Sampling Plans for Microbiological Testing – How to Construct and Role in Testing

Russell S. Flowers and Timothy A. Freier

Microbiological tests are performed to reach a decision or judgment. If the purpose for collecting a sample cannot be defined, then the analysis should probably not be done. The rationale for testing falls into four general categories: to determine safety, to determine adherence to Good Manufacturing Practices (GMPs), to determine the utility of a food or ingredient for a particular purpose, and to predict product stability. Microbiological testing also may be used to gather background information, such as baseline data. However in food safety and quality programs, decision-making based on microbiological data requires that limits be established to differentiate acceptable from unacceptable product. Limits including methods and sampling plans are defined as microbiological criteria, which include standards, guidelines and purchase specifications.

Standards are limits established by international, federal and regional laws. Exceeding a standard relating to a pathogen, such as *Salmonella* or *Listeria*, may lead to a product recall and/or punitive action. Guidelines are internal, advisory criteria established by a processor or by a trade association. A wide variety of criteria fit into this category, such as results on pre-op swabs from equipment, in-process samples of product or equipment, and environmental samples tested for pathogens. Failure to meet a guideline usually serves as an alert to the processor, indicating that remedial action should be taken. Purchase specifications are agreements between the vendor and buyer of a product as a basis for sale. Failure of the vendor to meet specifications can be used as a basis for product rejection.

The application of sampling plans for microbiological criteria was pioneered by the International Commission on Microbiological Specifications for Foods (ICMSF). The sampling developed by ICMSF are comprised of the following parameters: 1) the number of samples drawn from the lot and individually analyzed for the defect (n); 2) the number of samples permitted to exceed the limit (c); and 3) the limit (m). The sampling plan can be adjusted for stringency by varying these factors. As the number of samples increases, so does stringency. Likewise, as the number of samples that are allowed to exceed the limit decreases (c) so does the stringency of the plan.

Table 1 shows the probabilities of acceptance in a sampling plan where c = 0. Even when large numbers of samples within a lot are found negative, there is no inference of the complete absence of the target organism. Table 2 shows the BAM/FDA sampling plan for *Salmonella*.

The most stringent plan for foods is Category I, where sixty 25 g samples are analyzed. If each is negative, then the level of *Salmonella* is less than or equal to one in 500 g at the 95% confidence level. From Table 2 it is clear at the highest plan stringency (n = 60 x 25 g samples) that a lot containing 2% defective units would be accepted ca. 30% of the time. Thus, negative test results for the most stringent FDA sampling does not indicate the complete absence of *Salmonella*, only a certain confidence that the level is below the established limit.

The foregoing discussion of attribute sampling is related to two-class sampling plans. Here, results fall into two categories, good quality and defective quality. Three-class sampling plans classify results into three categories: acceptable, marginal, and unacceptable. In addition, three class plans include another parameter (M). This is a level at which there is "decisive concern." The value of M is a judgment based on knowledge of the product relating either to public health or spoilage. If any part of the lot contains product in the area of decisive concern, then it should be rejected, even if the number of marginal samples does not exceed that permitted by a two-class plan. Three-class plans are more discriminating than two-class plans. There is a tendency for those using three-class plans to assume a mathematical relationship between the values of m and M. This is not the case, as the value of m usually relates to GMPs, whereas M represents decisive concern and may be a value several magnitudes above m.

Attribute sampling requires random distribution of the defect within the lot; that is every sample has an equal chance of containing the defect. In the real world, this is often not the case. However, the sampling plan is independent of lot size, as long as the defect is randomly distributed within the lot.

The level and incidence of contamination must also be considered when selecting a sampling plan. When the microorganism is present at a lower incidence, not every test portion will contain the organism. Detection now depends upon the probability of selecting a contaminated portion (Table 1). In the case of a low probability event, such as contamination of ground beef with *E. coli* O157:H7, one positive result encountered in thousands of tests does not necessarily indicate

Silliker Laboratories Group, Inc.
900 Maple Avenue,
Homewood, IL 60430

TABLE 1. Two-class plans (c=0): probabilities of acceptance.

