Evaluation of the Impact of Skewness, Clustering, and Probe Sampling Plan on Aflatoxin Detection in Corn

Xianbin Cheng and Matthew J. Stasiewicz

Probe sampling plans for aflatoxin in corn attempt to reliably estimate concentrations in bulk corn given complications like skewed contamination distribution and hotspots. To evaluate and improve sampling plans, three sampling strategies (simple random sampling, stratified random sampling, systematic sampling with U.S. GIPSA sampling schemes), three numbers of probes (5, 10, 100, the last a proxy for autosampling), four clustering levels (1, 10, 100, 1,000 kernels/cluster source), and six aflatoxin concentrations (5, 10, 20, 40, 80, 100 ppb) were assessed by Monte-Carlo simulation. Aflatoxin distribution was approximated by PERT and Gamma distributions of experimental aflatoxin data for uncontaminated and naturally contaminated single kernels. The model was validated against published data repeatedly sampling 18 grain lots contaminated with 5.8–680 ppb aflatoxin. All empirical acceptance probabilities fell within the range of simulated acceptance probabilities. Sensitivity analysis with partial rank correlation coefficients found acceptance probability more sensitive to aflatoxin concentration (−0.87) and clustering level (0.28) than number of probes (−0.09) and sampling strategy (0.04). Comparison of operating characteristic curves indicate all sampling strategies have similar average performance at the 20 ppb threshold (0.8–3.5% absolute marginal change), but systematic sampling has larger variability at clustering levels above 100. Taking extra probes improves detection (1.8% increase in absolute marginal change) when aflatoxin is spatially clustered at 1,000 kernels/cluster, but not when contaminated grains are homogeneously distributed. Therefore, taking many small samples, for example, autosampling, may increase sampling plan reliability. The simulation is provided as an R Shiny web app for stakeholder use evaluating grain sampling plans.

KEY WORDS: Aflatoxin; clustering; probe sampling; sampling strategy; skewness

1. INTRODUCTION

Aflatoxin is a secondary metabolite produced mostly by toxigenic strains of Aspergillus flavus and Aspergillus parasiticus, which can infect preharvest grains under environmental stress for example, drought and insect infestation, or grains in storage facilities under suboptimal storage conditions, for example high temperature and moisture (Gnonlonfin et al., 2013). Consuming grains contaminated with a large quantity of aflatoxin may cause acute aflatoxicosis, a condition characterized by severe liver damage and possibly death. Chronic exposure to aflatoxin is associated with stunted growth in children and may result in liver cancer (IARC, 2018; Wu, Groopman, & Pestka, 2014). This work will focus on aflatoxin in corn as a test case. To ensure the safety of the corn, an action level of 20 ppb for aflatoxin has been set by the U.S. Food and Drug Administration (FDA, 2000). Any lot with ≥20 ppb of aflatoxin in the United States will be deemed unfit for human consumption or many types of feeds, and thus rejected or sent to uses that tolerate higher levels of aflatoxin, such as specific animal feeds like finishing
cattle (FDA, 2000). As a side note, this article uses the unit “ppb” instead of “μg/kg” or “ng/g” to be consistent with the FDA regulations and the industry convention.

Determining the aflatoxin level of grains in a lot requires bulk sampling. It is a process roughly summarized as taking representative samples from the grains in a lot, measuring the aflatoxin concentration in these portions, and making inference on the aflatoxin level in the entire batch of grains. Among several methods for obtaining representative samples, probe sampling is one of the most common methods adopted by the grain industry (USDA, 2020). A probe sample is taken by inserting a hollow probe into a grain container and entrapping grains inside the probe. Multiple probe samples of a prespecified weight will be pooled together, ground, and further divided into subportions (USDA, 2015). Eventually, a final subportion is assayed for aflatoxin, whose concentration is used to estimate the aflatoxin level for the entire lot of grains. Since only a few portions of the grains are used for estimating the overall contamination level, it is crucial that these portions are unbiased and can faithfully represent the lot. Efforts of ensuring unbiased sampling can be seen in the sampling guidance set by the Grain Inspection, Packers and Stockyards Administration (GIPSA), which designate a certain number of probing spots at regular intervals in grain containers such as trucks, barges, hopper cars, and so on (USDA, 2020). This sampling strategy is considered a systematic sampling (SS) method. Additionally, other sampling strategies have been proposed by the academics as conceptually feasible methods, including simple random sampling (SRS) and stratified random sampling (STRS) (Bouzembrak & van der Fels-Klerx, 2018; Jongenburger, de Besten, & Zwietering, 2015).

While randomized sampling strategies have been utilized extensively, research about their effects on representative sampling has been mostly restricted to theoretical derivation (Bouzembrak & van der Fels-Klerx, 2018; Jongenburger et al., 2015) without taking into consideration real-life factors such as skewed distribution of aflatoxin concentration and spatial clustering. Aflatoxin concentration distribution has been reported to be skewed in both bulk grains and single kernels (Chavez, Cheng, & Stasiewicz, 2020; Stasiewicz et al., 2017; Whitaker, Dickens, & Wiser, 1976), that is, a small fraction of the grain kernels are highly contaminated whereas the majority proportion remains uncontaminated or contaminated at a negligible level. This condition is known to cause large variances of the aflatoxin level in the test sample (Johansson et al., 2000) and can increase the risk of accepting a defective lot of grains, possibly resulting in foodborne diseases. Aside from skewness, spatial clustering of contaminated grains may also undermine the effort to sample representatively. Clustering refers to the phenomenon of spatial autocorrelation where kernels in close proximity tend to have similar contamination levels, which has been reported in literature (Krska & Molinelli, 2007; Oerke et al., 2010; Shotwell, Goulden, Bothast, & Hesselting, 1975). Several studies have attempted to address how clustering affects food product sampling. Casado, Parsons, Weightman, Magan, and Origgi, 2009a utilized geostatistical analysis to examine the spatial distribution of two mycotoxins (deoxynivalenol and ochratoxin A) in wheat and reported that deoxynivalenol was heterogeneously distributed. Their subsequent research (Casado, Parsons, Weightman, Magan, & Origgi, 2009b) compared the performance of random sampling and regular grid sampling and concluded that regular grid sampling provided more accurate estimation of deoxynivalenol concentration. Jongenburger, Reij, Boer, Gorris, and Zwietering, 2011b observed the clustered distribution of Cronobacter spp. in powdered infant formula and concluded that STRS would outperform SRS in facilitating the detection of localized contamination. While these studies have explored the relationship between clustering and sampling strategies, their systems were somewhat restricted in that clustering was either observed from individual bags of infant powdered formula along a production line or simulated as correlated numbers in a grid. Their findings may not be directly applicable to probe sampling in grains where clustering may occur among individual kernels that can exist in any part of a grain container.

