United States Department of Agriculture

Food Safety and Inspection Service

Report on *Campylobacter* Testing of Poultry Products – decision to suspend the qualitative test (30 mL)

Background

In 2009, FSIS began the process of considering a number of different options for reducing the incidence of *Campylobacter* on young chickens and young turkeys. Two primary factors influenced the decision for selecting the performance standard option:

- A desire to prevent raw poultry products with high levels of *Campylobacter* from reaching consumers (as compared to low levels)
- The need to integrate sample collection with the *Salmonella* testing in order to control costs and minimize the sample collection burden on field staff and slaughter facilities

For young chicken carcasses, the originally proposed guidance required a reduction in the annual incidence of contaminated carcasses. This initial proposed standard for *Campylobacter* differed from the *Salmonella* standard because two analyses on a sample are performed for testing of *Campylobacter*. The first analysis was a test using a 30 mL aliquot that could detect lower levels of *Campylobacter*. A second test using a 1 mL aliquot was performed at the same time and would only be able to detect higher levels. On average, the second test will be positive only when the levels of *Campylobacter* on the carcass are 30 times higher than on the first test. The proposed *Campylobacter* standard used the same 51 samples collected for *Salmonella* and allowed a maximum of 27 positives carcasses on the 30 mL test. The standard allowed only 8 *Campylobacter*-positive samples on the 1 mL sample.

During revisions of the proposed *Campylobacter* standard, the 30 mL test portion was dropped and only the remaining 1 mL test portion was considered. Therefore, a young chicken establishment would be determined compliant or non-compliant based on sample set criteria of the 1 mL portion only (8 *Campylobacter* positive samples are acceptable out of 51 sample sets). FSIS did, however, continue to perform the 30 mL test in order to allow for further evaluation of the originally proposed standard[1].

Because data are now available to compare the results from the 1 mL and 30 mL portions, FSIS is evaluating whether to continue analyzing the 30 mL portion or suspend this analysis. This

[1] For young turkeys, the original *Campylobacter* performance standard did not have a qualitative component (based on a 24mL swab) because the data did not warrant a second level criterion for compliance. Therefore, no performance standard at the 24mL portion could be evaluated.

decision hinges on the magnitude of improvement in detecting non-compliant slaughter establishments that might attend inclusion of the 30 mL results. The following analysis explains FSIS's findings with respect to this decision.

Methods

Multiple tests provide results that can be interpreted in parallel or in series. Parallel interpretation will consider a tested entity to be positive if any one of the tests is positive. Series interpretation will consider a tested entity to be positive only if all of the tests are positive. For *Campylobacter* sample sets, parallel interpretation of the 1 mL and 30 mL portions would classify a young chicken slaughter establishment as failing the performance standard if either there were more than 8 positive 1 mL samples or there were more than 27 positive 30 mL samples. If these results were interpreted in series, then both portions would need to be in excess of these criteria for the slaughter establishment to fail the performance standard.

In the following analysis, it is assumed that the 1 mL and 30 mL results are interpreted in parallel. Such an interpretation serves to increase the classification sensitivity of the combined tests. Classification sensitivity describes the probability that a non-compliant establishment fails the sample set criteria. A non-compliant establishment is a slaughter establishment whose process control is inadequate and such establishments are expected to fail the performance standard. Classification sensitivity increases in parallel interpretation because a non-compliant establishment can fail either of two ways.

Because parallel interpretation increases classification sensitivity, it necessarily reduces the classification specificity of the combined assays. Classification specificity is the probability that a compliant establishment passes the sample set criteria. A compliant establishment is one with adequate process control; such establishments are expected to pass the performance standard. Classification specificity is reduced because a compliant establishment might be misclassified in either of two ways.

Because parallel interpretation increases classification sensitivity, it may improve the public health effects of a performance standard based on both sample portions. By potentially detecting more non-compliant establishments, this interpretation might motivate more of these establishments to improve their process control, thereby reducing exposure of consumers to contaminated poultry meat. Because the *Campylobacter* performance standard's goal is to achieve improved public health, the parallel interpretation is appropriate.

In contrast, interpreting the 1 mL and 30 mL portion results in series would reduce classification sensitivity while increasing classification specificity. Such an interpretation might be rational if the goal of the performance standard were to minimize its effect on slaughter establishments by limiting the number of compliant establishments that are falsely classified as failing. Although the consequences for such establishments are not necessarily trivial, FSIS believes its mission is to focus on public health.