Composition of lot		Number of sample units tested				
% acceptable	% defective	5	10	20	60	100
98	2	.90	.82	.67	.30	.13
95	5	.77	.60	.36	.05	.01
90	10	.59	.35	.12	<	<
80	20	.17	.11	.01		
70	30	.03	.03	<		
50	50	.01	<			
40	60	<				
30	70					

that the lot testing positive is more highly contaminated than lots testing negative. The positive may simply have resulted from the 1 in 1,000s probability and this lot may be no more highly contaminated than the products testing negative. Re-sampling to confirm a positive result may not be meaningful. Confirmation of the positive result requires that a second positive sample be selected. The probability of selecting two positive samples is much lower than the probability of selecting the first positive portion. The probability that both will be positive is the square of the probability of a single portion testing positive (Table 3). The lower the incidence of contamination, the more difficult it will be to confirm. Confirmation will depend upon luck, or testing until the incidence of contamination is established. Very low incidence of contamination will be virtually impossible to confirm by re-sampling.

In the case of a routine testing of a low probability defect, batches may test negative but be contaminated at a low incidence. The USDA's "Moving Sum Rules" is a case in point (Table 4). As manufacturers fall within acceptable limits, there is a decreasingly lower possibility that a positive sample will be encountered. Thus, leading to the conclusion that when a positive sample is encountered, the proceeding lots may have been contaminated at a similar level but not detected. Although only one sample per day is analyzed, in order to be in compliance, the manufacturer must generally be producing product that will meet the acceptable limit if *n* (the window size) samples were tested each day. For example, ground beef has an *AL* of 2 in 38 days. To assure compliance, a manufacturer must produce product each day that would have an average incidence of contamination no greater than 2 in 38. Any given day's production might be higher or lower, but the cumulative incidence of contamination must be 2 or less positives in 38 samples.

In situations where the defect is randomly distributed, attribute sampling can be used to assess compliance with regulatory standards, purchase specifications, and to test the safety or utility of raw ingredients and finished product. If the distribution is not random, then attribute sampling may not be applicable, and the number of samples needed in order to reach a conclusion may be far greater. Alternatively, investigational sampling using a biased approach may be more productive if the investigator is knowledgeable about the distribution.

Frequently, single samples are utilized by regulators and by industry for decision making. These results mean very little if they indicate compliance with the standard or specification, as the confidence that can be placed in them is very low. Only if the results indicate non-compliance, is the information very meaningful.

TABLE 2. BAM/FDA sampling plans for *Salmonella.*

- **Defines 3 categories of product**
  - I = Product intended for aged, infirmed, infants, and immunocompromised
  - II = No lethal process between time of sampling and consumption
  - III = A lethal process exists between time of sampling and consumption
- **Sampling plans**

I = 60 x 25 g = 4 x 375 g = 95% confidence $\leq$ 1 *Salmonella* / 500 g
II = 30 x 25 g = 2 x 375 g = 95% confidence $\leq$ 1 *Salmonella* / 250 g
III = 15 x 25 g = 1 x 375 g = 95% confidence $\leq$ 1 *Salmonella* / 125 g

TABLE 3. Probability of detecting a confirming analyte at low contamination.

Probability of positive test	Probability of test and retest positive
1 in 2	1 in 4
1 in 5	1 in 25
1 in 10	1 in 100
1 in 20	1 in 400
1 in 50	1 in 2,500
1 in 100	1 in 10,00

Sampling and microbiological testing are useful to establish baseline data, screen raw materials and to verify control. However, it is often not practical to sample and test a sufficient number of samples to obtain meaningful information relative to a specific lot or batch. However, microbiological criteria can be established and are useful when an acceptable limit is determined that is within a range allowing practical application of sampling plans and testing protocols. However, it must be recognized that no feasible sampling plan can ensure the absence of a pathogen. All microbiological criteria involve both consumer and producer risk; i.e. the risk of accepting "bad" food and rejecting "good" food, respectively**.**

TABLE 4. Moving Sum Rules.

Commodity	Target (% positive for *Salmonella*)	Window size (n) in days	Acceptable Limits (AL)
Steers/ Heifers	1	82	1
Cows/Bulls	1	82	1
Raw ground beef	4	38	2
Fresh pork Sausages	12	19	3
Turkeys	15	15	3
Hogs	18	17	4
Broilers	25	16	5

Federal Register, Feb. 3, 1995, Vol. 60, No. 23, Proposed Rules for Pathogen Reduction; HACCP Systems.