Given the demand for more effective sampling plans to reduce food safety risk associated with aflatoxin-contaminated grains, this study has been proposed to evaluate the effect of sampling strategies, skewed distribution of aflatoxin, and spatial clustering on sampling effectiveness in a grain container. To mimic the sampling process in a grain container, we constructed a stochastic model in R which was utilized in conjunction with the Monte Carlo technique to examine the probability of acceptance at different combinations of sampling strategies (SRS, STRS, SS), number of probes (5, 10, 100), and clustering level (1, 10, 100, 1,000) in a truck of $3 \times 2 \times 1.5$ m$^3$. Finally, the impact of these input
Simulating Aflatoxin Sampling

Fig 1. The schematic diagram of the simulation process: (1) simulating contaminated corn kernels in a container; (2) taking probe samples by a specific sampling strategy; (3) assaying the samples; and (4) accepting or rejecting the container. The inner iteration layer produces an acceptance probability for one grain container and the outer iteration layer produces multiple acceptance probabilities, each corresponding to a different container.

parameters on the probability of acceptance was quantified by a sensitivity analysis. To allow for user interaction, a web app for the model was created in R Shiny and is publicly available.

2. MATERIALS AND METHODS

2.1. Simulation Model Framework

2.1.1. Overview

A simulation model framework was constructed in R and was designed to take user’s inputs that defined a specific sampling scenario and returned the probability of acceptance. A graphic user-interface was developed using R Shiny and can be accessed via https://go.illinois.edu/FoodSafetySampling. Mathematical derivation details beyond those in this main text are presented in Supporting Information Appendix A.

The simulation model framework is comprised of the following four modules: hazard, sampling, assay, and decision making (Fig. 1). The hazard module creates a virtual grain container with a user-defined hazard profile, such as the container size, the overall aflatoxin concentration in the grain container, aflatoxin clustering level, and so on. Upon creation, this virtual grain container is passed as an input to the sampling module, where certain number of simulated probe samples are taken. In the assay module, these simulated probe samples are subjected to grinding, subsampling, and assaying. Based on the assay result and the FDA action level, the whole lot of grains will either be accepted or rejected.

A set of input parameters were created to delineate sampling scenarios in detail (Table I). While these input parameters were built to accept a wide range of values, some parameters had built-in default values that were chosen based on literature, empirical data, or arbitrary assumptions. The following sections will elaborate on essential parameters and rationale for default values.

2.1.2. Hazard Module

Hazard module is intended to populate a 3D space with points that represent contaminated grains. Parameters in this module control the concentration and spatial distribution of these “contaminated grains.” This article uses corn sampling as an example.

To simulate the skewness in concentration, the aflatoxin level for uncontaminated kernels (<20 ppb) and contaminated kernels (≥20 ppb) were assumed to follow different distributions. Based on our previous study (Cheng, Chavez, & Stasiewicz, 2020) where data of 432 individually-collected aflatoxin-contaminated kernels were used for distribution fitting, aflatoxin concentration in healthy kernels could be approximated by a modified PERT distribution (\( \text{min} = 0, \text{mode} = 0.7, \text{max} = 19.99, \text{shape} = 80 \)) and concentration in contaminated kernels could be fitted by \( 20 + \Gamma(\alpha = 2, \theta = 39, 980) \). Due to limited computation capacity, the aflatoxin concentration in contaminated kernels was simplified to a constant number of 40,000 ppb, which matched the most probable number from the aforementioned distribution. As a result, the final aflatoxin distribution would be a mixture of both distributions with the weights proportional to the number of contaminated and healthy kernels. The validation of simulated aflatoxin distribution was described in detail in (Cheng et al., 2020).

Clustering was depicted as corn kernels spreading around a contamination source. Their coordinates were random vectors drawn from a
**Table I. Definitions of the Input Parameters in the Simulation Model**

| Variable         | Description                                                                 | Value Range/Distribution          | Default Value | Unit | Source                |
|------------------|------------------------------------------------------------------------------|-----------------------------------|---------------|------|-----------------------|
| $x_{lim}, y_{lim}, z_{lim}$ | Dimensions of the grain container (length, width, height) | $>0$                              | 1, 1, 1       | m    | User input            |
| $c_{hat}$        | Estimated aflatoxin concentration in the container                           | $>0$                              | 3             | ppb  | User input            |
| dis_level        | Aflatoxin level distribution in the contaminated kernels                     | Constant or modified Gamma        | 40000         | ppb  | Experimental data     |
| $n_{affected}$   | Number of kernels clustering on a contamination source                        | $\geq 0$                          |               |      | Assumed               |
| covar_mat        | Covariance matrix that describes clustering                                  | positive semidefinite matrix       |               |      | Assumed               |
| method_sp        | Sampling strategy                                                            | SRS, STRS, SS                     | SRS           |      | User input            |
| $n_{sp}$         | Number of probes (for SRS and STRS)                                          | $>0$                              | 5             |      | User input            |
| $n_{strata}$     | Number of strata (for STRS)                                                  | $>0$                              | NA            |      | User input            |
| by               | Stratification direction (for STRS)                                          | row, column, 2d                    | NA            |      | User input            |
| $d$              | Inner diameter of the probe                                                  | $>0$                              | 0.04          | m    | Commercial            |
| sp_radius        | Sampler radius                                                                | $=d/2$                            | 0.02          | m    | Derived               |
| $L$              | Length of the probe                                                          | $0 < L \leq z_{lim}$              | $z_{lim}$     | m    | Assumed               |
| rho              | Density of corn kernels                                                       | $>0$                              | 1.28          | g/cm$^3$ | Literature            |
| m_kbar           | Average mass of a single kernel                                              | $>0$                              | 0.3           | g    | Experimental data     |
| homogeneity      | Homogeneity of ground kernels                                                 | 0–1                               | 0.1           |      | Assumed               |
| method_det       | Detection method                                                              | ELISA                             | ELISA         |      | User input            |
| Mc               | Maximum permitted level of aflatoxins                                         | $>0$                              | 20            | ppb  | FDA                   |
| conc_neg         | Aflatoxin concentrations from uncontaminated kernels                          | Modified PERT                     | PERT (min = 0, mode = 0.7, max = 19.99, shape = 80) |      | Experimental data     |

SRS = simple random sampling, STRS = stratified random sampling, SS = systematic sampling, ELISA = Enzyme-linked Immunosorbent Assay, NA = not applicable.
multivariate normal distribution \( N(\mu, \Sigma) \), where \( \mu \) = coordinates of the contamination source, \( \Sigma \) is variance–covariance matrix that described the dispersity. We selected \( \Sigma = \text{diag}(0.0004, 0.0004, 0.0004) \) as default, assuming the \( x \)-, \( y \)-, and \( z \)-coordinates were independent, and each dimension had a standard deviation of 2 cm.

