Analysis of 2002 FSIS Bovine AMR Products Survey Results

Food Safety and Inspection Service
United States Department ofAgriculture
February 2003

Summary of results:

1) Some of the establishments (4 of 34 or 12%) from which samples were collected were able to consistently produce advanced meat/bone separation and meat recovery (AMR) products derived from beef vertebrae, that did not contain unacceptable nervous tissues (spinal cord (SC) or dorsal spinal nerve root ganglia (DRG)). This result was based on not detecting unacceptable nervous tissues (UNT) in 6 or more samples taken randomly over about a seven-month period. On the other hand, some establishments' samples were nearly all positive for UNT, suggesting that much of the AMR product derived from beef vertebrae in these establishments would contain unacceptable nervous tissues.
2) For the study, approximately 35% of the finished AMR product samples had unacceptable nervous tissues detected; 29% of the samples had SC tissue detected; and 10% had DRG tissue detected. However, the percentages of positive samples were significantly different for different periods of the survey. For the last third of the survey, the percentage of positive (for any unacceptable nervous tissues) post-desinewing samples was about 44%, versus 31% for the first two-thirds of the survey.
3) The presence of SC and DRG tissues in these samples were not significantly correlated, suggesting that there may be different factors that cause their occurrence in this product.
4) The type of bones used in processing may contribute to the likelihood of unacceptable nervous tissues being present in the finished product; however, regardless of the type of bone used, product can be produced with a low likelihood of the presence of unacceptable nervous tissues.
5) On histologic examination, nearly all (96.5%) of the AMR product samples were found to contain hematopoietic cells, which indicates the presence of bone marrow. For the presence of hematopoietic cells, there was no significant difference between the two different basic designs (two manufacturers) of AMR machines that were used to produce AMR products.
6) The average of the calcium values for post-desinewing samples was 91.7 mg/100g; the highest value was 159 mg/100g. For establishments for which all samples were negative for SC and DRG, the average of the calcium levels was 77.0 mg/100g.
7) The average of the excess iron (iron – 0.138(1.1) protein) measurements was 2.81 mg/100g. For establishments for which all samples were negative for SC and with one exception, DRG, the average of the excess iron levels was 2.20 mg/100g, where, from 27 results, 3 were greater than 3.1 mg/100g.
8) Excess iron and calcium were positively correlated, suggesting a common set of factors that influence their levels. In general, higher levels of these variables were associated with a higher likelihood of UNT being in the product, suggesting that these variables reflect processing parameters that might be related to the likelihood of unacceptable nervous tissues being present in the product. However, there were significant expectations to general trends.
9) Relationships between excess iron and calcium and machine operating parameters and product type were equivocal.

10) Protein levels decreased, on average, by about 3%, as a result of the desinewing step. Since the protein that was being removed was most likely iron-deficient, this finding suggests that the excess iron measure be adjusted to account for this loss of protein.
11) Thirteen percent of the post-desinewing samples that were negative for unacceptable nervous tissues had positive finding for the matched pre-desinewing sample. This suggests that the quantity of sample used for determining the presence of unacceptable nervous tissues may be too small, or composite samples were not sufficiently homogeneous in the distribution of UNT.
12) An ELISA (enzyme-linked immunosorbent assay) method was compared with the direct FSIS histologic method (immunohistochemical microscopic examination) of detecting UNT. While there was a correlation between the ELISA results and the findings of UNT in samples by the direct method, there was also a significant number of false negative findings. Using a cutoff value for determining positive samples that provides an approximate 25% false positive rate, the false negative rate on the all-positive samples, as determined by the direct method, was about 30%; for samples from establishments for which most of their samples were positive, the false negative rate was about 20%. Of further concern, though, is that false negative rates seemed to be related to the establishment from which the samples were taken; no discernable "reason" was found to help explain possible causes for this establishment-specific dependency.
13) The Agency observed a wide variation in the physical character of AMR products of the study. The AMR-product variation in texture and consistency ranged from similar to ground beef derived from hand-deboned product to similar to thick tomato soup or sauce. However, these observations were not related to analytical results obtained on samples.
14) Twenty-three establishments that had positive SC samples submitted 68 follow-up AMR product samples for verification (which were not included in the survey results). Most of the follow-up samples were collected after the survey. All, but two, establishments had no more than 3 follow-up samples; the two exceptional establishments had 12 and 19 samples. For these two establishments, the percentages of positive survey and follow-up samples were nearly the same, of about 40-50%. For the other 21 establishments the percentages of positive follow-up samples were generally less than the percentages for the survey samples. Overall, approximately 1/3 (22 or 67) of the follow-up samples was positive for SC tissue.

Analysis of results of survey of AMR products derived from bovine vertebrae.

The purpose of this survey was to characterize AMR products derived from beef vertebrae, in particular, to establish a baseline for the prevalence of unacceptable nervous tissues (UNT), consisting either of spinal cord (SC) and dorsal spinal nerve root ganglia (DRG), in AMR products. These two types of tissues have been identified in the Harvard BSE Risk Assessment (2002) as specific risk materials for bovine spongiform encephalopathy (BSE). The results of this survey are to be used for a regulatory impact analysis, as required by Congress.

