 121 °C. The resulting sterile dairy solids extract (DSE) was stored at 4 °C until used.

2.2.1.2. Strains and culture conditions used. A rifampicin-resistant (RifR), non-pathogenic *E. coli* strain (TVS 355) was graciously provided by Dr. Trevor Suslow at the University of California Davis. The strain was cultured from frozen stocks stored at −80 °C on MacConkey agar (Neogen, Lansing, MI) supplemented with 80 μg/ml rifampicin (Sigma Aldrich, St. Louis, MO) (MACR) and incubated at 42 °C for 24 h. Three to five colonies of the *E. coli* strain were inoculated to 200 ml of tryptic soy broth (Neogen) with 80 μg/ml rifampicin (TSBR) and incubated at 37 °C for 24 h. A 200-ml portion of each 24-h culture was added to separate 7-L carboys of sterile DSE and incubated at 37 °C for 48 h. The population density of *E. coli* after the 48-h incubation was determined by preparing serial dilutions with sterile 0.1% peptone water (Becton Dickson, Sparks, MD) and plating onto MacConkey Agar (Becton Dickinson, Sparks, MD) with 80 μg/ml rifampicin (MACR) and 3M™ Petrifilm™ *E. coli*/Coliform Count films. The plates and Petrifilms were incubated at 37 °C for 18–24 h, and *E. coli* levels enumerated. The carboys were transported to the field location, mixed by manual shaking, and were poured into a 13-L backpack sprayer (H.D. Hudson Manufacturing Company, Chicago, IL) immediately prior to spray inoculation of the experimental field.

2.2.1.3. Inoculum application. Prior to inoculation of the field, 20 samples were taken to determine any background incidence of rifampicin-resistant *E. coli*. None were detected.

Inoculation was carried out by backpack application onto leaves. *Escherichia coli* levels in the inoculum were 2.3×10^7 CFU/ml. Inoculations were performed on a sunny day. The field was divided into 30 (5 × 6) plot and each plot was subdivided into 3 × 3 1.0-m subplots. The total number of subplots per field was $N_{subplot}$ = 270. The field sites to be inoculated were assigned using a random number generator and the plot and subplot numbers recorded. The GPS coordinates of each

location (latitude and longitude) of each subplot center was calculated by computer in advance. For the first trial, a GPS meter (Bad Elf GNSS Surveyor (BE-GPS-3300)) was used to locate each of the ten contamination plot/subplots to be inoculated. In the second trial, stakes and strings were used to identify the 270 subplots. Ten subplots were assigned to be inoculated for each field.

After reaching the designated location of the plot/subplot selected randomly, the backpack sprayer was used to apply the inoculum within the 1.0-m^2 area. Aliquots of 417 ml (25 s by sprayer) of inoculum were applied to each selected 1-m^2 subplot. Care was taken to avoid spraying any inoculum outside the designated subplot. The team applying the inoculum recorded the exact positions within the subplot that was inoculated but did not share that information with the testing team until after all of the analyses were completed. The frequency of spraying of each subplot depended on the frequency they were selected (i.e., a subplot sprayed twice if it was selected twice).

2.2.2. Sampling in the field

The sampling team sampled the field the day after inoculation. The sampling team took a total of 30 samples for each of the three sampling methods from each of the three replicate fields. The sampling sites were selected using a random number generator. For random sampling, a new set of random numbers of plot (1–30) and subplot (1–9) were randomly generated. For stratified random sampling, a sample was taken from each of the 30 plots, with the subplot to be sampled within each plot selected by a random number generator. For Z-pattern sampling, the plots sampled were fixed. Plots 1, 2, 3, 4, 5, 6, 15, 16, 25, 26, 27, 28, 29, and 30 were sampled twice. Plots 11 and 20 were only sampled once. The specific subplots to be sampled were again selected using a random number generator. After recording the plot and subplot numbers (and GPS locations for the first trial), the sampling team proceeded to either the designated GPS location (trial 1) or the designated grid location (trial 2) and conducted the sampling.

2.2.3. Sample processing

At each of the 30 locations for each of three sampling methods, leaf samples of approximately 25 g (10 leaves) were collected, transferred to a sterile sample bag, placed in a cooler, and transported to the laboratory. Samples were processed within 24 h of collection. For each sample, 25 g of leaves were weighed and transferred to a sterile Whirlpak bag (Nasco, Jackson,WI). A 50-ml aliquot of sterile buffered 0.1% peptone water (BPW) was added to each bag and stomached in a laboratory stomacher (Seward, Stomacher 400 circulator, U.K.) for 2 min at 250 rpm at room temperature. A 1-ml aliquot was then pipetted onto 3M™ Petrifilm™ *E. coli*/Coliform Count films and MACR plates. The presence/absence results were determined after 24–48 h incubation.

Data analyses were performed with Matlab. The number of positive samples for each sampling method was calculated. The spatial relationship between positive samples and contamination sites were analyzed visually and statistically as a means of evaluating the results for systematic errors.