Our model allows the end user to define the estimated overall aflatoxin concentration and number of kernels surrounding each contamination source. With these two parameters and the dimensions of the container, algorithms (Supporting Information Appendix A.1. and A.2.) are utilized to determine the number of contaminated kernels and their locations, which will be used in the subsequent modules.

2.1.3. Sampling Module

Sampling module controls the quantity and locations of probe samples. While the number of probe samples is mainly chosen by the end user, their locations are determined by the sampling strategy. Three sampling strategies are included in this model: simple random sampling (SRS), stratified random sampling (STRS), and systematic sampling (SS). SRS is defined as taking random probe samples in a grain container. STRS is defined as dividing the grain container into several strata and taking same number of random samples in each stratum. SS is defined as taking samples at locations specified by the GIPSA protocol (USDA, 2020). For simplicity, all probes were assumed to be long enough to reach the bottom of the container. In other words, the sampling location varies in the \( x \)- and \( y \)-dimension, covering the entire \( z \)-dimension space.

2.1.4. Sample Preparation and Assay Module

With coordinates and concentrations from the hazard and the sampling module, sample preparation and assay module simulates a series of procedures including compositing probe samples, grinding, subsampling, and concentration measuring.

To simulate compositing probe samples, we first defined how a kernel was captured. Any contaminated kernel would be deemed captured by a probe when it falls within the radius of the probe, or more precisely, when

\[
\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \leq r,
\]

where \((x_1, y_1)\) = coordinates of the contaminated kernel on the \(X-Y\) plane, \((x_0, y_0)\) = coordinates of the probe central axis, \(r = \) probe radius (m). We assumed the radius was 2 cm.

After the kernels are captured, they are subjected to comminution, which is simulated mathematically using the concept of moving averages. More specifically, the captured kernels’ concentration values are arranged in a vector, and a moving window with a certain width will slide through the vector, calculating the mean concentration in that window. The window width controls the smoothness of the aflatoxin concentration values and represents the homogeneity of ground kernels (see homogeneity in Table 1). A wider window will force more numbers to regress toward the average, making the values smoother. In this model, we assumed the window width to be 10% of the vector length as it created a smoothing effect that was most satisfying among all the widths examined in a small-scale trial.

Subsampling is simulated based on the GIPSA protocol (USDA, 2015). It specifies that the ground sample should be divided into a 500 g work sample, which is then subdivided into a 50 g test sample. This process is implemented by taking a random subset from the parent vector. The size of the subset should be determined by

\[
\frac{m_{\text{sample}}}{\bar{m}_k},
\]

where \(m_{\text{sample}}\) is sample mass (500 g or 50 g, depending on the type of sample), \(\bar{m}_k\) is average mass of grain kernels. Based on our empirical data, \(\bar{m}_k\) was set at 0.3 g.

Finally, measuring the concentration of aflatoxin is simply simulated by calculating the mean of the concentration values in the test sample vector. For this study, the limit of detection (LOD) was set at 1 ppb to align with the enzyme-linked immunosorbent assay (ELISA) kit used to generate the single kernel data adopted to validate the simulated aflatoxin concentration in (Cheng et al., 2020). It is appropriate to also simulate ELISA method testing for bulk levels because many specific ELISA kits and testing methods are approved by the Federal Grain Inspection Service (FGIS) (USDA, 2018) for bulk toxin testing. In fact, this model is indifferent to the type of assay utilized for measurement if the assay LOD is lower than the threshold concentration for rejection. Thus, any assay method with the same LOD would result in a similar simulation. Assay performance metrics are not considered because they vary from assay to assay.
Simulating, instead, a perfect assay makes this study more generalizable.

2.1.5. Decision Module

Decision module determines whether the entire grain container should be accepted or rejected. If the aflatoxin concentration in the test sample is $\geq 20$ ppb, which is the aflatoxin action level (denoted as $M_c$) of food products intended for human consumption (except for infants) (FDA, 2000), the whole container is deemed contaminated and thus rejected.

2.1.6. Iteration

To calculate the probability of acceptance, this simulation model framework is nested within a two-layer iteration framework. In the inner layer, the hazard locations are fixed while the sampling locations are randomized, allowing for the calculation of acceptance probability, which is defined by

$$P_a = \sum_{i=1}^{n_1} I_a,$$

where $P_a$ is probability of acceptance, $I_a$ is binary indicator of acceptance with 0 denoting rejection and 1 denoting acceptance, $n_1$ number of iterations in the inner layer. This represents the probability of accepting a grain container after sampling it for $n_1$ times.

In the outer layer, the hazard locations are also randomized, allowing for the calculation of acceptance probabilities in different grain containers, which are mathematically represented as a set of $P_a$'s:

$$\left\{ P_{a_1}, P_{a_2}, \ldots, P_{a_{n_2}} \right\},$$

where $P_{a_j}$, the acceptance probability of sampling the $j^{th}$ container, $n_2$ number of different grain containers. This layer represents the action of examining $n_2$ grain containers with each container sampled for $n_1$ times. These $n_2$ containers have the same overall aflatoxin concentration and clustering level, but the locations of hotspots vary.

2.2. Model Validation

To prove the simulation model could accurately predict the probability of acceptance, model validation was conducted using empirical data reported in (Johansson, Whitaker, Giesbrecht, Hagler, & Young, 2000a). The general validation procedure started with extracting input values from the literature study and mapping them to our simulation’s input parameters. Once the simulation was complete, an operating characteristic (OC) curve was plotted and compared against the empirical acceptance probabilities reported by the literature study. Our simulation model is only deemed valid when the empirical probabilities fall within the range spanning from the minimum to the maximum of simulated probabilities. This validation method was adapted from the boxplot method described in (Sargent, 2010).

With the attempt to faithfully replicate the sampling scenario of the literature study, we simulated a cubic grain container containing 100 pounds of corn, where 32 random samples (2.5 pounds each) were ground in a Romer mill and subdivided into a total of 64 subsamples (50 g each). Any subsample with $\geq 20$ ppb of aflatoxin was considered positive. Due to lack of information about clustering and ground corn homogeneity, these parameters were given default values shown in Table I. To generate the OC curve, we virtually sampled 18 lots with aflatoxin concentration varying from 5.8 to 676.6 ppb, which were concentrations reported from (Johansson et al., 2000). Each sampling scenario was iterated for 10,000 times.