A primary question was what processing factors affect the likelihood of SC tissue and DRG tissue being found in samples of AMR products. In addition, FSIS was concerned about the presence of excessive iron (exFe) in the AMR products, as evidence of the presence of more than negligible amounts of bone marrow tissue (i.e., hematopoietic cells), and calcium, as a measure of minutely sized bone particles that are not normally seen in hand-deboned beef. While the measurement of heme iron was a known alternative analysis method for determining the amount of bone marrow tissue (i.e., hematopoietic cells) in AMR products, the Agency believed that the methodology needed further development to standardize the testing methodology, and to lower the cost of the testing procedure. Few commercial laboratories were known to be accredited to analyze for the amount of heme iron in AMR products. In a proposed regulation (1998), FSIS proposed requirements regarding the level of iron in AMR products that was considered to be excess over what would exist in the corresponding hand-deboned products, as a means of evaluating whether more than negligible bone marrow exists in the AMR products. In Attachment A is an analysis of iron levels in the hand deboned data from the 1996 FSIS AMR survey. Based on this data, it was derived that an excess iron measurement of 2.8 mg/100g or more for a single sample would imply that the sample is, with more than 99.9% confidence, not from a hand-deboned product processed under good manufacturing practices. This conclusion is based on an analysis of variance of the excess iron levels for the hand-deboned product collected in the 1996 survey and took into account the analytical measurement error associated with iron and protein. Thus, the identification of the factors that influence the levels of excessive iron and calcium are also desired. Variables, for which information was collected, are: machine pressure and dwell times during processing; types of non-vertebra bones (rib, pelvis, flat, or any other types); type of vertebra bones (neck or back only); and food chemistry (iron, protein and calcium).

Data on AMR products before and after desinewing were collected. The regulatory (present and proposed) requirements apply to the AMR products after desinewing; however, FSIS is interested in the relationships of the products before and after desinewing.

Thirty-four establishments were identified as producing AMR products derived from beef vertebrae. The sample design called for 6 pairs of samples to be collected, one sample before and one after desinewing for each AMR machine operating in the establishment. The times of sampling were from the middle of January 2002 to the end of August 2002, with the last sample being received for analysis on September 18, 2002. However, no samples were collected during the last week of February and the month of

March. If an establishment was not producing AMR products derived from beef vertebrae at the time that the sample was requested, an additional sample request form was sent to that establishment. The follow-up sample request form apparently was the cause of there being more than 6 samples in some establishments; samples for the original form and also for the additional form were collected. Furthermore, the planned design was not completed because of test and labor resource constraints. Toward the end of the survey period, the designated pre-desinewing samples were not collected. In addition, for some of the later samples, food chemistry analyses were not done. Hence the distribution of samples over the establishments is not balanced.

Pathological examination for the presence of SC and DRG tissues is time consuming, relatively expensive and requires an expert technician. These burdens motivated the development of an ELISA (enzyme-linked immunosorbent assay) procedure that would be relatively quick and simple to conduct.

Since the data was collected from establishments over a 7-month period, it is not possible to identify causal relationships of variables, as would or might be the case in a controlled study. Identified correlations between two variables can occur as a result of unknown causal variables affecting both. Furthermore, variables that are causal, in the sense that changes in values of a variable x affect the values of y, given everything else being equal or held constant, may not be detected because of the existence of other, unknown, variables that affect values of y, thus masking the causal relationship. Among establishments, there are many variables that could affect the values of variables of interest, thus masking relationships that exist, or creating correlations of variables that are not truly related in a causal fashion. In an attempt to eliminate some of the effects of unknown establishment-specific variables, statistical analyses of relationships are performed within establishments and summaries of these results are used to determine the existence of possible causal relationships. But even within establishments, there exist variables that could affect the results and possibly mask true relationships or create non-causal correlations. In addition, in an attempt to eliminate the effect of possible deviant or extreme results, non-parametric statistical tests are relied upon for evaluating the strength of relationships. Thus the statistical analysis consists of examinations of the consistency of relationships within establishments, and among establishments. Statistical significant levels that are quoted are two-sided. Statistical analyses were performed using PC-SAS®, release 8.0.

An attachment presents an analysis of data from the FSIS, 1996 AMR survey of AMR product derived from beef neckbones, where samples were collected from 7 establishments producing AMR product and two establishments producing hand-deboned product. The results of this analysis can be used to derive an excess iron performance standard for AMR product.

General overview of results

Before presenting a more detailed analysis of the relationships of the variables that were studied in this survey, summary results, by establishment, are presented in Table 1. The prevalences of spinal cord (SC) or dorsal spinal nerve root ganglia (DRG) tissues in all 394 samples (before and after desinewer), by establishments, are presented in Table

1a. The prevalences of SC or DRG tissues in the 256 final AMR product samples are presented in Table 1b. Included in Table 2 are: the fraction of samples that were positive for either spinal cord (SC) or dorsal spinal nerve root ganglia (DRG) tissues, designated by "unacceptable nervous tissues" (UNT), as well as the fraction that were positive for each one; the mean of the pressures and dwell times used for the collected samples; and the median calcium and excess iron result. The exact formula for the excess iron will be given below. The determination of the fraction of samples that were positive with respect to the specific type of nervous tissue or any of them as follows: if any of the matched pre- or post-desinewing samples were positive for a nervous tissue, then the sample (considered as the pair) was counted as a positive. The fraction then is just the ratio of the number of positive samples divided by the number of samples, where in the case the sample consists of a matched pair it is counted as one sample. The fraction is computed in this way because it is assumed that if UNT occurs in the pre-desinewed product, then it would occur in the post-desinewed product, and that a negative result on the sample from the latter product would occur because of sampling error. As discussed below, about 13% of the samples that were positive for the pre-desinewed sample were negative for the matched post-desinewed sample.