The decision to continue or suspend the 30 mL portion hinges on the magnitude of improvement in classification sensitivity between a performance standard using the 1 mL portion only and a performance standard using the 1 mL and 30 mL portions interpreted in parallel. In addition, the laboratory costs of analyzing the 30 mL portion should be understood. There are a number of approaches for assessing classification sensitivity from available data. The following sections outline the data used to inform this decision, as well as one method used to complete the analysis.

Data

Campylobacter testing began in the third quarter of 2011. Quarterly results for the young chicken data suggest that the percent positive 30 mL portions is always greater than the percent positive 1 mL portions (Table 1 and Figure 1). Furthermore, the patterns across time for both portions are similar. The quarterly percent positive results for turkey data are consistently less than corresponding young chicken results. In addition, there are some quarters where the percent positive 1 mL portions are slightly greater than the 24 mL portions.

For the purposes of assessing classification performance, *Campylobacter* testing data completed between 2011 and January 2013 were summarized into sample sets. Because data from turkey establishments is limited, this analysis only considers the young chicken testing data.

There were 338 young chicken sets reported during this interval, but many sets were incomplete. There were 215 sets with at least 51 samples. To increase the number of sets in the analysis, any set with 41 or more samples (80% of the target) was included and all results were standardized to a 51-sample set. Therefore, a total of 248 sets from young chicken establishments were considered.

For each set, the number of positive 1 mL and 30 mL samples were summarized. In the aggregate, there were 12,648 samples analyzed of which 6% (791) and 17% (2160) were positive for the 1 mL and 30 mL portions, respectively (Table 2).

Table 1. Percent Positive *Campylobacter* Tests in the PR/HACCP Verification Testing Program by Product Class and Calendar Year (CY) Quarter, 2011-2013.

Year	CY Quarter	Positive 1mL	Negative 1mL	1mL % Positive	Positive 30mL*	Negative 30mL*	30mL % Positive*
Young chicken							
2011	3	85	919	8.5	233	770	23.2
	4	141	1287	9.9	331	1085	23.4
2012	1	141	1492	8.6	391	1241	24.0
	2	145	2408	5.7	389	2165	15.2
	3	206	3161	6.1	539	2828	16.0
	4	264	2954	8.2	519	2699	16.1
2013	1	212	3574	5.6	429	3357	11.3
Turkey							
2011	3	24	362	6.2	15	371	3.9
	4	22	681	3.1	26	672	3.7
2012	1	8	438	1.8	9	437	2.0
	2	9	458	1.9	15	452	3.2
	3	13	510	2.5	14	509	2.7
	4	18	660	2.7	22	656	3.2
2013	1	9	501	1.8	19	491	3.7

*For turkey, the sample portion volume is 24 mL rather than 30 mL.