3. Results

3.1. Simulation-based analysis

The theoretical analysis of the mean detection probability was validated by computer simulation. The performance of the three sampling methods was characterized by running a series of simulations with different numbers of contamination sites and different numbers of samples. The detection probabilities as a function of the number of contamination sites by probability theory are depicted in Fig. 2. As expected, the detection probability is higher with increasing numbers of contaminated subplots ($\leq N_C$) and increasing number of samples analyzed.

The characteristics of the three sampling methods were further evaluated by considering the distribution of the detection probabilities among the 100 simulations (each with 100 iterations) by box plots (Fig. 3). No significant differences in median detection probabilities were observed among the three sampling methods; they all agreed well with the theoretical predictions when a sufficient number of simulations (and iterations) were performed. However, there were differences in the minimum, first quartile, third quartile and maximum values observed among the three sampling methods, i.e., the Z-pattern testing consistently had larger dispersions.

Next, the number of subplots per plot was evaluated for its effect on the mean detection probability. In the original simulations, each plot was divided into 3 × 3 subplots. In this evaluation the number of subplots per plot was varied from 2 × 2 to 10 × 10, so the total subplots in the field ranged from 120 to 3000, respectively (Fig. 4). Assuming the area of the main plot is fixed, the number of subplots is inversely related to the sampling area. We assumed that single contamination sites cover a single subplot regardless of the size of the subplot. Thus, a 10 × 10 subplot scheme would have a sample area 25 times smaller than a 2 × 2 subplot scheme. The mean detection probability decreases rapidly as a function of total number of subplots. The area contaminated similarly decreased with increasing number of subplots. This relationship is roughly linear on a log-log plot (Fig. 4). There was no difference among the three sampling methods in terms of the mean probability of detection.

As indicated above, while the median detection probabilities among the three sampling methods are similar, differences of the detection probabilities among the simulations were noted (Fig. 3). The distribution of mean detection probabilities across the 100 iterations per simulation was further evaluated as a function of contamination sites and number of samples taken. Fig. 5a–d provides examples of the impact of contamination site numbers (2 vs. 6) and number of samples taken (10 vs. 30). The detection probability dispersion of the Z-pattern sampling was consistently greater than that observed with the other sampling methods. This difference was most evident when the number of contamination sites was small, where the Z-pattern sampling had zero detection probability for a larger fraction of the simulations. This is due, in part, to the Z-pattern sampling only including a fraction of the field's subplots. If a contaminated site lies in the Z-pattern, it would have greater detection probability than the other sampling methods, but if the contamination site was not in the sampled plots, sampling would fail to detect the contamination completely, regardless of sample numbers.

We additionally used simulation modeling to consider when there are multiple “samplers” or when a field is sampled multiple times (assuming that new contamination sites have not been introduced between sampling times). The impact of multiple samplers or sampling times was then combined to calculate overall mean detection probability (i.e., the probability that at least one sampler detects at least one positive sample). An example of the impact of multiple samplers (3–6) each taking 30 samples in conjunction with 1–6 contamination sites was evaluated via a single simulation, each with 100 iterations (Fig. 6a–d). As anticipated, increasing the number of samplers, and thus the total number of samples, increased the mean detection probability. Also as anticipated, increasing the number of contamination sites increased the mean detection probability. The results with random and stratified random sampling were similar whereas the Z-pattern sampling had lower mean detection probabilities, and this difference got greater as the number of samplers increased. When there were three samplers (Fig. 6a), the mean detection probability was similar to the other sampling methods, as was observed with one sampler in Fig. 5b and d. However, as the number of samplers increased (and thus the number of samples) differences became apparent. This reflects the Z-pattern sampling method only collecting samples in the “Z” area. As the number of samples and the number of contamination sites increase, the random and stratified random sampling methods will increasing