The median values for the food chemistry variables are the results on the post-desinewing sample. The establishment identification numbers (id) were determined by assigning random numbers from 1 to 34 to the establishments. The observations in Table 1 are ordered from the smallest to the largest fractions of UNT positive samples, and, within groups that had the same fraction, by the median excess iron result.

Table 1: Summary of results for selected variables, by establishment. Observations are ordered in ascending fraction of samples detected with unacceptable nervous tissues.

Estab-lish-ment Id	Mach-ine Type	Number Samples	Mean Pressure (psi)	Mean Dwell Time (sec)	Frac. UNT +	Frac. SC +	Frac. DRG +	Number of Food Chemistry Samples	Median Calcium (mg/100g)	Median Excess Iron (mg/100g)
6	H	6	2902	4.2	0.00	0.00	0.00	5	54.0	2.00
21	H	13	3300	25.5	0.00	0.00	0.00	8	74.0	2.56
15	P	7	1658	1.1	0.00	0.00	0.00	5	95.0	1.54
13	H	6	2757	3.0	0.00	0.00	0.00	4	109.5	2.91
4	H	11	2233	4.2	0.09	0.00	0.09	5	58.0	0.99
26	H	7	2443	1.0	0.14	0.00	0.14	5	63.0	1.21
23	H	7	2621	3.0	0.14	0.14	0.00	5	97.0	2.62
10	P	7	1223	1.0	0.14	0.14	0.00	4	116.0	2.21
33	H	12	2853	4.0	0.17	0.00	0.17	8	84.0	2.90
27	H	6	2921	7.1	0.17	0.00	0.17	2	88.5	3.91
25	H	5	3135	3.5	0.20	0.20	0.00	2	116.0	3.55
34	H	12	3429	1.0	0.25	0.17	0.08	7	88.0	3.46
24	P	6	2322	1.8	0.33	0.00	0.33	5	105.0	3.48
17	H	6	2902	3.9	0.33	0.17	0.17	5	93.0	4.25
29	P	6	1451	0.1	0.33	0.33	0.00	4	50.0	0.93
28	H	6	2500	5.0	0.33	0.33	0.00	5	66.0	1.40
3	H	6	2757	2.8	0.33	0.33	0.00	4	102.0	3.68
31	P	10	2345	2.3	0.40	0.20	0.20	5	121.0	2.57
9	P	7	2467	4.0	0.43	0.29	0.29	5	132.0	3.71
12	H	7	3213	1.0	0.43	0.43	0.14	4	90.5	4.69
18	H	11	3004	4.6	0.45	0.45	0.00	8	74.0	1.48
20	P	6	2878	1.0	0.50	0.17	0.33	5	91.0	2.54
16	H	6	2983	10.0	0.50	0.50	0.00	4	97.0	4.15
22	H	6	3000	1.0	0.50	0.50	0.00	5	100.0	2.61
5	P	6	2909	0.5	0.67	0.33	0.50	4	87.0	2.41
8	H	6	2612	5.0	0.67	0.67	0.00	5	89.0	2.99
19	H	6	2868	2.8	0.67	0.67	0.00	4	99.5	3.03
2	H	6	2394	2.0	0.67	0.67	0.33	5	118.0	3.08
14	H	6	2612	1.6	0.67	0.67	0.00	4	123.5	4.86
11	H	5	2840	1.0	0.80	0.80	0.00	2	104.0	3.30
32	H	16	3376	3.9	0.81	0.50	0.31	12	96.0	2.92
7	H	6	2902	1.0	0.83	0.83	0.17	4	119.0	3.94
1	H	13	3104	1.0	0.85	0.85	0.08	8	94.0	1.90
30	H	7	2600	10.0	0.86	0.86	0.43	3	65.0	1.79

An examination of the results presented in Table 1 reveals that: 1) there is virtually no correlation between the fraction of SC and DRG positive samples; 2) there are moderate correlations - determined by significance levels ranging from about 0.10 to 0.25 for Spearman correlations of approximately 0.2 in absolute value - between the fraction of positive UNT samples with the median excess iron, calcium, and the mean machine

pressure and dwell times; 3) there is virtually no correlation of the fraction of positive DRG samples and the variables identified in 2), and 4) there are high positive correlations (P- value < 0.05) of the median excess iron with median calcium and machine pressure. The lack of correlation of the fraction of SC and DRG tissue findings could imply that distinct factors contribute to the likelihoods of SC and DRG tissue.

Samples were classified regarding the likelihood that they would be positive with respect to UNT by considering the fraction of samples detected to have UNT for the establishment from which they were sampled. As depicted in Table 1, five classifications were made: 1) establishments for which less than 10% of the samples were positive for UNT; 2) establishments for which the fraction UNT positive samples was greater than 0.10, but less than or equal 0.25; 3) establishments for which the fraction UNT positive samples is greater than 0.25, but less than or equal 0.50; 4) establishments for which the fraction UNT positive samples is greater than 0.50 but less than 0.8; and 5) establishments for which the fraction of UNT samples is greater than or equal to 0.8. For class 1, there are 5 establishments; for class 2, 7 establishments; for class 3, 12 establishments; for class 4, 5 establishments; and for class 5, 5 establishments. To distinguish positive UNT samples from negative ones, for a sample, a value of ½ was added to the class level if the sample were positive for UNT.

