USDA

FSIS Estimation of Pathogen Rates in the Population Represented by the Samples Collected in Each Sampling Project for Fiscal Year 2015

Introduction

FSIS created a SAS program to calculate aggregate sampling pathogen rates. This document describes the methods.

Types of Calculations

FSIS computed one of three types of calculations for a sampling project:

- Prevalence – a calculation for sampling projects in which the selected samples represented all product produced by all establishments that produced the product (except for establishments that were exempted for religious purposes or because their production volume did not meet the threshold for inclusion).
- Volume-weighted percent positive – a calculation applied to sampling projects that did not include all establishments; inference can be made only to the entire product produced by the included establishments.
- Percent positive –calculated as the percent of samples that tested positive. No weighting was done to account for the volume of product produced by the sampled establishments. This type of calculation was generated for sampling projects with small sample sizes or for exploratory sampling projects.

Sampling Projects

Table 1 contains the list of sampling projects covered by this report.

Table 1: FSIS Sampling Projects Included in This Report

Project Code	Project Name
EM31	Egg Products Sampling - Pasteurized - Egg Whites – *Salmonella*
EM32	Egg Products Sampling - Pasteurized - Whole Egg or Yolks – *Salmonella*
EM33	Egg Products Sampling - Pasteurized - Whole Eggs with Added Yolks or Whole Egg Blends – *Salmonella*
EM34	Egg Product Sampling - Pasteurized - Whole Eggs or Yolks – *Salmonella*
EM34	Egg Product Sampling - Pasteurized - Whole Eggs or Yolks with > 2% salt or sugar added - *Salmonella*
EM35	Egg Products Sampling - Pasteurized - Dried Yellow Egg Products – *Salmonella*
EM36	Egg Products Sampling - Pasteurized - Dried Egg Whites – *Salmonella*
EM37	Egg Products Sampling - Pasteurized - Pan Dried Egg Whites – *Salmonella*
EXP_CH_MSK01	Exploratory Sampling for Mechanically Separated Chicken
EXP_TU_MSK01	Exploratory Sampling for Mechanically Separated Turkey
HC_CH_CARC01	HACCP Verification for Young Chicken Carcasses
HC_CH_COM01	Sampling for Ground and Other Comminuted Chicken (not Mechanically Separated)
HC_CPT_LBW01	Sampling for Chicken Parts – Legs, Breasts, and Wings
HC_TU_CARC01	HACCP Verification for Young Turkey Carcasses
HC_TU_COM01	Sampling for Ground and Other Comminuted Turkey (not Mechanically Separated)
MT43	Risk-based Sampling of Raw Ground Beef or Veal Products - *E.coli* O157:H7 & *Salmonella*

Project Code	Project Name
MT54	Sampling of Raw Ground Beef or Beef Patty Components (other than trim)
MT55	Sampling of Bench Trim for further use in ANY raw, non-intact beef products
MT60	Sampling of Beef Manufacturing Trimmings
MT64	Sampling of Raw Ground Beef or Beef Patty Components (other than trim)
MT65	Sampling of Bench Trim for further use in ANY raw, non-intact beef products
RTEPROD_RAND	RTEPROD Sampling - Random RTE Products
RTEPROD_RISK	RTEPROD Sampling - Risk-based RTE Products

Data Extracted from FSIS Databases

For each sampling project in Table 1, FSIS extracted all samples collected in fiscal year 2015 (October 1, 2014 to September 30, 2015) with lab result "Negative", "Positive" or "Confirmed Negative of a Previous Presumptive". FSIS restricted the samples to those with test codes for pathogens *Salmonella*, *Campylobacter*, *Listeria monocytogenes* (*Lm*), *Escherichia coli* (*E. coli*) O157:H7, or non-O157 Shiga Toxin-producing *E. coli* (non-O157 STEC).

In addition to lab sample results data, FSIS extracted the slaughtered totals for chicken and turkey carcasses (sampling projects HC_CH_CARC01 and HC_TU_CARC01) for each establishment that produced these products. These data were used to statistically weight the samples in the computation of prevalence for these two sampling projects. The slaughter totals were for the months May to September of 2015 since these sampling projects went into effect in May of 2015.

FSIS also extracted data on establishment product volume and production days from its databases. These data were used to compute statistical weights to estimate prevalence or volume-weighted percent positive for all project/pathogen pairs for which either of these two types of estimates was computed (other than chicken or turkey carcasses). As noted in the preceding paragraph, slaughtered totals were used for these two poultry products.

FSIS updates its databases regularly so sample data can change. Fiscal year 2015 data for this report were as of December 11, 2015.

Statistical Methods

FSIS used the SAS Institute Inc., Cary, North Carolina software SAS 9.4 (TS1M2) to extract the data from the PHIS and DW databases and for all computations.

Data Preparation and Consolidation of Sampling Projects

Since FSIS extracted sample data from two databases (Public Health Information System [PHIS] and Data Warehouse [DW]), FSIS prioritized inclusion of samples from PHIS as it is the contemporary FSIS database. Duplicated samples from the DW were removed. FSIS then grouped all egg projects into a single project. FSIS also combined MT54 with its replacement project MT64, and combined MT55 with its replacement project MT65.