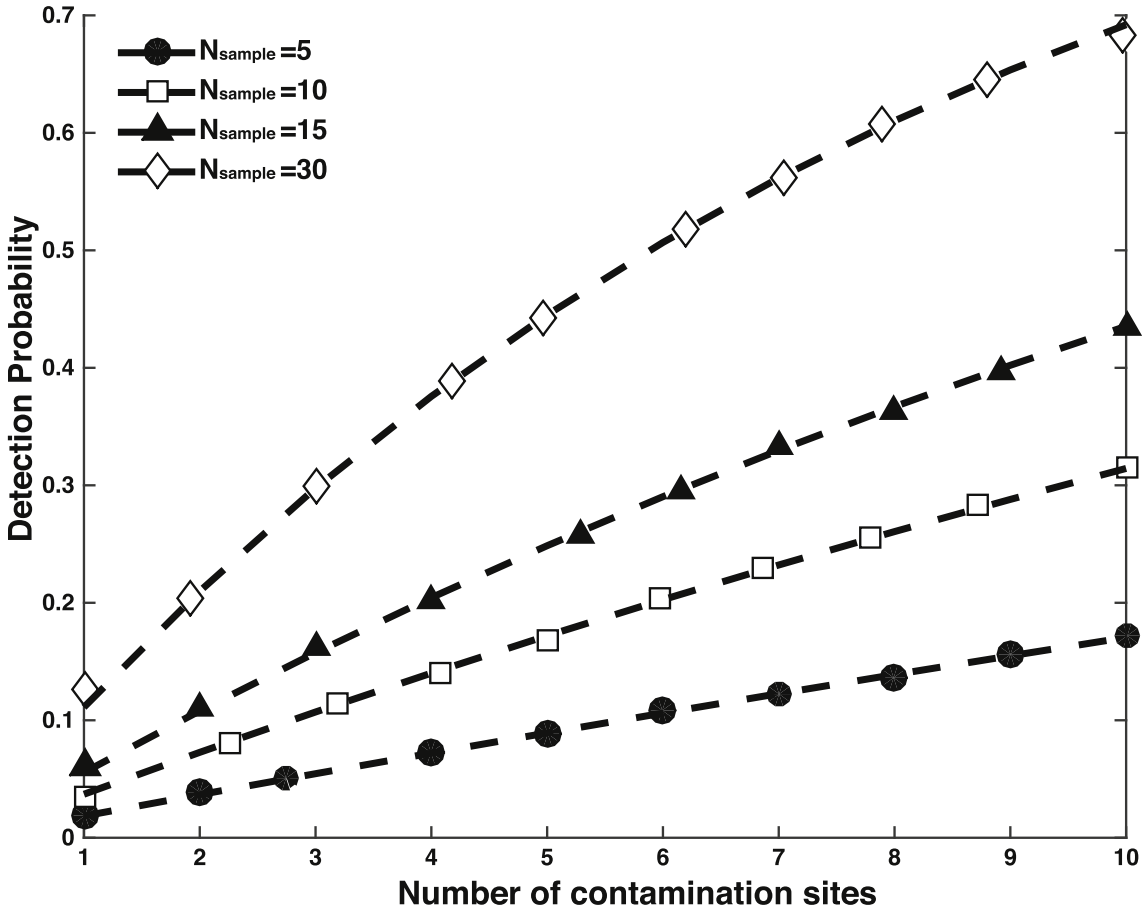

Fig. 2. The theoretical detection probability of detecting at least one positive sample as a function of the number of samples collected and the number of contamination sites (generated with replacement)(see equations (4), (5), and (9)). Detection probability (without replacement) increases as a function of the number of contamination sites (x-axis) and the number of samples (shown as different symbols).

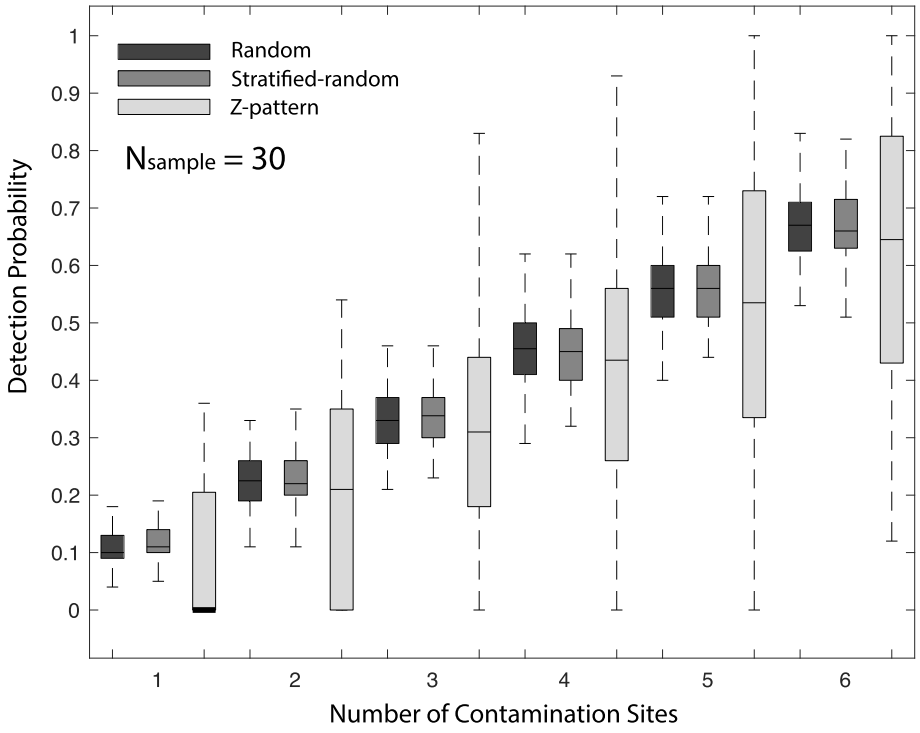

Fig. 3. Detection probability (without replacement) as a function of the number of contamination sites (generated with replacement). Three boxes are the simulation results for the three sampling methods (N_{sample} = 30). Each box plot represents the detection probability from 100 simulations. The simulation generated randomly contaminated fields and estimated the detection probability by repeating the three sampling methods 100 times (100 iterations/simulation). The box plot (a.k.a. box and whisker diagram) is a non-parametric way of displaying the distribution of data based on the five number summaries: minimum, first quartile, median, third quartile, and maximum.