2.3. Simulated Experiments

Once the simulation model was validated against the literature, simulated experiments were designed and conducted in an attempt to elucidate the impact of various factors on the probability of acceptance, including the number of probes, the chosen sampling strategy, and the clustering level (Table II).

Specifically, two simulated experiments were conducted to evaluate sampling in a $3 \times 2 \times 1.5$ m$^3$ truck under four levels of clustering (1, 10, 100, 1,000 kernels clustering on a contamination source). For easier interpretation, these clustering levels were converted to percentage of clusters (50%, 9%, 1%, 0.03% respectively) based on the equation (10) in the Supporting Information Appendix A.2. For example, a percentage of clusters at 50% means half of the contaminated kernels are contamination sources and the remaining kernels are equally distributed to each source. A larger percentage of clusters indicates more contamination sources, and thus lower clustering level. Six aflatoxin levels (5, 10, 20, 40, 80, 100 ppb) were examined in each experiment to construct the OC curves. It was assumed in this experiment that each truck was fully filled with grains, that is, the grain depth equaled the height of the truck.
### Table II. Chosen Input Values for the Two Simulation Experiments

| No. | Aflatoxin Level ($c_{hat}$) (ppb) | Number of Probes ($n_{sp}$) | Sampling Strategy ($method_{sp}$) | Clustering Level ($n_{affected}$) |
|-----|----------------------------------|----------------------------|----------------------------------|----------------------------------|
| 1   | 5, 10, 20, 40, 80, 100           | 10,000                     | 1, 10, 100, 100                  | 1, 10, 100, 100                  |
| 2   | 5, 10, 100, 100                  | 10,000                     | 1, 10, 100, 100                  | 1, 10, 100, 100                  |

*These numbers of probes are determined by the GIPSA protocol, which requires seven probes for trucks with grain >4 ft (1.21 m) deep. It is assumed in this experiment that each truck is fully filled with grains, or the grain depth equals the height of the truck.*

$P_{a_{ij}}$ is the probability of acceptance at $i$ ppb for the $j$th container, $n$ is the total number of containers $n_2$.

Experiment 1 focused on the effect of having different sampling strategies (SRS with seven probes, STRS with seven strata and one probe/stratum, GIPSA SS for trucks). Since the GIPSA SS protocol specifies that seven probes should be taken from trucks with grain >4 ft (1.21 m) deep (USDA, 2020), we also set the number of probes to seven for both SRS and STRS for consistency. Meanwhile, experiment 2 was designed to investigate the effect of different number of probes (5, 10, 100) while fixing the sampling strategy to SRS. Taking five or 10 probes represented a sampling event with manual probing whereas taking 100 probes was intended to represent a sampling event with an autosampler. Other input parameters were given default values listed in Table I. As a side note, the total mass of composite probe samples before grinding in experiment 2 was not equal for different number of probes. This is consistent with the GIPSA protocol for manual probing where there is no upper limit for the total mass of composite probe samples; extra probes are allowed to ensure sample representativeness (USDA, 2020). In summary, these two experiments assessed 72 sampling scenarios respectively with each scenario iterated for 10,000 times ($n_1 = n_2 = 100$).

As part of the exploratory analyses, sampling was also simulated in a $1 \times 1 \times 1$ m$^3$ tote and a $14 \times 3 \times 2.5$ m$^3$ hopper car with inputs equivalent to those in experiment 1 and 2. These results were included in the Supporting Information Appendix B (Fig. S1 and Fig. S2).

### 2.4. Data Analysis

#### 2.4.1. OC Curve Slope Estimation

A slope estimation algorithm was developed and implemented to quantitatively compare the OC curves summarized from the simulated experiments. The algorithm was designed to estimate the OC curve slope at 20 ppb aflatoxin by

$$
\hat{k}_{20} = \frac{1}{2n} \times \left( \sum_{j=1}^{n} \frac{P_{a_{10j}} - P_{a_{20j}}}{10 - 20} + \sum_{j=1}^{n} \frac{P_{a_{20j}} - P_{a_{40j}}}{20 - 40} \right),
$$

where $\hat{k}_{20}$ is an estimated slope at 20 ppb, $P_{a_{ij}}$ is the probability of acceptance at $i$ ppb for the $j$th container, $n$ is the total number of containers $n_2$.

The intuition behind the algorithm is based on the fact that an ideal OC curve should be a step function where the acceptance probability maintains at 100% when aflatoxin concentration is <20 ppb and
abruptly decreases to 0% when the concentration is \( \geq 20 \) ppb (Johansson, Whitaker, Giesbrecht, Hagler, & Young, 2000b). However, OC curves in reality are usually continuous curves that decrease gradually as the concentration increases. A sampling plan can be considered more effective if the OC curve is more analogous in shape to a step function. In other words, a preferable OC curve should be steep at the threshold concentration (20 ppb), which can be quantitated by the slope. As an OC curve was approximated by line segments of discrete acceptance probabilities, we could estimate the slope at 20 ppb by averaging the slopes of the two segments adjacent to 20 ppb, that is the slope between 10 and 20 ppb and between 20 and 40 ppb. Subsequently, by averaging the slopes at the threshold level across all the containers, we could quantitatively estimate the sampling power of a sampling plan. For easier interpretation, the estimated slope was presented as the absolute marginal change in acceptance probability at 20 ppb; a larger marginal change indicated a steeper slope.

2.4.2. Sensitivity Analysis

To quantify the impact of aflatoxin concentration, number of probes, clustering level, and sampling strategies on the probability of acceptance, sensitivity analysis was performed using the partial rank correlation coefficient (Marino, Hogue, Ray, & Kirschner, 2008).

Specifically, simulated results of experiment 1 and 2 were pooled together to produce the partial rank correlation coefficient (PRCC) sensitivity analysis plot for sampling in a 3\( \times \)2\( \times \)1.5 m\(^3\) truck. The function “pcc” from the R package “sensitivity” was utilized to calculate the PRCCs for the four input parameters mentioned above with a 95% confidence interval.

2.4.3. Reproducibility

The simulation framework, the simulated experiments, and all the downstream analyses were performed in R (3.6.3) on Windows 10 (64-bit) with the following packages: caTools (1.18.0), spatstat (1.63-3), mc2d (0.1-18), mvnfast (0.2.5), rlang (0.4.5), scales (1.1.0), reshape2 (1.4.3), MASS (7.3-51.5), ggforce (0.3.1), tidyeverse (1.3.0), sensitivity (1.21.0). The random number seed was set at 123. The R code is publicly accessible via https://github.com/ericxbcheng/ILSI_remote_repo.