Table 1a: Advanced Meat/Bone Separation and Recovery (AMR) Study:
Laboratory Results of Beef AMR Product Samples*
Testing for the Presence of Spinal Cord (SC)** and DRG*** Tissues
Establishments Producing AMR Product
(Based on test results of 394 samples of AMR Products Produced from January to September 2002)

Establishments where SC detected		Establishments where SC not detected, yet DRG detected	Establishments where SC or DRG not detected	Total Establishments Tested
25 (73.5%)		5 (14.7%)	4 (11.8%)	34 (100%)
Detected 1 time only	Detected 2 or more times			
6 (17.6%)	19 (55.9%)	9 (26.5%)		34 (100%)

Footnotes: * AMR product (before and after desinewer) derived from beef vertebrae of 34 establishments
** SC – spinal cord tissue – a central nervous system (CNS) tissue
*** DRG - dorsal spinal nerve root ganglia

Source: FSIS Eastern Laboratory, Athens, Georgia
AMR Products Survey 2002

Table 1b: Advanced Meat/Bone Separation and Recovery (AMR) Study:
Laboratory Results of 256 Final Beef AMR Product Samples*

Testing for the Presence of Spinal Cord (SC)** and DRG*** Tissues

(AMR Products Produced from January to September 2002)

SC (only) Detected	DRG (only) Detected	SC & DRG Detected	SC or DRG Not Detected	Total Samples Tested
64 (25%)	16 (6%)	9 (4%)	167 (65%)	256 (100%)

Footnotes: * AMR product (after desinewer) derived from beef vertebrae of 34 establishments
** SC – spinal cord tissue – a central nervous system (CNS) tissue
*** DRG - dorsal spinal nerve root ganglia

Source: FSIS Eastern Laboratory, Athens, Georgia
AMR Products Survey 2002

Analysis of beef AMR product characteristics after desinewing.

Nervous tissues and hematopoietic cells

Since the regulation is to apply to AMR products after desinewing, and the number of samples after desinewing is nearly balanced with respect to machines, the following analysis is on the post-desinewing samples. There were 256 results from samples collected after desinewing. Of these, there were 28.9% of them detected having SC tissue, 9.8% having DRG tissue, and 96.5% having hematopoietic cells (i.e., bone marrow). The percentage of samples positive for any unacceptable nervous tissue (UNT) was 35.2%. These numbers are consistent with that premise that occurrences of SC and DRG tissues are not correlated, since, if it is assumed that the occurrences of SC and DRG tissues are independent, then there would be an expected 35.9% of the samples that would be positive for either tissue. Within-establishment correlations of the occurrences of SC and DRG were computed for the post-desinewing AMR products derived from beef vertebrae. Of the 9 within-establishment correlations that were possible to compute, 6 were positive and 3 were negative, which is not statistically significant (P- value of about 0.5).

A further analysis of the data revealed that over the time of the survey the percentage positive samples increased. For the first two months of the survey (Jan-Feb. 2002) 31 samples were collected. The sampling was interrupted during the month of March, and continued from April to the end of August. The samples were divided into 6 time

periods: the first period consisting of samples received by the laboratory during January and February; and the other 5 periods were assigned by dividing the samples into periods of approximately equal numbers, assuring that the samples arriving at the lab on the same dates were in the same period (the actual collection dates were not recorded for all samples). The following table presents the fraction of positive post-desinewing samples for any unacceptable nervous tissues (UNT), by time period and likelihood of positive samples, defined from Table 1. The results presented in Table 2 clearly show that the percentages of positive samples generally increased over time for all likelihood classes. The latter third of the survey generally consisted of samples collected during the summer; thus, it was possible that environment, specifically temperature, may have had an impact on the likelihood of samples being positive for UNT. However, there is no known scientific basis for ambient temperatures of summer to be associated with an increased likelihood of positive UNT results. All AMR samples were to be chilled (and not frozen) after collection, and the chilled samples were to be shipped with sufficient frozen "chill-packs" to have the AMR product sample arrive overnight at the FSIS Eastern Laboratory (via a commercial courier) in a chilled condition. Temperature damaged putrefied samples were to be discarded and not tested. Only a few AMR samples arrived in a putrefied state - in such a condition that they could not be tested.

Table 2: Fractions of post-desinewing samples that were UNT positive, by likelihood classes of positive samples and time periods of survey. The lower time period designation means the samples were analyzed earlier in the survey.

	Likelihood Class											
	1		2		3		4		5		All	
	N	Frac. UNT	N	Frac. UNT	N	Frac. UNT	N	Frac. UNT	N	Frac. UNT	N	Frac. UNT
Period												
1	6	0.000	5	0.000	11	0.273	4	0.500	5	0.600	31	0.258
2	7	0.000	9	0.222	20	0.300	6	0.500	6	0.500	48	0.292
3	7	0.000	9	0.222	14	0.143	10	0.700	8	0.750	48	0.354
4	7	0.000	11	0.091	14	0.429	2	0.700	8	0.625	42	0.310
5	10	0.000	10	0.300	10	0.400	5	1.000	11	0.818	46	0.457
6	6	0.167	11	0.091	14	0.429	3	0.667	7	1.000	41	0.415
All	43	0.023	55	0.164	83	0.325	30	0.667	45	0.733	256	0.352

A further analysis (not presented here) indicated that the pattern seen above in Table 2 did not hold for the presence of DRG tissue, and (consequently) did hold for the presence of SC tissue. The existence of a time factor effect would not have a significant impact on the conclusions, regarding information collected for establishments, insofar as the samples, within establishments, over time are nearly balance; the variation associated with this variable is not large compared to that associated with the establishment facto;

and the machine settings (pressure and dwell time) are not significantly correlated with time, though for the dwell time there were some establishments for which the dwell times were higher in the latter third of the survey. Because food chemistry results were not collected for the latter third of the survey and the percentage of positive samples were higher in the latter third of the survey, the estimated relationship of the food chemistry results and the likelihood of positive samples derived from these data may not be accurate. The time effect is included in the subsequent analyses by defining a variable to be equal 0 when the sample is within the first two-thirds of the survey, and equal 1 when the sample is in the last third of the survey.