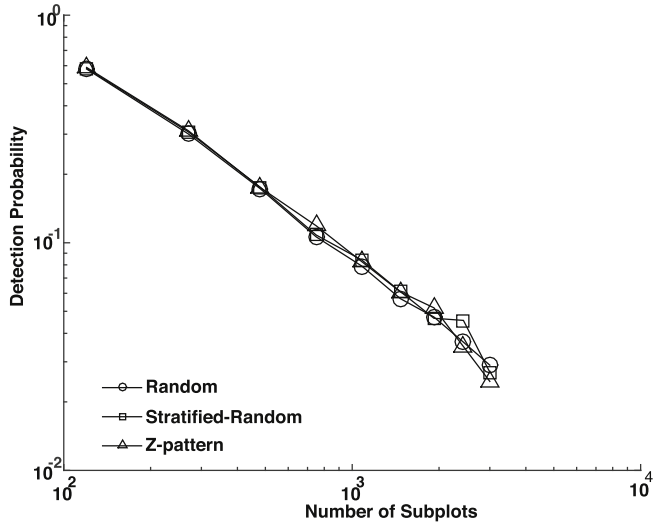

Fig. 4. Log-log plot of mean detection probability as a function of total number of subplots per field for the three sampling methods (N_C = 6, N_{sample} = 30).

identify contamination sites across the entire field. However, the Z-pattern sampling method never samples subplots outside the "Z" area, resulting in an over sampling of the "Z" area when there are large numbers of samples, but not detecting contaminated subplots in the "non-Z" area. However, in terms of practical sampling methods, it is highly unlikely that this differences in mean detection probabilities among the sampling methods would be noted, because is highly unlikely that > 30 pre-harvest samples would be taken from a single field.

3.2. Validation trials

Ten contamination sites were randomly assigned in each of the three fields. Thirty samples were subsequently collected for each of the three sampling methods in each field. The numbers of positive samples for each sampling method are depicted in Fig. 7. The total numbers of positive samples for the three sampling methods were similar when summed over the three replicate experimental fields. The first trial sampled soybeans fields using GPS positioning for locating inoculation and sampling sites. There were a total of 9 positive samples that yielded rifampicin-resistant *E. coli* with random sampling, 7 for stratified random sampling and 9 for Z-pattern sampling. However, there is a considerable amount of variation on the number of positive sample across fields. The Z-pattern sampling plan had the highest variation