3. RESULTS

3.1. Model is Validated Against Previously Reported Empirical Data

To prove the accuracy of the model, validation was performed by simulating a sampling event equivalent to the scenario from (Johansson et al., 2000a) and comparing their experimental acceptance probabilities against our simulated probabilities. As a reminder, the model would only be considered validated when the experimental acceptance probabilities fall within the entire range of the simulated ones.

The acceptance probabilities from the simulation were summarized and plotted as an OC curve in Fig. 2. Each dot in the OC curve was the median acceptance probability at that aflatoxin concentration and the shaded area marked the range from the minimum to the maximum. The 18 empirical acceptance probabilities from the study whose corresponding aflatoxin concentration ranged from 5.8 to 676.6 ppb were overlaid onto the plot as triangles. As it is apparent that all the empirical acceptance probabilities fall within the range of simulated probabilities, we can confirm that our simulation model is successfully validated.

3.2. OC Curve Analysis

In an attempt to interpret the performance of a sampling plan, both the OC curve’s decreasing trend and the confidence interval were taken into account. As a reminder, a sampling plan is considered desirable when the OC curve is steep at the threshold concentration (20 ppb), which can be quantified by the absolute marginal change in acceptance probability (Table III). A larger value indicates a steeper curve at 20 ppb. The following sections will present OC curves with 95% confidence intervals for each experiment to facilitate useful comparisons. An OC curve with narrower confidence interval is preferred as it indicates higher stability and thus more consistent performance.

3.2.1. SRS and STRS are More Stable than SS at High Level of Clustering

OC curves for different combinations of clustering levels (1, 10, 100, 1,000 kernels around each cluster source or 50%, 9%, 1%, 0.03% of the contaminated kernels being cluster sources) and sampling strategies (SRS, STRS, SS) were plotted to examine
Simulating Aflatoxin Sampling

Fig 2. The comparison between the entire range of the simulated acceptance probabilities (shaded area) and the observed acceptance probability from the literature (triangle). The shaded area spans the range from minimum to maximum of the simulated probabilities at each concentration and the solid dots represent the median acceptance probabilities.

Table III. The Absolute Marginal Change of the Acceptance Probability at 20 ppb of Aflatoxin for the Two Simulation Experiments

| Primary Parameter | Input Value | 1 (%) | 10 (%) | 100 (%) | 1000 (%) |
|-------------------|-------------|-------|--------|---------|----------|
| Sampling strategy (method_sp) | SRS         | 3.5   | 3.2    | 2.0     | 0.8      |
|                    | STRS        | 3.5   | 3.2    | 2.0     | 0.8      |
|                    | SS          | 3.5   | 3.2    | 1.9     | 0.7      |
| Number of probes (n_sp) | 5           | 3.4   | 3.1    | 1.8     | 0.6      |
|                    | 10          | 3.6   | 3.4    | 2.2     | 0.9      |
|                    | 100         | 3.6   | 3.6    | 3.4     | 2.4      |

the interaction between sampling strategies and clustering level (Fig. 3).

When the clustering level was at the lowest level where only one kernel clustered around a contamination source (50% of the contaminated kernels being cluster sources), the three OC curves had analogous decreasing trend (Fig. 3, top left panel). In detail, almost all the containers were accepted until the overall aflatoxin concentration was 10 or more ppb. As the concentration increased to the threshold level of 20 ppb, only around 50–60% of the containers were accepted. When the concentration reached 40 ppb, almost none of the containers were accepted. In terms of the decreasing rate at 20 ppb, all three sampling strategies resulted in 3.5% absolute change in acceptance probability (Table III), indicating
equivalent sampling power. However, SS had noticeably wider confidence interval especially at 20 ppb, suggesting large variability in sampling power at the threshold level.

As the clustering level increased, the OC curves tended to decrease more gradually, which was reflected in Table III where the absolute marginal change decreased from 3.5% to around 0.8% for all three strategies, indicating that increased clustering may hinder aflatoxin detection. The OC curves of SRS and STRS exhibited almost the same decreasing trends as well as confidence intervals. Meanwhile,
the OC curve of SS showed a much more irregular trend with an extremely wide confidence interval.

At the highest clustering level (1,000 kernels around each cluster source or percentage of cluster = 0.03%), around 80% of the containers at the threshold concentration would be accepted when using SRS or STRS. When adopting SS to sample the containers at the threshold level, the acceptance probabilities could range from 0% to 100% with a median at 100%, meaning that it is common to see all the containers accepted with a few extreme exceptions where all the containers are rejected. While the OC curves between SS and the other two strategies were visually distinct, their averaged marginal change in acceptance probability at the threshold level were extremely similar (around 0.8%), suggesting equivalent averaged sampling performance despite much larger variance in the performance of SS.

### 3.2.2. Extra Probes Facilitate Detection at High Level of Clustering

To elucidate the interaction between clustering level and number of probes, OC curves for different combinations of clustering level (1, 10, 100, 1000 kernels around each cluster source or 50%, 9%, 1%, 0.03% of kernels being cluster sources) and probe quantity (5, 10, 100) were plotted in Fig. 4.

Similar to the findings in the previous section, OC curves at the lowest clustering level were analogous in the decreasing trend as well as the confidence interval. Around 50–60% of the containers at the threshold level were accepted regardless of the number of probes used for sampling. Their absolute marginal change in acceptance probability at the threshold level were also similar (around 3.5%), also suggesting that probe quantity has a negligible impact on sampling when the aflatoxin distribution is homogeneous. The confidence intervals of the three OC curves almost overlapped one another, indicating similar variability in performance.

As the clustering level increased, the three OC curves began to separate in trend and decrease at a lower rate as evidenced by their less extreme marginal changes (Table III). At the clustering level of 1,000 kernels per cluster source (percentage of cluster = 0.03%), the OC curve of fewer probes was almost entirely above the OC curve of more probes. More specifically, while around 80–90% of the containers at the threshold level would be accepted using five or 10 probes, the probability would reduce to around 60% when using 100 probes. This indicates that taking more probes at a high clustering level would facilitate aflatoxin detection. Such a finding was corroborated by the observation that the OC curve of 100 probes was much steeper at the threshold level than that of five probes as evidenced by a 1.8% increase in the absolute marginal change in acceptance probability, suggesting extra probes may exhibit larger sampling power in a container with high level of clustering. Interestingly, the confidence interval for the three OC curves did not widen much in response to the increasing clustering level. Nor did the confidence interval differ noticeably between OC curves of different probe quantity at a fixed clustering level. Given that the sampling strategy was fixed to SRS in this experiment, it is reasonable to speculate that variability in sampling performance mainly stems from the choice of sampling strategy instead of the probe quantity.