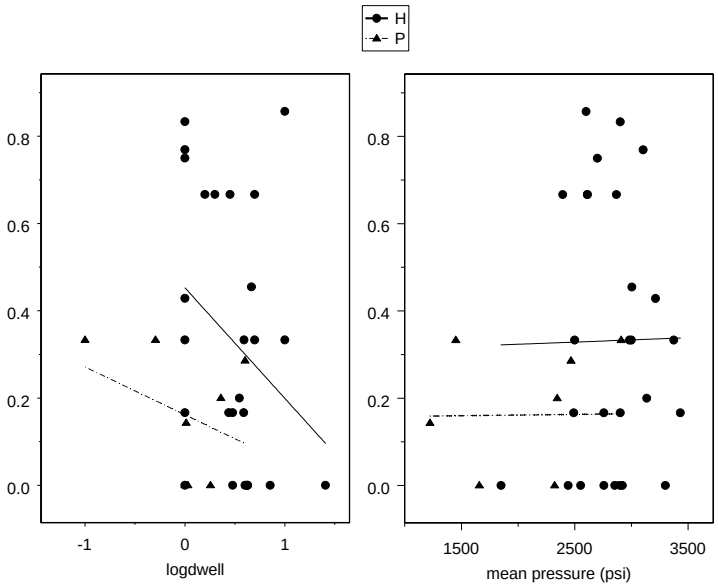
Two types of AMR machine, Protocon (P) and Hydrosep (H), were identified in the 34 establishments producing the AMR product: 26 were using Hydrosep machines and 8 were using Protocon machines. The means of the establishment-specific percentage of positive UNT samples were: for the Hydrosep, 37.7%, and, for the Protocon, 30.5%. An analysis of variance on the variable with value equal 1 when UNT was detected, and 0 otherwise, assuming establishment as a random factor, and including a time effect distinguishing samples within the first two-thirds of the survey from the latter third, did not indicate significant machine type effects. In addition, machine effects, within establishments, were not statistically significant for the H machines, but marginally so for the P machines (P- value = 0.11).

The difference of the percentages of UNT positive samples for the two machine types is not statistically significant. However, the pattern of the differences of the percentages of SC and DRG positive samples for the two machines are dissimilar. For the Hydrosep machines, the establishment-specific mean of SC positive samples is 34.5% compared to 16.2% for the Protocon. This difference is significant at about the 0.19 level, based on analysis of variance on the raw results, treating establishment as a random factor, and accounting for the time effect. The Wilcoxon test statistic (on the establishment-specific mean values) had a significance level of 0.16. For the presence of the DRG tissue, the direction of the machine effect is reversed: For the Hydrosep machines, the DRG establishment specific percentage is 7.6% compared to 16.1% for the Protocon machines (P- value = 0.13 for the ANOVA, and 0.30 for the Wilcoxon). The difference of the differences is significant at the 0.05 level, based on an ANOVA, with establishment as a random factor and the time effect. This finding reinforces the premise that different factors are contributing to the presence of the two types of unacceptable nervous tissue. For the presence of hematopoietic cells (i.e., bone marrow), there was no significant difference, where the percentage of positive samples is about 97% for the H machine, and 94% for the P machine.

An important question is whether there is a relation between the percentage of SC or DRG tissue positive samples and values of operating parameters for the machine: pressure and dwell time. One establishment had two machines that had pressures that were more than slightly different, so in the following analysis, these machines were considered as two establishments. The averages of these for a given machine within an establishment were assigned to the machine and to samples for which the information was missing. The averages of the machine pressures and dwell times for the Hydrosep machines are slightly higher than those for the Protocon machines. The higher pressures

and dwell times might provide a "possible reason" for the higher fraction of positive cns tissues and the lower fraction of positive DRG tissues for the Hydrosep machine samples. Figure 1 presents a scatterplot of the establishment-specific fraction of samples with SC tissue versus the $\log_{10}$ of the mean dwell time and the mean pressure, with different symbols indicating the type of machine. Also the linear regression lines for the two machines are depicted.

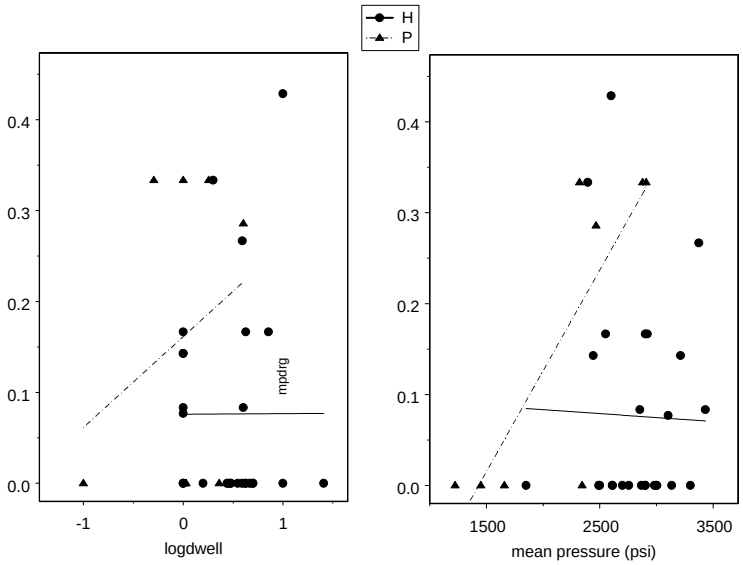
Figure 1: Scatterplot of fraction of establishment specific positive SC tissue samples versus $\log_{10}$ mean dwell time (s) (left graph) and mean pressure, (right graph), together with linear regression lines, by type of machine. The y-axes are the fraction of samples that are positive for SC tissue; the x-axes are labeled.