Computation of Sampling Weights

Projects for which FSIS calculated prevalence or volume-weighted percent positive first required an estimate of each establishment's total product volume for the year. Each establishment's estimated annual production volume for its relevant products was calculated from the following two components from establishment product profiles in PHIS:

1) The average daily production volume for the relevant product groups within a given establishment. (Each product group is assigned a mean daily value of one of seven volume group ranges, as per FSIS PHIS Directive 5300.1).
2) The maximum number of production days in a month for the relevant products in each establishment.

The product of these two values yielded a monthly estimated volume which was then multiplied by 12 to get the estimated annual production volume. Another step followed for projects such as MT43 and MT60 where samples of a specific weight (325 grams for these two projects) were tested. The estimated annual volume was converted to the number of 325-gram units of the particular product, as 325-gram portions of the product were sampled.

As an example of the computation of the total annual volume and the statistical sampling weights, suppose an establishment's estimated mean daily volume was 50,000 pounds of ground beef subject to sampling under the MT43 project, and its maximum number of production days in a month was 20. Since 325 grams of the product was tested for each sample, FSIS computed the total number of 325-gram units of the product in the population as follows:

$$Total\ volume \approx \frac{50{,}000\ pounds}{day} \times \frac{20\ days}{month} \times \frac{12\,months}{year} \times \frac{453.6\,grams}{pound} \times \frac{1\,sample\,unit}{325\,grams} = 16748307.69\ \text{ units of } 325\,\text{grams}$$

If 50 samples were tested for the establishment in the year, each sample was assigned a statistical sampling weight of: 16,748,307.69/50 = 334,966.15. No rounding was done in the programs.

FSIS computed the sampling weights for projects HC_CH_CARC01 and HC_TU_CARC01 (chicken and turkey carcasses, respectively) by dividing the slaughtered total by the number of samples for each establishment covered by these two projects. This was done separately for each of these two projects. The samples and slaughter totals for these projects were for the months May to September of 2015 as these projects went into effect in May 2015. The sampling weights for these two projects reflected the actual production volumes for the establishments.

Estimation of Percentages, Standard Errors (Variances) and Confidence Intervals

All sampling units (carcasses, parts, or 325-gram portions of a product) in all establishments did not have an equal probability of selection because the volume of product varied greatly across establishments. While the selected samples for an establishment (for a given project) were equally weighted because they had an equal chance of selection within the establishment, the statistical weights varied across establishments. Unequal weighting of samples applied to each project for which prevalence or volume-weighted percent positive was estimated.

FSIS employed a SAS procedure designed to produce calculations and standard errors of the calculations by taking into account the complexities of the sample design. FSIS estimated the variances, hence the standard errors, of the percentages with the Jackknife, a statistical method commonly used to get estimates of standard errors for complex sample designs.

The commonly used linear confidence interval for a proportion or percent can result in a confidence limit that is not between 0 and 100 percent for estimates near 0 or 100, leading to artificial truncation of the confidence interval. Logit-transformed confidence intervals are guaranteed mathematically to be between 0 and 1 for a proportion so there is never any artificial truncation of a confidence interval; no limit would need to be adjusted because it is negative or exceeds 1 (100%).

Logit-transformed confidence intervals are obtained by first computing the confidence limits on the logit scale and then transforming them back to a proportion or percent scale. Obtaining the confidence interval on the logit scale and transforming the limits to get the confidence interval for a proportion is more accurate than the common linear interval, especially for a proportion far from one-half (Fleiss et al., p. 295). As the percent of samples that tested positive for a pathogen in FSIS sampling projects generally fell below 10 percent, FSIS decided to use logit-transformed confidence intervals for the estimated percentages.

The logit is the natural logarithm of the odds of a positive test. The odds of a positive test is the ratio of the probability of a positive test to the probability of a non-positive test. If p is the estimated proportion positive of a given pathogen for a particular product, the logit of p is defined as:

$$\log it(p) = \ln \frac{p}{1-p} \text{ and } p = \frac{e^{\log it(p)}}{1+e^{\log it(p)}}$$

where e is the exponential function. Fleiss et al. (p. 295) showed that the estimated standard error of p is:

$$se(p) = p(1-p)se[\log it(p)]$$

From the last equation, it follows that the standard error of the logit of p is:

$$se[\log it(p)] = se(p)/(p(1-p))$$

On page 195 of their book, Fleiss et al. provided the following formula for the limits of an approximate 100(1-α) percent confidence interval for the logit of p:

$$\log it(p) \pm z_{1-\alpha/2} se(\log it(p))$$

The level of confidence for confidence interval estimation is 100(1-α) percent. To get 95 percent confidence intervals, $\alpha=.05$. The inverse transformation of the confidence limits for the logit yields the following formulas for the confidence limits for a proportion (m in the formulas is the logit of p):

Lower confidence limit: $\left[1+e^{-m+z_{1-\alpha/2}\frac{se(p)}{p(1-p)}}\right]^{-1}$

Upper confidence limit: $\left[1+e^{-m-z_{1-\alpha/2}\frac{se(p)}{p(1-p)}}\right]^{-1}$

Proportions and their confidence limits would then be multiplied by 100 to transform them to the percent scale. The SAS procedure that we utilized for estimation provided logit confidence limits as an option that FSIS selected.

Reference

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions* (3rd ed.). Hoboken, New Jersey: John Wiley & Sons.