### 3.3. Acceptance Probability is Sensitive to Natural Factors

The sensitivity analysis was conducted to assess the impact of the aflatoxin concentration, number of probes, clustering level, and sampling strategies on the acceptance probability. The impact was quantified by the partial rank correlation coefficient (PRCC) with a 95% confidence interval (Fig. 5).

In terms of the magnitude of impact, the aflatoxin concentration had the largest influence (−0.87) on the acceptance probability, followed by the clustering level (0.28), the number of probes (−0.09), and the sampling strategy (0.04). The aflatoxin concentration and the clustering level are the natural factors; while higher aflatoxin concentration would cause a dramatic decrease in acceptance probability, higher clustering level would increase acceptance of aflatoxin. Conversely, the human factors, including the probe quantity and the sampling strategy, had a smaller effect on the acceptance probability than the natural factors do. While taking more probes would be conducive to aflatoxin detection, choosing STRS or SS would only make it less likely to reject a contaminated container compared to using SRS. Taking into consideration the OC curves from Fig. 3, one may draw the conclusion that it was SS that prevented aflatoxin detection, because the OC curve of STRS had almost the same trend and confidence interval as the OC curve of SRS.

We further explored the sensitivity of the acceptance probability to these factors in grain containers of different sizes, including a tote (1 × 1 × 1 m³)
Fig 4. The OC curves under 5, 10, 100 probe samples and four clustering levels (1, 10, 100, 1,000 kernels/cluster source or 50%, 9%, 1%, 0.03% of contaminated kernels being cluster sources). The sampling strategy was fixed at SRS and homogeneity after grinding was fixed at 10%. Each scenario was iterated for 10,000 times; points in the OC curves are the median acceptance probabilities whereas the shaded area represents the 2.5th-97.5th percentile range. The dashed line indicates the FDA aflatoxin action level (20 ppb).

(Supporting Information Appendix Fig. S1) and a hopper car (14 × 3 × 2.5 m³) (Supporting Information Appendix Fig. S2). Interestingly, their PRCC plots demonstrate a pattern similar to the one in Fig. 5 where the natural factors were predominant, and the human factors had less influence. It is worth noting that the sampling strategy’s PRCC indices for the tote and the hopper car were extremely close to 0, indicating that the impact of sampling strategies may be negligible compared to other factors.
4. DISCUSSION

OC curve analyses regarding sampling strategies or probe quantity were conducted to determine how these factors would affect the sampling performance.

In terms of sampling strategies, despite their relatively small impact on acceptance probability as shown in the sensitivity analysis, they are a variable under operator control. Therefore, comparing sampling strategies is still valuable to provide clear guidance for improving grain sampling strategies. Evaluating the effect of these sampling strategies offers theoretical guidance for establishing and improving sampling protocols. The OC curves as well as the slope analyses suggested that they had almost equivalent averaged sampling power to detect aflatoxin in a container. However, using SS would result in larger variability in acceptance probability, especially when aflatoxin was highly clustered. Although few studies investigated the effect of sampling strategies on grain sampling, a simulation study evaluating sampling strategies on produce sampling (Xu & Buchanan, 2019) could be compared with our study. Xu and Buchanan examined the performance of SRS, STRS, and Z-pattern SS (i.e., taking random samples along the edges and the diagonal line of a produce field). Their results indicated that while the three strategies exhibited similar sampling capability, adopting the Z-pattern SS would lead to much higher variability in detection. Overall, our findings were consistent with their conclusion that SRS and STRS may be preferable over SS due to much smaller variance in detection. Interestingly, the substantial variability brought by SS seems to be irrelevant of the geometric pattern (i.e., Z-shape versus uniformly spaced probing). One possible explanation of the large variability is that sampling locations in an SS scenario are either fixed or confined to a small portion of the entire lot, which could make the samples more biased. Such a bias may be intensified by the existence of hotspots as evidenced by the widening confidence interval in response to the increasing clustering level. The reason for this may be that highly clustered kernels are likely either captured by probes altogether or not captured at all. These conjectures
regarding sampling bias coincide with those reported by (Whitaker & Slate, 2012), who pointed out that such a bias could be reduced by randomizing probing locations, increasing probe quantity, and thoroughly mixing the grains. However, comparison of sampling strategies in other food systems may result in different conclusions. (Jongenburger, Reij, Boer, Gorris, & Zwietering, 2011a) compared the performance between SRS and SS when sampling powdered infant formula with localized contamination and concluded that SS was superior to SRS. Meanwhile, another study on sampling powdered products by (Thevaraja, Govindaraju, & Bebbington, 2021) argued that both SRS and SS are equally powerful and taking many incremental samples was more effective for detection than taking a few grab samples of larger sizes. In both cases, whether sampling is random or systematic depends on whether the incremental samples are taken at random time points or at fixed time intervals, which is different from the SS definition in the probe sampling context. Both studies have adopted a theoretical approach to analyzing detection. While (Jongenburger et al., 2011a) combined binomial distribution and Poisson distribution in the calculation, (Thevaraja et al., 2021) utilized the Markov chain method. These methods will provide an analytical solution to estimating the acceptance probability but may be less flexible than the Monte Carlo method used in this article, which could consider the uncertainty and variability in a food system and reflect that in the variance of acceptance probability. In summary, while SS may cause large variability to probe sampling in grains compared to SRS and STRS, its effect in other locally contaminated food systems may be subjected to the assumptions about clustering and how SS is defined.

Corroborating the conclusion by (Whitaker & Slate, 2012), our sensitivity analysis also suggests that increasing the number of probe mechanisms could reduce contamination detection. This is in agreement with the studies about detecting other mycotoxins such as deoxynivalenol and ochratoxin A (Casado et al., 2009a; Oerke et al., 2010). Furthermore, the advantage of extra probes is most prominent when the aflatoxin is mostly accumulated in a few hotspots, which is demonstrated by the diverging OC curves between 5, 10, and 100 probes at higher clustering levels (Fig. 2). When the clustering level is low, the OC curves resemble one another in the decreasing trend as well as the confidence interval, indicating that in a container with uniformly distributed aflatoxin, using fewer probes to detect aflatoxin contamination is just as effective as using more probes. These findings may substantiate the call for more stringent sampling plans for detecting highly clustered hotspots. Currently, the GIPSA protocol recommends using a relatively small number of probes for sampling static grain containers (e.g., seven or nine probes for trucks depending on the depth of grains) (USDA, 2020). These sampling plans may perform well only on the condition that the grains have been sufficiently homogenized so that the target contaminant is uniformly distributed in the container. However, grains in static grain containers are seldom homogenized prior to probe sampling because many static grain containers are primarily designed for transportation (e.g., trucks, barges, hopper cars, etc.) or storage (e.g., storage containers, sacks) and are not equipped with mixing devices. Under these circumstances, it is advisable to utilize a larger number of probes to ensure sample representativeness, especially when the target contaminant is known to exist in hotspots like aflatoxin.