For the mean pressure, the lines are nearly parallel to the x-axis indicating the lack of correlation of the likelihood of SC tissue and machine pressure. For the dwell time, the correlation appears to be negative: higher likelihoods of SC tissues are associated with lower dwell times.

A similar examination of the effects of pressure and dwell time on the presence of DRG yielded no consistent results. Figure 2 presents scatterplots and linear regression of lines of the establishment-specific fraction of samples detected positive for DRG tissue versus the $\log_{10}$ of the dwell time and the machine pressure. While a positive correlation is seen for the Protocon machines, none is seen for the Hydrosep machines.

Figure 2: Scatterplot of fraction of establishment specific positive DRG tissue samples versus $\log_{10}$ mean dwell time (s) (left graph) and mean pressure, (right graph), together with linear regression lines, by type of machine. The y-axes are the fraction of samples that are positive for DRG tissue; the x-axes are labeled.



To examine further the relationship of these variables with the presence of unacceptable nervous tissue, relationships within the establishments were examined. To help decide the significance of any pattern, the following approach was used. Assume random variables, x and y, measured on a sample, where x takes on values of 0 (for a negative sample) and 1 (for a positive sample) and where y is a variable that can take on any numerical value. To determine if there is a correlation between x and y, the average of the ranks of the values of y for positive samples is compared to the average rank, (n+1)/2, where n is the number of samples. Specifically, the statistic computed for each establishment is

$$d_k = (\bar{r}_{m_k} - \frac{n_k + 1}{2})\delta_k \qquad (1)$$

where the index k specifies an establishment, m_k is the number of positive samples out of n_k samples of the establishment, $\delta_k = m_k/n_k$ is the fraction of positive samples, $\bar{r}_{m_k}$ is the average rank of the of the m_k positive samples among the n_k values of y (the lowest value being assigned the lowest rank of 1, and the next lowest a rank value of 2, and so forth, and ties are set equal to the average rank). This statistic is symmetric about $\delta = ½$. Note that d_k is zero when $\delta_k = 0$ or 1, or when all the rank scores of y are the same. The variance of d_k, when the null hypothesis of zero correlation is true, assuming no ties, is

$$\text{var}(d_k) = (n_k + 1)\delta_k(1 - \delta_k)/12. \qquad (2)$$

The test statistic computed is

$$T = \sum_{k=1}^{K} n_k d_k \qquad (3)$$

where K is the number of establishments. The variance of T is

$$\text{var}(T) = \sum_{k=1}^{K} n_k^2 \, \text{var}(d_k). \qquad (4)$$

Observations for which there were no differences in the rank values were deleted. Hence, to gauge the significance of the value of T for testing whether there is a relationship (rejecting the null hypothesis of no relationship), a z-value is computed,

$$Z = \frac{T}{\sqrt{\text{var}(T)}} \qquad (5)$$

which is compared to the percentiles of the normal distribution.

None of the 6 values of Z, of Eq. 5, for determining the significance of comparisons: the presence of UNT, SC, and DRG with dwell times and machine pressure, were statistically significant at the two-sided 0.10 level or less. The six values are: Z(UNT, dwell) = 0.667; Z(SC, dwell) = 1.30; Z(DRG, dwell)= -0.873; Z(UNT, press)= 0.287; Z(SC, press)= 1.53; Z(DRG, press) = -1.61. The number of establishments for which there were non-zero d_k's was small: for the DRG and dwell time, only 4 values were non-zero, and all were negative. The sign for the correlation between SC and dwell time was positive, which was not the same as the sign for the between-establishment correlation of these seen above (Fig. 1). Of some interest though is the difference of the signs of the Z-values for the SC and DRG, again reaffirming the possibility of different factors affecting the presence of SC and DRG tissue.

Product type used in processing

One of the factors that might affect the likelihood of SC or DRG tissue is the type of product used in processing, in regards to the vertebrae or other non-vertebrae bones that are in the pre-processed product. Information was collected on whether rib or pelvis or other types of non-vertebrae bones were used; another field recorded whether neck or back vertebrae were used, as opposed to just the whole vertebrae (which might have included neck or back vertebrae). For each establishment, the fraction of samples from product that were processed using neck or back vertebrae only, and included other non-vertebrae bones was computed. For the most part, within establishments, added non-vertebrae bones were not used (less than 20% of the samples) or most of the time added non-vertebrae bones were used (greater than 60%).

For designated neck or back vertebrae, one clear pattern was found: all high likelihood of UNT establishments (classes 4 or 5) had very few samples that indicated only neck or back vertebrae were used. For added non-vertebrae bones, a similar pattern can be seen, where only 2 of the 10 high likelihood establishments indicated no added non-vertebrae

bone usage. However, there were a few establishments with high fractions of added non-vertebrae bones and non-specified vertebrae that had a low fraction of samples with UNT (likelihood class 1 or 2), thus, no general causal relationship involving these variables can be established from these data. The number of samples for vertebrae designated as only neck or back is small, thus it is difficult to make conclusions regarding the effect of only using neck vertebra bones. Further the samples not designated as using only neck or back vertebrae could, in any case, consist mostly of neck or back vertebrae, adding to the difficulty of inferring possible causal relationships involving this variable. Table 3 gives the fraction of UNT positive samples by machine type, vertebrae type and whether other non-vertebrae bones were added or not.