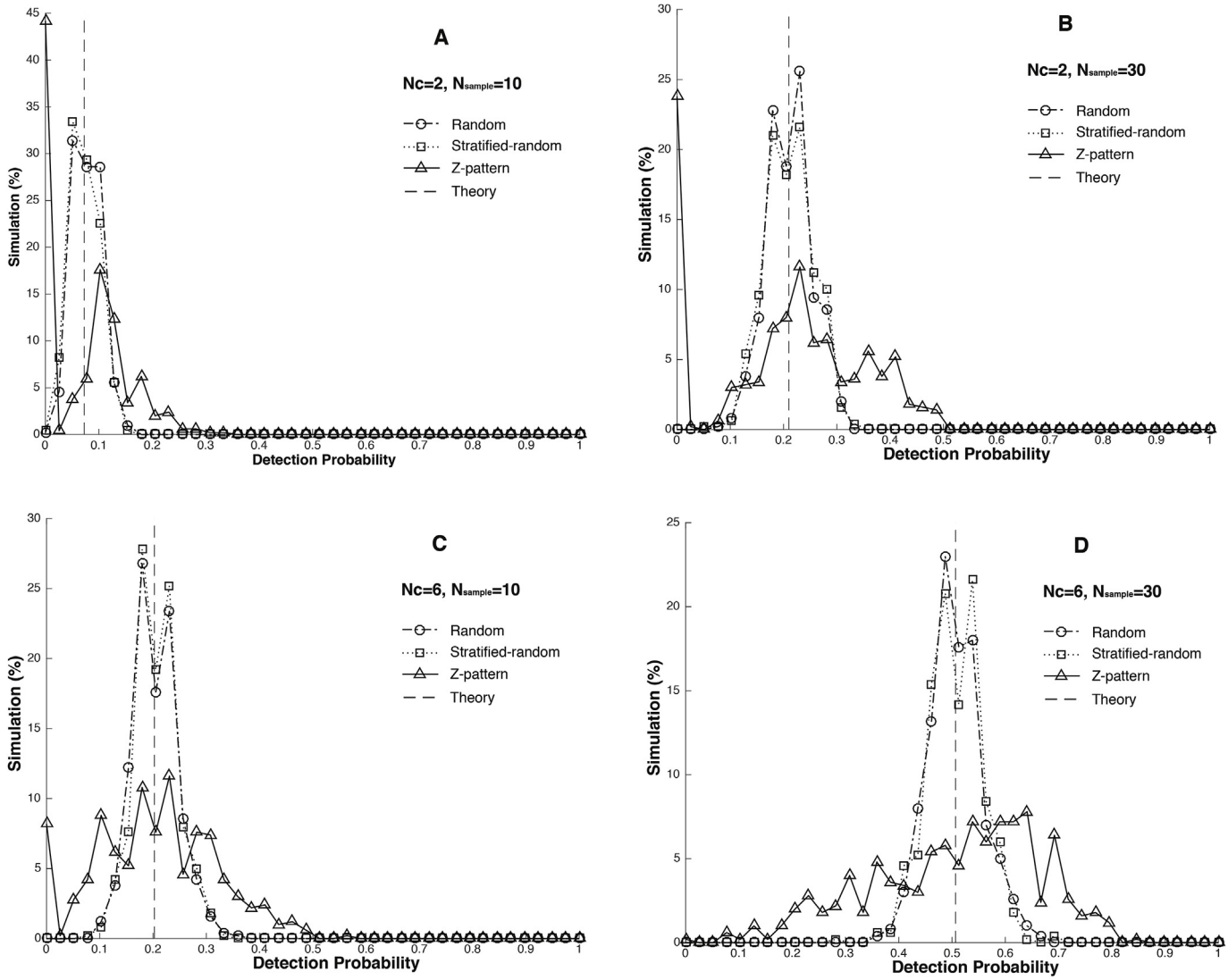

Fig. 5. Distribution of detection probabilities for the three sampling methods for four combinations of contamination sites (N_c = 2 (A & B) and N_c = 6 (C & D) and numbers of samples (N_{sample} = 10 (A & C) or 30 (B & D)). The vertical dashed line represents theoretical value for the mean detection probabilities.