Performing probe sampling on a large scale can be challenging, given that probe sampling is usually carried out manually. Thus, alternative sampling methods that can yield a large quantity of samples with relatively less labor required should be considered, such as autosampling. Autosampling is often implemented for dynamic lots where samples are taken from a moving stream of grains by an automatic sampler (Whitaker & Slate, 2012). Common automatic samplers such as cross-cut samplers are diverters that pass through the grain stream at fixed intervals and capture an incremental amount of grains (USDA, 2020). A study by (Mallmann et al., 2014) revealed that compared to manual probing, autosampling may not only reduce the variance in aflatoxin quantification, but also result in much smaller producer and consumer risks as evidenced by a remarkably steeper OC curve. Its advantage over manual probing may be attributed to the fact that any grain in the moving stream could be captured by the sampler with almost equal chance. Furthermore, the physical interaction between kernels during mass flow may randomize the previously clustered contaminated kernels, which resembles the effect of partial mixing and leads to lower clustering levels. These hypotheses about autosampling could be tested by our model. While our model is designed for manual probe sampling and does not directly simulate mass flow, the OC curve of taking many probes randomly may provide a rough estimate of the autosampling performance for detecting aflatoxin. In this study, our
analysis of using SRS to take 100 probes is a proxy for autosampling for the following reasons. First, adopting SRS ensures equal probability to capture any kernel in the container, which corresponds to the first advantage mentioned above. Second, taking as many as 100 probes can be seen as similar to autosampling with a high sampling frequency. Based on these assumptions, it is reasonable to extrapolate that autosampling may almost always improve aflatoxin detection due to its highest absolute marginal change in acceptance probability at the threshold level, regardless of clustering levels. Furthermore, if partial mixing does exist in autosampling, we may justify autosampling with a relatively lower sampling frequency. This is because a lower clustering level would diminish the difference in sampling power among sampling plans with differing probe quantities. Using a lower sampling frequency to achieve a similar sampling power may help reduce the sampling cost.

One limitation of the simulated experiments is that the total mass of composite probe samples was not controlled across different number of probes. While this approach is consistent with the GIPSA protocol, holding the total mass of composite probe samples constant would result in a different, also useful, comparison between sampling plans with different number of probes. This is because the effect of increasing probes would be no longer associated with the total mass of composite probe samples and would be purely attributable to the dispersion of probes. If interested this comparison, one could use the tool developed here for comparisons that fix the total mass of composite probe samples by reducing the probe diameter for increasing number of probes.

5. CONCLUSIONS

A validated Monte-Carlo simulation model has been designed and built to simulate the entire sampling procedure for detecting aflatoxin in grain containers. Using the model, we examined the effects of sampling strategies (simple random sampling, stratified random sampling, systematic sampling in compliance with the U.S. GIPSA sampling schemes), number of probes (5, 10, 100), and clustering level (1, 10, 100, 1000 kernels/cluster source or percentage of clusters at 50%, 9%, 1%, 0.03%) on the probability of acceptance over 10,000 iterations. Based on the sensitivity analysis with partial rank correlation coefficients, the sampling performance is more sensitive to natural factors (aflatoxin concentration and clustering level) rather than human factors (probe quantity, sampling strategy). The OC curves reveal that while the three sampling strategies exhibit similar sampling power on average, systematic sampling brings much higher variability especially when aflatoxin is concentrated in hotspots. Furthermore, taking extra probes would be particularly useful for detecting aflatoxin when hotspots exist. When aflatoxin is uniformly distributed in the container, taking fewer probes may be just as effective as taking more probes. As aflatoxin distribution in grain containers is usually skewed and clustered, taking a large number of samples, such as by autosampling, may allow more efficient sample collection and improve aflatoxin detection.

This work could guide how to improve manual probing for aflatoxin detection, as well as provide indications on when autosampling based approaches are likely to lead to the largest sampling plan performance gains. Furthermore, given input parameters specific to a contaminated grain container, one can use this model to estimate the probability of rejection (1 – probability of acceptance) when the container is sampled with a chosen sampling strategy and number of probes.

ACKNOWLEDGMENTS

This work was supported by the International Life Sciences Institute (ILSI) North America Food Microbiology Committee. ILSI North America is a public, nonprofit science foundation that provides a forum to advance understanding of scientific issues related to the nutritional quality and safety of the food supply. ILSI North America receives support primarily from its industry membership. ILSI North America had no role in the design, analysis, interpretation, or presentation of the data and results.

References

Bouzembrak, Y., & van der Fels-Klerx, H. J. (2018). Effective sampling strategy to detect food and feed contamination: Herbs and spices case. Food Control, 83(Supplement C), 28–37. https://doi.org/10.1016/j.foodcont.2017.04.038
Casado, M. R., Parsons, D. J., Weightman, R. M., Magan, N., & Origgi, S. (2009a). Geostatistical analysis of the spatial distribution of mycotoxin concentration in bulk cereals. Food Additives & Contaminants: Part A, 26(6), 867–873. https://doi.org/10.1080/02652260902788812
Casado, M. R., Parsons, D. J., Weightman, R. M., Magan, N., & Origgi, S. (2009b). Modelling a two-dimensional spatial distribution of mycotoxin concentration in bulk commodities to design effective and efficient sample selection strategies. Food Additives & Contaminants: Part A, 26(9), 1298–1305. https://doi.org/10.1080/02652260902402517
Chavez, R. A., Cheng, X., & Stasiewicz, M. J. (2020). A review of the methodology of analyzing aflatoxin and fumonisin in single corn kernels and the potential impacts of these methods on food security. *Food Sec., 9*(3), 297. https://doi.org/10.1007/s12571-019-00309-z

Cheng, X., Chavez, R. A., & Stasiewicz, M. J. (2020). When to use one-dimensional, two-dimensional, and shifted transversal design pooling in mycotoxin screening. *PLOS ONE, 15*(8), e0236668. https://doi.org/10.1371/journal.pone.0236668

FDA. (2000). *Guidance for industry: Action levels for poisonous or deleterious substances in human food and animal feed*. Retrieved from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-action-levels-poisonous-or-deleterious-substances-human-food-and-animal-feed

Gnonlonfin, G. J. B., Hell, K., Adjoji, Y., Fandohan, P., Koundande, D. O., Mensah, G. A., … Brimer, L. (2013). A review on aflatoxin contamination and its implications in the developing world: A sub-Saharan African perspective. *Critical Reviews in Food Science and Nutrition, 53*(4), 349–365. https://doi.org/10.1080/10408398.2010.535718

IARC. (2018). *Aflatoxins: IARC monographs—100F*. Retrieved from https://monographs.iarc.fr/wp-content/uploads/2018/06/mono100F-23.pdf

Johannson, A. S., Whitaker, T. B., Hagler, W. M., Giesbrecht, F. G., Young, J. H., & Bowman, D. T. (2000). Testing shelled corn for aflatoxin, Part I: Estimation of variance components. *Journal of AOAC International, 83*(5), 1264–1269.