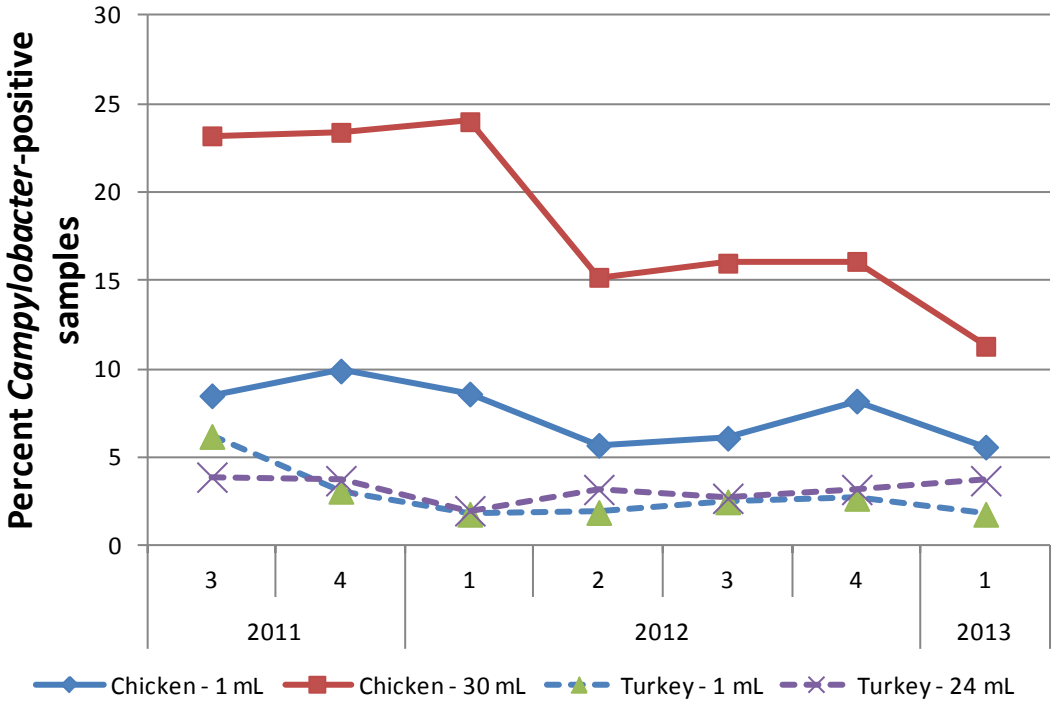

Figure 1. The results summarized in Table 1 are displayed graphically.

Because the current performance standard is based on the 1 mL portion, a set was determined to be from a non-compliant establishment if 9 or more 1 mL portions were positive among 51 samples. In this manner, 22 (~9%) establishments were classified as non-compliant. In the aggregate, there were 1122 samples analyzed from non-compliant establishments of which 36% (398) and 60% (668) were positive for the 1 mL and 30 mL portions, respectively. From the remaining establishments classified as compliant, 11,526 samples were analyzed of which 3% (390) and 13% (1492) were positive for the 1 mL and 30 mL portions, respectively.

Table 2. Aggregated sample results for 1 mL and 30 mL portions from FSIS *Campylobacter* testing; 2011 through Jan 2013. Overall results and results stratified by non-compliant and compliant sets are presented.

Portion	Overall	NonCompliant	Compliant
fraction positive 1 mL	6% (791 / 12,648)	36% (398 / 1122)	3% (390 / 11,526)
fraction positive 30 mL	17% (2160 / 12,648)	60% (668 / 1122)	13% (1492 / 11,526)

Classification sensitivity and specificity

The sensitivity is determined from the non-compliant establishment data. The cumulative binomial probability for test results greater than 8 and 27 positives in 51 samples determines the classification sensitivity of the 1 mL and 30 mL portions.

$$Sensitivity_{1ml} = \sum_{s=9}^{s=51} \binom{51}{s} p_{1ml}{}^{s} (1 - p_{1ml})^{51-s}$$

$$Sensitivity_{30ml} = \sum_{s=28}^{s=51} \binom{51}{s} p_{30ml}{}^{s} (1 - p_{30ml})^{51-s}$$

The specificity is determined from the compliant establishment data. The cumulative binomial probability for test results less than 8 and 27 positives in 51 samples determines the classification specificity of the 1 mL and 30 mL portions.

$$Specificity_{1ml} = \sum_{s=0}^{s=8} \binom{51}{s} \pi_{1ml}{}^{s} (1-\pi_{1ml})^{51-s}$$

$$Specificity_{30ml} = \sum_{s=0}^{s=27} \binom{51}{s} \pi_{30ml}{}^{s} (1-\pi_{30ml})^{51-s}$$

For the above calculations, the binomial probabilities are determined from the estimates in Table 2. Therefore, the probabilities of positive results among non-compliant establishments are $p_{1ml} = 0.36$ and $p_{30ml} = 0.60$. Among compliant establishments, these probabilities are $\pi_{1ml} = 0.03$ and $\pi_{30ml} = 0.13$.

To estimate the sensitivity of using both the 1 mL and 30 mL portions to classify establishments, and assuming that an establishment fails the performance standard if it fails either the 1 mL or the 30 mL sample set criteria, we use the following calculation:

$$Sensitivity_{1ml+30ml} = Sensitivity_{1ml} + Sensitivity_{30ml} - (Sensitivity_{1ml} \times Sensitivity_{30ml})$$

To estimate the specificity of using both the 1 mL and 30 mL portions to classify establishments, we use the following calculation:

$$Specificity_{1ml+30ml} = Specificity_{1ml} \times Specificity_{30ml}$$

Cost of 30 mL portion

The cost of laboratory materials needed to complete analysis of a 30 mL sample was estimated. This cost included a standard unit cost per sample for the qualitative culture set-up. In addition, there were unit costs that pertained to confirmatory analysis needed when the qualitative analysis suggests *Campylobacter* may be present. The confirmatory costs factor into the total laboratory material costs based on the estimated fraction of samples that are positive for the 30 mL portion.

$$Total\ Cost\ per\ 30\ mL\ sample = Total\ Qualitative\ Costs + Total\ Confirmatory\ Costs * Fraction\ of\ 30\ mL\ samples\ positive$$

Results

Using just the 1 mL portion sample set criteria, the estimated classification sensitivity is 99.85% (Table 3). This sensitivity implies that essentially all non-compliant establishments are detected using the 1 mL results solely. If the 30 mL portion was used alone to classify establishments, then the estimated classification sensitivity is 79.33%. By combining the 1 mL and 30 mL

portion results, classification sensitivity is improved to 99.97% (i.e., 0.12% more than the sensitivity just using the 1 mL portion).