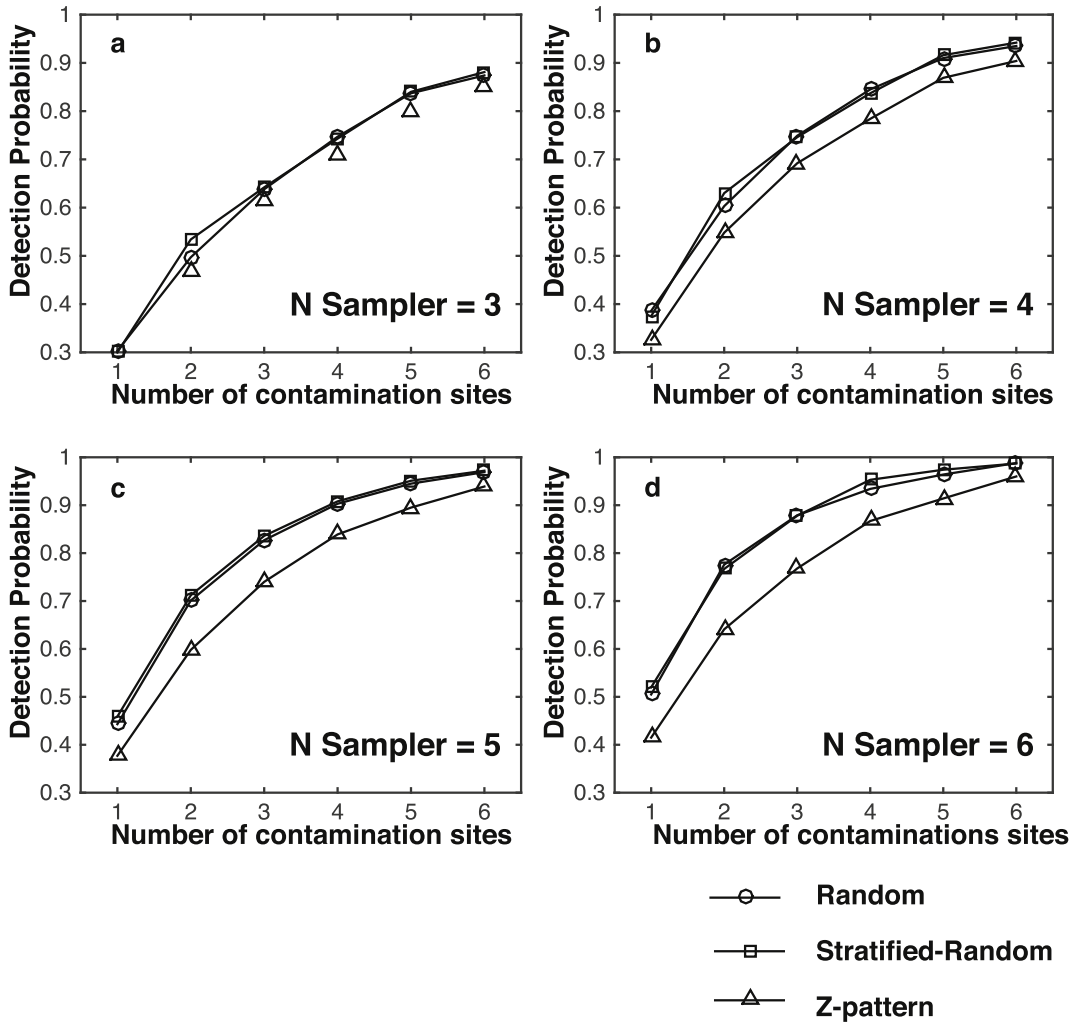

Fig. 6. Detection probability of three sampling methods as a function of the number of contamination sites (x-axis) and number of samplers ($N_{sampler}$). It is assumed that each sampler collects samples independently. The detection probabilities were estimated using one simulation with 100 iterations for each sampler. The results for each sampler were combined and detection probability calculated by the number of iterations where at least one positive sample was detected for any of the samplers.

(standard deviation of number of positive samples = 2.6). Stratified sampling plan had the lowest variation (standard deviation of number of positive samples = 1.2). Random sampling had intermediate variation (standard deviation of number of positive samples = 2.0). When the locations of rifampicin-resistant *E. coli* samples were compared to the original sites of inoculation there was not an exact match, but "near neighbor" analysis indicated that positive samples were isolated from adjacent subplots. Review of the results suggested that this was likely the result of the limited resolution of the GPS meter.

The second validation trial used a second set of three fields planted with lettuce in conjunction with a more traditional "stake and string" grid system. Across the three fields, random sampling identified 4 rifampicin-resistant samples with random sampling, 3 with stratified random sampling, and 4 for Z-pattern sampling. In this instance the positive samples precisely corresponded to inoculated subplots. Again there was considerable variation in the number of positive samples across fields. The Z-pattern sampling had highest variation (standard deviation of the number of positive samples = 1.5). Random sampling had lowest variation (standard deviation of the number of positive samples = 0.58). Stratified sampling had intermediate variation (standard deviation of the number of positive samples = 1.0). The Z-pattern sampling had the highest variability and thus the highest level of uncertainty in both trials.

4. Discussion

The goal of this study was to compare the effectiveness of three widely used sampling methods for the detection of bacteria in preharvest leafy green fields. The results help quantify the impact that the number of contamination sites and the number of samples analyzed have on the likelihood of positively detecting a targeted pathogen or indicator microorganism. Further, the results demonstrate that basing decisions related to choice of sampling method cannot rely solely on mean detection probabilities but also must consider the uncertainty associated with those values. The use of relatively simple simulation models proved to be highly effective for estimating that uncertainty.

Quantifying the variability (and thus uncertainty) of detection probabilities provides a more objective means of estimating the relative effectiveness of sampling methods and clearly points out the limitations of sampling when small numbers of samples are used to evaluate fields that are sparsely contaminated. Small numbers of contamination sites in combination with small number of samples ($N \leq 5$) typically resulted in detection probabilities below 0.2 (Fig. 2). Similarly, the likelihood of detecting a target bacterium is very small when the sample size/area becomes increasingly small (Fig. 4). Simulation modeling provided new insights into the characterization of the performance of different sampling methods by allowing the variability of the sampling methods to be quantified along with the mean detection probability. This is clearly