Physical characteristic of hand-deboned meat versus AMR products

The Agency observed a wide variation in the physical character of AMR products of the study. The AMR-product variation ranged from similar in texture and consistency to ground beef derived from hand-deboned product to similar in texture and consistency to thick tomato soup or sauce. Thus, a measure of viscosity is one example of a parameter that would better define AMR products. The measurement of viscosity or other physical characteristics (i.e., slump factor) would assure that AMR products are closer in physical appearance to ground beef.

Food Chemistry

Protein (p %), iron (Fe mg/100g) and calcium (Ca mg/100g) measurements were made on 170 of the 256 samples. For each sample, a single determination was made, which may in itself, increase the variability of the iron results since the repeatability of iron measurements is about 0.16 mg/100g (see attachment A). Thus, a 95% confidence interval for the true level in the sample based on a single analysis would have a range of 0.64 mg/100g. However, this estimate was based on an analysis of data for which a handful of "outlier" results were deleted. In practice, it was recommended that duplicate analysis be made (see attachment), and the mean value used as an estimate of the concentration of iron in the product, provided the individual results are not "too" far apart.

For each sample, the iron to protein ratio (ipr) was computed and an excess iron measurement was determined as: ex Fe = Fe – (0.138)(1.1)p. The factor 0.138 is the ipr value determined from ground product collected in the FSIS AMR 1996 survey, representing product that was used for AMR processing during the survey. The factor 1.1 is an adjustment (or fudge) factor to account for the loss of iron-depleted protein during desinewing (See the attachment and the next section).

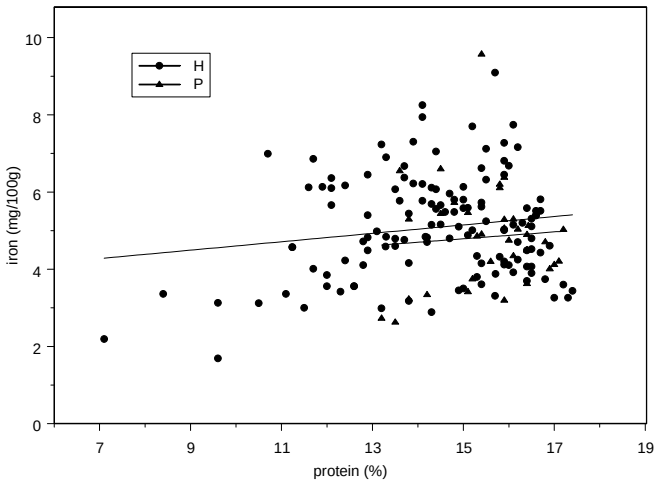
Table 3: Summary of fraction of samples with CNS tissues, added non-vertebrae bones and vertebrae types, machine type, and level of dwell time (defined in table and in text).

Likeli-hood UNT Class	Machine Type	Estab-lish-ment id	Fraction UNT	Fraction non-specified vertebrae	Fraction added non-vertebrae bones	Mean pressure (psi)	Mean dwell time (sec)
1	H	17	0.00	0.20	0.00	1850	4.24
1	H	17	0.17	0.33	0.17	2552	4.23
1	H	32	0.00	1.00	0.00	2757	3.00
1	H	16	0.00	1.00	0.08	3300	25.54
1	H	33	0.00	1.00	1.00	2902	4.17
1	P	28	0.00	0.57	0.00	1658	1.07
2	H	3	0.17	0.33	0.17	2491	3.00
2	H	8	0.14	0.71	0.86	2443	1.00
2	H	26	0.08	0.92	0.92	2853	4.00
2	H	6	0.20	0.40	1.00	3135	3.52
2	H	34	0.25	1.00	1.00	3429	1.00
2	H	30	0.17	1.00	1.00	2921	7.14
2	P	15	0.14	0.00	0.00	1223	1.03
3	H	25	0.33	1.00	0.00	3000	1.00
3	H	5	0.17	0.50	0.17	2902	3.88
3	H	18	0.17	1.00	1.00	2757	2.75
3	H	7	0.33	1.00	1.00	2983	10.00
3	H	11	0.33	0.83	0.00	2500	5.00
3	H	22	0.43	1.00	1.00	3213	1.00
3	H	9	0.45	1.00	0.91	3004	4.65
3	P	14	0.33	0.67	1.00	2878	1.00
3	P	23	0.20	0.00	0.00	2345	2.29
3	P	27	0.33	0.83	0.67	2322	1.80
3	P	1	0.43	0.86	0.00	2467	4.00
3	P	20	0.33	0.17	0.83	1451	0.10
4	H	12	0.67	1.00	0.00	2394	2.00
4	H	19	0.67	1.00	0.67	2612	5.00
4	H	13	0.67	1.00	0.67	2868	2.83
4	H	24	0.67	1.00	0.83	2612	1.58
4	P	21	0.67	0.50	0.33	2909	0.51
5	H	29	0.86	0.86	0.00	2600	10.00
5	H	31	0.75	0.75	0.75	2700	1.00
5	H	10	0.83	0.83	0.83	2902	1.00
5	H	4	0.77	0.92	0.92	3104	1.00
5	H	2	0.60	0.93	0.93	3374	3.91