Johannson, A. S., Whitaker, T. B., Giesbrecht, F. G., Hagler, W. M., & Young, J. H. (2000a). Testing shelled corn for aflatoxin, part II: Modeling the observed distribution of aflatoxin test results. *Journal of AOAC International, 83*(5), 1270–1278. https://doi.org/10.1093/aoac/83.5.1270

Johannson, A. S., Whitaker, T. B., Giesbrecht, F. G., Hagler, W. M., & Young, J. H. (2000b). Testing shelled corn for aflatoxin, part III: Evaluating the performance of aflatoxin sampling plans. *Journal of AOAC International, 83*(5), 1279–1284. https://doi.org/10.1093/aoac/83.5.1279

Jongenburger, I., den Besten, H. M. W., & Zwietering, M. H. (2015). Statistical aspects of food safety sampling. *Annual Review of Food Science and Technology, 6*(1), 479–503. https://doi.org/10.1146/annurev-food-022814-015546

Jongenburger, I., Reij, M. W., Boer, E. P. J., Gorris, L. G. M., & Zwietering, M. H. (2011a). Random or systematic sampling to detect a localised microbial contamination within a batch of food. *Food Control, 22*(8), 1448–1455. https://doi.org/10.1016/j.foodcont.2011.03.009

Jongenburger, I., Reij, M. W., Boer, E. P. J., Gorris, L. G. M., & Zwietering, M. H. (2011b). Actual distribution of Cronobacter spp. in industrial batches of powdered infant formula and consequences for performance of sampling strategies. *International Journal of Food Microbiology, 151*(1), 62–69. https://doi.org/10.1016/j.ijfoodmicro.2011.08.003

Krska, R., & Molinelli, A. (2007). Mycotoxin analysis: State-of-the-art and future trends. *Analytical and Bioanalytical Chemistry, 387*(1), 145–148. https://doi.org/10.1007/s00216-006-0797-3

Mallmann, A. O., Marchioro, A., Oliveira, M. S., Rauber, R. H., Dilkin, P., & Mallmann, C. A. (2014). Comparison of the efficiency between two sampling plans for aflatoxins analysis in maize. *Brazilian Journal of Microbiology, 45*(1), 35–42. https://doi.org/10.1590/S1517-838220140000100006

Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology, 254*(1), 178–196. https://doi.org/10.1016/j.jtbi.2008.04.011

Oerke, E.-C., Meier, A., Dehne, H.-W., Sulyok, M., Krska, R., & Steiner, U. (2010). Spatial variability of fusarium head blight pathogens and associated mycotoxins in wheat crops. *Plant Pathology, 59*(4), 671–682. https://doi.org/10.1111/j.1365-3059.2010.02286.x

Sargent, R. G. (2010). Verification and validation of simulation model 297. *Proceedings of the 2010 Winter Simulation Conference Baltimore, MD. Pp. 166–183. https://doi.org/10.1109/WSC.2010.5679166

Shotwell, O. L., Goulden, M. L., Bothast, R. J., & Hesseltine, C. W. (1975). Mycotoxins in hot spots in grains. I. Aflatoxin and zearalenone occurrence in stored corn. *Cereal Chemistry, 52*(5), 687. https://pubag.nal.usda.gov/catalog/29962

Stasiewicz, M. J., Falade, T. D. O., Mutuma, M., Mutiga, S. K., Harvey, J. J. W., Fox, G., … Nelson, R. J. (2017). Multi-spectral kernel sorting to reduce aflatoxins and fumonisins in Kenyan maize. *Food Control, 78*, 203–214. https://doi.org/10.1016/j.foodcont.2017.02.038

Thevaraja, M., Govindaraju, K., & Bebbington, M. (2021). Modelling the effect of sampling methods on detection tests for powdered products. *Food Control, 120*, 107512. https://doi.org/10.1016/j.foodcont.2020.107512

USDA. (2015). Chapter 4 sample preparation procedures. In *Mycotoxin handbook* (pp. 32–41). Washington, DC: USDA, Agricultural Marketing Service, Federal Grain Inspection Service.

USDA. (2018). FGIS’s role in aflatoxin testing | agricultural marketing service. Retrieved from https://www.ams.usda.gov/publications/content/fgis%2E exploding%20role-aflatoxin-testing

USDA. (2020). Chapter 2 probe/trier sampling. In *Grain inspection handbook: Sampling* (pp. 28–48). Washington, DC: USDA, Agricultural Marketing Service, Federal Grain Inspection Service.

Whitaker, T. B., Dickens, J. W., & Wiser, E. H. (1976). Monte Carlo technique to simulate aflatoxin testing programs for peanuts. *Journal of the American Oil Chemists’ Society, 53*(8), 545–547. https://doi.org/10.1007/BF02586257

Whitaker, Thomas B., & Slate, A. B. (2012). Sampling and sample preparation for mycotoxin analysis. In E. M. Binder & R. Krska (Eds.), *Romer labs guide to mycotoxins* (4th ed.). Leicestershire, UK: Anytime Publishing.

Wu, F., Groopman, J. D., & Pestka, J. J. (2014). Public health impacts of foodborne mycotoxins. *Annual Review of Food Science and Technology, 5*(1), 351–372. https://doi.org/10.1146/annurev-food-030713-092431

Xu, A., & Buchanan, R. L. (2019). Evaluation of sampling methods for the detection of pathogenic bacteria on pre-harvest leafy greens. *Food Microbiology, 77*, 137–145. https://doi.org/10.1016/j.fm.2018.09.007

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Table A1.** Glossary of mathematical notations.

**Fig. S1.** The PRCC sensitivity analysis plot for sampling in a $1 \times 1 \times 1 \text{ m}^3$ tote with 10,000 iterations. Each PRCC index was computed with 95% confidence interval.

**Fig. S2.** The PRCC sensitivity analysis plot for sampling in a $14 \times 3 \times 2.5 \text{ m}^3$ tote with 10,000 iterations. Each PRCC index was computed with 95% confidence interval.