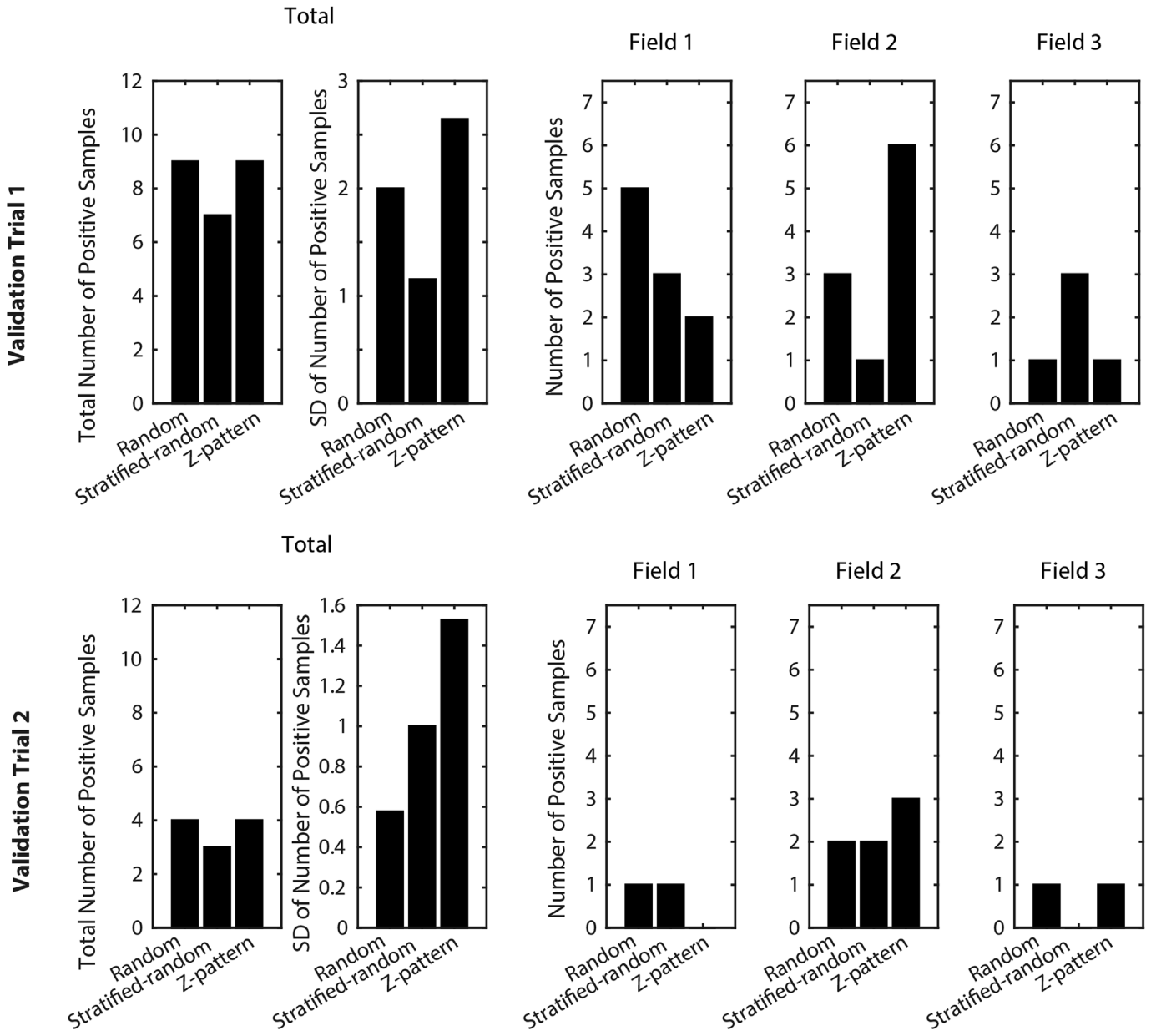

Fig. 7. Number of positive rifampicin-resistant *E. coli* samples and standard deviation of positive samples for different sampling methods in two field trials. Each of the three fields had 10 contamination sites and 30 sampling sites for potential combined total of 30 positive samples.

demonstrated that while Z-pattern sampling yielded the same mean detection probabilities as the other sampling methods, the variability indicated a systematic over- and under-estimation of the degree of contamination based on whether the contamination sites were within the sampled plots, thereby leading to more erratic results.

Z-pattern sampling has been found to be efficient in field sampling for insect control (Zehnder, 2014) and soil composition analyses (Paetz and Wilke, 2005). The obvious advantage of Z-pattern sampling is the increased ease of collection of samples. The total distance travelled with a Z-pattern sampling strategy would be typically shorter than either random or stratified random sampling. This is undoubtedly one of the reasons for the wide spread adoption of Z-pattern sampling for pre-harvest assessment of leafy greens for microbiological contaminants. The current study clearly indicates that the mean detection probabilities of the three sampling methods are similar if a sufficiently large number of samples are evaluated (Figs. 2–4). However, the current study also clearly establishes that the convenience of Z-pattern testing must be balanced against its increased variability, particularly if relying on a small number of samples when the numbers of contamination sites are likely to be low. The consistency of random or stratified random sampling was superior to Z-pattern sampling.

The limited validation trials were consistent with the results achieved by simulation modeling. However, strict validation would require a substantially greater number of field trials of commercial leafy greens cultivation sites since the frequency of contamination is low and the number of samples taken are typically few. Future validation trials should also consider potential differences in the types of leafy greens. This might explain the approximate doubling of positive rifampicin-resistant *E. coli* detections with soybeans versus lettuce. Since soybean is much taller than lettuce, it was not easy to inoculate exactly within the selected area when spraying. However, the differences in the two validation trials are also attributable to the uncertainty associated with the uncertainty associated with the GPS locations in the first validation trial.

This study assumed that the results would be positive if the samples were collected in the contaminated subplot. However, in reality, the

results could be negative even though they were collected in a contaminated subplot. In Fig. 7, we can see that the average number of positive samples with three sampling methods ranged from 1 to 3. The P_d of positives theoretically is around 0.7 (Fig. 2). The expected number of positives (0.7*10 = 7) is higher than the observed number of positives. Many factors could lead to this, such as limited experiment times, false negatives, etc. This could also result from either the extent of contamination being low (e.g., < 10 CFU/cm^2) or the actual contaminated area within the subplot being small. Since we only collected leaf samples in part of the subplot, it is possible to get false negative results.

The current study is based on the assumption that contamination is the result of a random event(s). However, there are a number of scenarios where non-random events could be the root cause of leafy green field contamination. Studying the contamination of powdered infant formula, Jongenburger et al. (2011) reported that stratified sampling is preferred over random sampling for detection of localized contamination. Similar results have been reported by other investigators (Habraken et al., 1986; Rivas Casado et al., 2009). However, a potential drawback of stratified sampling is the potential to miss a systematic, reoccurring contamination event (Jongenburger et al., 2015). In such instances sampling performance can be enhanced by moving to stratified random sampling. It is also worth noting the performance of random and stratified random sampling are similar when the number of samples ≥30 (Jongenburger et al., 2011, 2015).

