Calculating Sample Sizes and Thresholds

A challenge when trying to implement sampling programs within a feed mill is defining proper calculation of the necessary sample size to detect pathogens of interest. If interested in pre-determined sample size recommendations and thresholds, refer to "Sample Size for Surveillance and Suspected Contamination." However, if a production system or feed mill has a general idea of prevalence rate for the pathogen of interest and would like to defer from the general recommendations, this resource aims to explain how to calculate sample size, basis for sample size, and how to set thresholds for sampling feed mills. If interested in how to transition from surveillance to suspected contamination sampling, refer "Transitioning from Surveillance Sampling to Suspected Contamination Sampling."

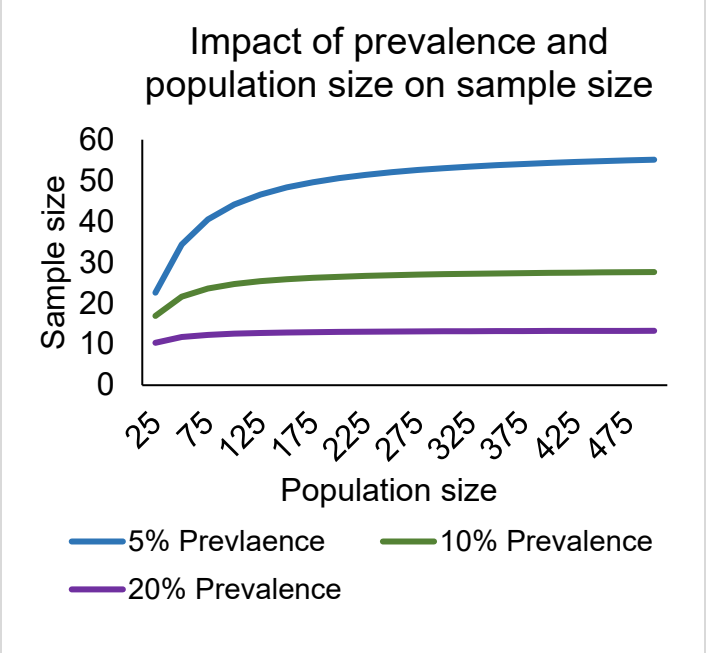
Sample size formula

The formula to determine sample size is the same as that which is used for sampling to detect disease in animals (Dahoo et al., 2014). This formula is used to determine the sample size necessary to have confidence in the outcome while minimizing interpretation error. It takes in to account the number of animals from within a population that must be sampled from a population to have a given level of confidence that at least one sample would be positive based on a given prevalence level. The sample size, n, is determined by the confidence interval (α = 1 – confidence level), the population size (N), and estimated minimum number of diseases animals in the group (D = estimated prevalence population size):

$$n = \left[\left(1 - \alpha^{1/D}\right) \times \left(N - \frac{D-1}{2}\right)\right]$$

However, when considering feed mills and the presence of a pathogen of interest, the feed mill is more concerned about detecting the pathogen of interest within feed or feed mill environment and not clinical disease within an animal. So in this instance, the sample size could be thought of as the number of total samples to take at the feed mill, the population could be thought of as the possible number of samples to take either in feed or in the environment, and the prevalence as the perceived prevalence of the pathogen of interest within the feed mill. When interpreting the formula in a scenario like this, it can be inferred that the number of possible samples to be collected approaches infinity because samples could be taken per ton, per pound, or per gram.

As shown in the graph on the right, when utilizing the sampling to detect disease formula, the sample size for a desired prevalence level will plateau. This is due to there being minimal changes within the sample sizes as the population size increases.



Houston, Grace E., Gebhardt, Jordan T., Jones, Cassandra K., Woodworth, Jason C., Paulk, Chad B., and Dritz, Steve S. 2022. Kansas State University Feed Safety Sampling Resources: *Calculating Sample Sizes and Thresholds*.

So if you are looking for a desired prevalence level within a population, identify where the sample size plateaus for the prevalence level, and utilize the sample size corresponding with the start of the plateau. As illustrated in the graph on the previous page, as prevalence rate increases, the sample size decreases which is because the pathogen of interest is present at higher percentages and thus, requires a smaller sample size to detect it within the population. When consulting the pre-determined sample sizes for surveillance, the basis for sample size recommendations was based on this same principle.

Sample size references

When trying to set sample sizes for feed and environmental samples, there are some peer-reviewed works to reference (Table 1). These sources evaluated the prevalence, or the number of samples containing detectable pathogen of interest divided by the total number of samples collected, for some pathogens within feed mills. Feed samples are defined as samples pulled directly from the source (either from feed ingredients or complete feed) while environmental samples are samples from surfaces. The data from these published works give reference ranges for prevalence of pathogens and can help guide surveillance sampling.

Thresholds

Thresholds designate action because the prevalence rate has increased during surveillance sampling. Thresholds can be adjusted to be stricter or more lenient with increased prevalence rates depending upon the production system and how risk averse they choose to be.

References

Dahoo, I., Martin, W., and Stryhm H. 2014. Sampling. In: S Marget McPike, editor, Veterinary Epidemiologic Research. VER Inc. Charlottetown, Prince Edward Island, Canada. p. 33-56.

Table 1. Prevalence rates from available published data where pathogens were naturally present in feed ingredients, complete feed, or feed mill environments.

Item	Reference doi	Pathogen[1]	Prevalence, %	Total Number of Samples Collected
Environmental Samples:				
Elijah et al., 2022	10.54846/jshap/1250	PDCoV	2.33	86
Elijah et al., 2022	10.54846/jshap/1250	PEDV	2.33	86
Garrido-Mantillo et al., 2022	10.1111/tbed.14354	PEDV	37.50	8
Gebhardt et al., 2021	10.1111/tbed.14335	ASFV	0.73	2186
Magossi et al., 2019	10.1002/mbo3.711	Salmonella sp.	66.24	237
Magossi et al., 2019	10.1002/mbo3.711	*Salmonella enterica*	19.75	157
Feed Samples:				
Gebhardt et al., 2021	10.1111/tbed.14335	ASFV	0.70	142
Leme et al., 2019	10.111/tbed.13215	SVA	25.93	27
Wu el al., 2021	10.1111/tbed.14209	PEDV	14.29	77

Environmental sample summary:	**Feed sample summary:**
Minimum prevalence: 0.73% Average prevalence: 21.48% Maximum prevalence: 66.24%	Minimum prevalence: 0.70% Average prevalence: 13.64% Maximum prevalence: 25.93%

[1]Abbreviations defined as: African swine fever virus (ASFV), Porcine deltacoronavirus (PDCoV), Porcine epidemic diarrhea virus (PEDV), Seneca valley virus (SVA)