The estimated classification specificity of using just the 1 mL portion sample set criteria is 99.99%. This specificity implies that essentially all compliant establishments will be classified as negative using the 1 mL results. If the 30 mL portion alone was used to classify establishments, then the estimated classification specificity is 100%. By combining the 1 mL and 30 mL portion results, classification specificity is reduced to 99.99% (i.e., no change from specificity just using the 1 mL portion).

Table 3. Estimated classification parameters for *Campylobacter* performance standards based on 1 mL, 30 mL and combined 1 mL and 30 mL portions are shown. The estimates for the combined performance standard assume that the results are interpreted in parallel.

Parameter	1mL only	30mL only	1mL+30mL
Classification sensitivity	99.85%	79.33%	99.97%
Classification specificity	99.99%	100.00%	99.99%

With respect to classification sensitivity, a graphical display of the sample set results suggests that the 30 mL results add little value to detection of non-compliant establishments (Figure 2). This figure illustrates that the number of 1 mL and 30 mL positives in a set are positively correlated. For sets with less than 9 positive 1 mL portions, the correlation between 1 mL and 30 mL positives is substantial, but this correlation is less substantial for sets with 9 or more positive 1 mL portions. Furthermore, the sets in the lower right quadrant (defined by the 1 mL and 30 mL sample set criteria) all had 1 mL portion results that exceeded the sample set criterion but did not exceed the 30 mL portion sample set criterion. Therefore, adding the 30 mL portion results does not change the classification of these sets.

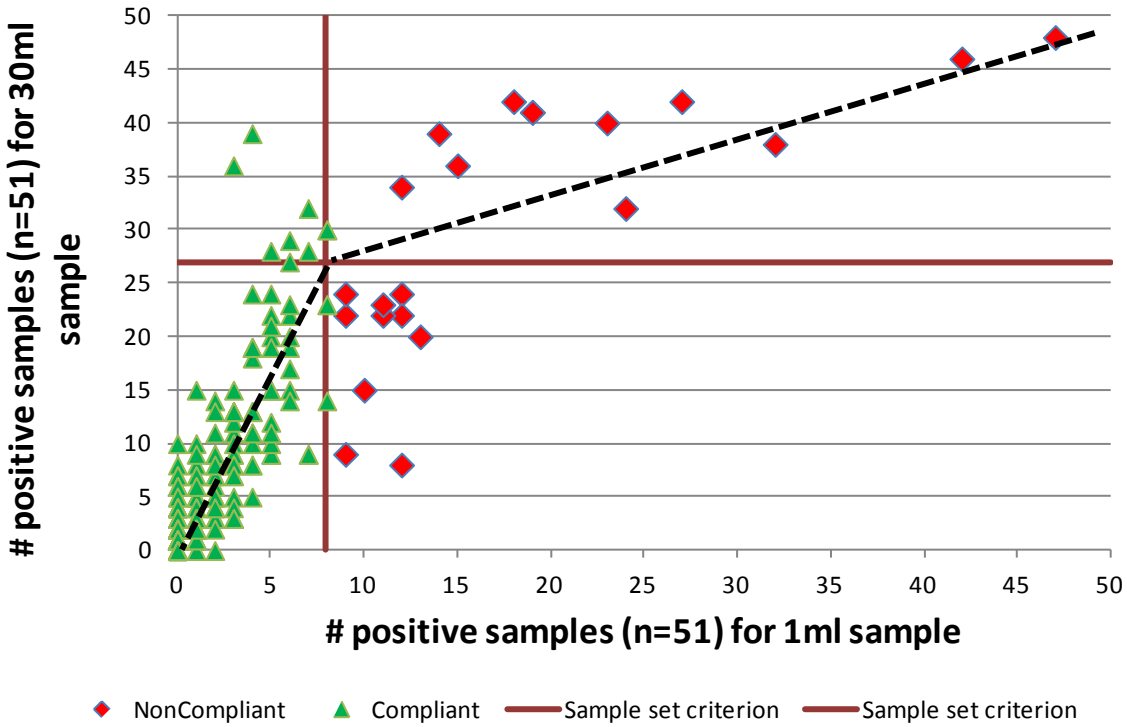

Figure 2. Sample results are plotted for 248 sets. Thresholds show the sample set criteria for the 1 mL (x-axis) and 30 mL (y-axis) portions. The dashed lines suggest the linear relationships between numbers of positive 1 mL and 30 mL samples in sets. A less steep linear relationship is evident for sets in which the number of positive 1 mL samples is greater than 8.