The current evaluation was based on the commonly used underlying assumption that the sampler has no *a priori* knowledge of potential sources of contamination. However, knowledge of the leafy green production environment and likely sources of contamination could potentially increase the effectiveness of testing regimes (Zagory, 2014). It is feasible that assessment of a leafy green field by a sampler could take advantage of expert knowledge of potential contamination sources to improve the statistical basis for pre-harvest testing. For example, the presence of overhead wires, adjacent animal facilities, areas of periodic flooding, and prevailing winds are all factors that can lead to non-homogeneous contamination of leafy green fields. This suggests that pre-harvest sampling could be enhanced by more systematically and quantitatively studying sources of contamination of pre-harvest produce to focus sampling on the conditions that are likely to increase contamination risks. This would allow use of hybrid sampling methods which combine systematic sampling with "samples of opportunity." Such sampling could involve collecting a portion of the samples via a standardized sampling method, while a second portion of the samples would increase the extent of sampling in areas which the sample collector observed factors that are associated with an increased contamination. A study comparing the relative effectiveness of sampling methods that combine a statistically based sampling with such "samples of opportunity" is currently underway.

5. Conclusions

This study's use of simulation modeling techniques to evaluate the performance of sampling methods for leafy greens demonstrated that the detection probabilities for random and stratified sampling plan are more stable than those for Z-pattern sampling. Similarly random and stratified random sampling performed more consistently than Z-pattern sampling when fields are sampled multiple times. In blinded field verification experiments, Z-pattern sampling had higher variability than other two samplings, which is consistent to the model results. The study demonstrates the utility of simulation modeling for evaluating the performance of different sampling methods used in pre-harvest testing.

Acknowledgments

We sincerely thank Dr. Donald Schaffner, Dr. Abani Pradhan, and Dr. Abhinav Mishra for their helpful suggestions. This work was supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture (grant number 2011-51181-30767). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the USDA National Institute of Food and Agriculture.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.fm.2018.09.007.

References

European Commission, 2007. Commission regulation (EC) No 1441/2007 on microbiological criteria for foodstuffs. J. Eur. Union 1441/2007, L322/312-L322/328.

Grinstead, C.M., Snell, J.L., 2003. Grinstead and Snell's Introduction to Probability, second ed. American Mathematical Society, Providence, RI.

Habraken, C., Mossel, D., van der Reek, S., 1986. Management of *Salmonella* risks in the production of powdered milk products. Neth. Milk Dairy J. 40, 99–116.

Jongenburger, I., Reij, M.W., Boer, E.P.J., Gorris, L.G.M., Zwietering, M.H., 2011. Random or systematic sampling to detect a localised microbial contamination within a batch of food. Food Contr. 22, 1448–1455.

Jongenburger, I., den Besten, H.M.W., Zwietering, M.H., 2015. Statistical aspects of food safety sampling. Annu. Rev. Food Sci. Technol. 6, 479–503.

Macarisin, D., Patel, J., Bauchan, G., Giron, J.A., Ravishankar, R., 2013. Effect of spinach cultivar and bacterial adherence factors on survival of *Escherichia coli* O157:H7 on spinach leaves. J. Food Protect. 76, 1829–1837.

Paetz, A., Wilke, B.-M., 2005. Monitoring and Assessing Soil Bioremediation, Soil Biology. Springer-Verlag, Berlin/Heidelberg.

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 stage: effects of farm management and environmental factors. Appl. Environ. Microbiol. 79, 4347–4358.

Rivas Casado, M., Parsons, D.J., Weightman, R.M., Magan, N., Origgi, S., 2009. Modelling a two-dimensional spatial distribution of mycotoxin concentration in bulk commodities to design effective and efficient sample selection strategies. Food Addit. Contam. 26, 1298–1305.

Zagory, D., 2014. Microbial Testing of Fresh Produce: where Is the Value? | Food Safety News. [WWW Document]. URL. http://www.foodsafetynews.com/2014/10/microbial-testing-of-fresh-produce-where-is-the-value/#.VsEM5zYrI0o, Accessed date: 14 February 2016.

Zehnder, G., 2014. Overview of Monitoring and Identification Techniques for Insect Pests - EXtension. [WWW Document]. Extension. URL. http://www.extension.org/pages/19198/overview-of-monitoring-and-identification-techniques-for-insect-pests#.ViqhaBNViko, Accessed date: 23 October 2015.

Zwietering, M.H., Ross, T., Gorris, L.G.M., 2014. Food Safety Management. Food Safety Assurance Systems: Microbiological Testing, Sampling Plans, and Microbiological Criteria. Encyclopedia of Food Safety, vol. 4. Academic Press, MA , pp. 244–253.