The laboratory material cost to qualitatively analyze a 30 mL sample is $7.94 (Table 4). The material cost to confirm a qualitatively positive sample is $5.65 per such sample. This confirmatory cost will only apply to the estimated 17% of samples positive on the 30 mL portion. Therefore, total cost per sample is $8.90.

Table 4. Laboratory material cost estimates per 30 mL sample analyzed by FSIS.

Campy Qualitative cost	Unit Cost
Bolton Broth	$ 3.00
Campy gas packs	$ 0.44
Cell Flasks	$ 2.00
Pipets	$ 1.00
Cefex Plates	$ 1.00
Spreaders	$ 0.50
SUBTOTAL	**$ 7.94**

Positive Confirmation (applicable to 17%)	Unit
Microscopic	$ 0.50
Latex Agglutination	$ 5.00
Loops	$ 0.15
SUBTOTAL	**$ 5.65**
TOTAL*	**$ 8.90**

* As explained in Methods, total cost = $7.94 + 0.17*$5.65.

Discussion

This analysis suggests that the 1 mL portion for the young chicken performance standard is more sensitive than the 30 mL portion in detecting non-compliant establishments. The 30 mL portion sample set criterion could be reduced to improve its classification sensitivity. The criterion used in this analysis (i.e., 27 positives are acceptable) was developed previously from a baseline survey conducted in 2007-2008. Nevertheless, the estimated classification sensitivity for the 1 mL portion is so large that any improvement is generally trivial. Therefore, this analysis suggests that regardless of the classification sensitivity of the 30 mL portion, the combination of 1 mL and 30 mL portions – interpreted in parallel – will generate an improvement in classification sensitivity that is negligible relative to a performance standard based on the 1 mL portion by itself.

If the estimated classification sensitivity for the 1 mL portion is accurate (i.e., 99.85%) and there are roughly 20 non-compliant establishments among approximately 200 young chicken establishments, then the performance standard based on the 1 mL portion is expected to detect 19.97 non-compliant establishments. The estimated classification sensitivity for the combined 1 mL and 30 mL portions (i.e., 99.97%) expects to detect 19.99 non-compliant

establishments. Such an improvement does not warrant the cost of including the 30 mL portion.

The approach used here aggregates testing data to determine the probability of a positive 1 mL or 30 mL sample given that an establishment is compliant or non-compliant. An alternative approach was to fit distributions to the two populations of data and examine the effect of sampling randomly across each distribution. This approach assumes a carcass with a randomly chosen concentration of *Campylobacter* is rinsed. A 1 mL and 30 mL portion is cultured assuming Poisson and binomial processes are applicable. Following Monte Carlo simulation of this sampling process, the classification sensitivities were nearly equivalent to those described above.

Another alternative approach used re-sampling of the set results to estimate the classification sensitivity. In this case, each of the 22 non-compliant sets were sampled based on the fraction of positive samples in each set for both 1 mL and 30 mL portions. A set with more positives than the sample set criterion for each portion was considered detected. Classification sensitivity was estimated as the sum of detected sets divided by 22. This sensitivity was 88% for the 1 mL portion, 57% for the 30 mL portion and 89% for the combined 1 mL and 30 mL portions. Therefore, the improvement in sensitivity from including the 30 mL portion was not substantial using this approach.

This analysis did not explore other approaches such as so-called “no-gold standard” methods because initial analyses found inconsistent results that could not be improved without more data or dramatic assumptions. No-gold standard methods are attractive because there rarely are definitive methods for determining the true status of members in a population. These methods seek to find the most likely parameters (e.g., sensitivity, specificity, prevalence) that might explain the available data. Nevertheless, the limited number of sets considered in this analysis does not enable robust solutions to the no-gold standard algorithm.

In general, this analysis suggests that any improvement in classification sensitivity that might occur by using the 30 mL portion results would be slight. Furthermore, because the current performance standard is based on just the 1 mL portion, any application of the 30 mL results would require a formal revision of the current performance standard. Additionally, the laboratory resources expended for 30 mL analyses should be used more effectively in other sampling efforts. For these reasons, FSIS has decided to suspend the 30 mL qualitative analysis.