Iron versus protein

The assumption that iron levels are correlated with protein levels in hand-deboned beef provides the motivation for using excess iron or the iron-to-protein ratio as a measure of product quality with regards to the evaluation of whether there are more than negligible amounts of bone marrow (i.e., hematopoietic cells) in the product. This assumption was valid for the hand-deboned data of the 1996 survey referred to above. If excess iron were present in the AMR product, then the correlation for AMR product would be smaller, in so far as the added bone marrow or hematopoietic cells (that is the assumed cause of the added iron) would not contribute to the protein levels. Thus, for a high(er) protein product, high(er) levels of iron would be expected, but the percentage increase due to the addition of bone marrow (i.e., hematopoietic cells) would be less than that of a low(er) protein product. Hence, for AMR product, the lack of a significant correlation does not in itself represent evidence that invalidates the assumption. For these data, there is not a large correlation when computed over all samples. Figure 3 is a scatterplot of iron versus protein, by machine type, together with the linear regression lines.

Figure 3: Scatterplot of iron versus protein, by machine type, with linear regression lines. The shorter line is the linear regression line for the data from the Protocon machines.



For the 170 samples, the Pearson correlation of iron and protein is 0.13, and of excess iron and protein, equal to –0.09. However, the non-parametric correlations (Spearman and Kendall) of iron and protein were nearly zero, and those of excess iron and protein, were negative.

A fairer evaluation of the hypothesis of positive correlation between iron and protein, given everything else being equal, is based on the correlations that exist within each establishment and machine. For each machine, the Spearman correlations were computed and were for the most part positive: 30 of the 41 correlations computed were positive, 1 was zero, and 10 were negative. This pattern is significant at approximately the 0.0139

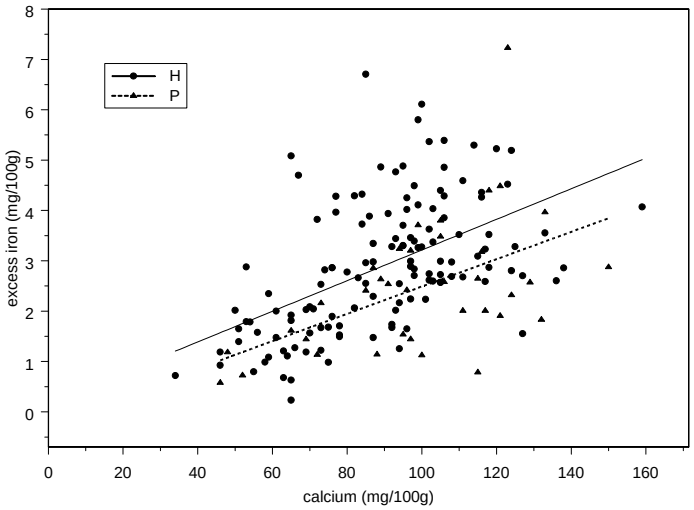
level, based on the sign test. For the Pearson correlation, the results are similar: for the sign test the significance is 0.05. When the same statistical test is computed using the excess iron measurement instead of the iron measurement, the significance level is 0.76 for the sign test. For the Spearman correlation the significance level is 0.14. The conclusion is that iron is correlated with protein.

Excess iron and calcium

The mean of the calcium results for the Hydropsep machine is about 89 mg/100g; for the Protocon, it is about 100 mg/100g. For iron, there was one reported result at 15.7 mg/100g; the next highest result was 9.6 mg/100g. The 15.7 mg/100g result is considered an outlier and thus deleted from the analyses. The mean of the excess iron results for the Hydropsep machine is about 2.90 mg/100g; for the Protocon, it is about 2.48 mg/100g. As will be discussed below, these differences are not statistically significant. The highest value for calcium is 159 mg/100g, with a corresponding excess iron value of 4.1 mg/100g. The sample came from a class 4 establishment, a Hydrosep machine, and the sample was positive for UNT. The second highest calcium result is 150 mg/100g, from a class 3 establishment, Hydrosep machine, however, the excess iron result for that sample is 2.9 mg/100g, which is about the average, and the sample was negative for UNT. The highest excess iron result (excluding the one sample with iron result of 15.7 mg/100g) is 7.23 mg/100g and the calcium value is also a relatively high at 123 mg/100g. This sample came from a class 2 establishment, a Protocon machine, with a positive unacceptable nervous tissue sample. The second highest excess iron result is 6.7 mg/100g, from a class 3 establishment, Hydrosep machine, however, the calcium result for that sample is a moderate 85 mg/100g and the sample was negative for UNT.

First, the relationship of excess iron and calcium and their relationships with the likelihood of UNT in samples is explored. The Pearson correlation of the machine-specific median calcium and excess iron values over all machines is about 0.63, indicating that calcium and excess iron results are positively correlated across machines. For the Hydrosep machines, the correlation is 0.72 and for the Protocon machines, the correlation is 0.71. Figure 4 is a scatterplot of excess iron versus calcium, together with linear regression lines for data from the two machine types. It is seen from the graph that the excess iron levels on the average are higher for the data from the Hydrosep machines; the calcium levels are nearly the same, and the positive correlations of excess iron and calcium for data from the two machines are about the same. This result suggests that there are production factors that may have significant deleterious influence on both calcium and excess iron results.

Figure 4: Scatterplot of excess iron versus calcium, with linear regression lines for data from the Hydrosep and Protocon of machines.



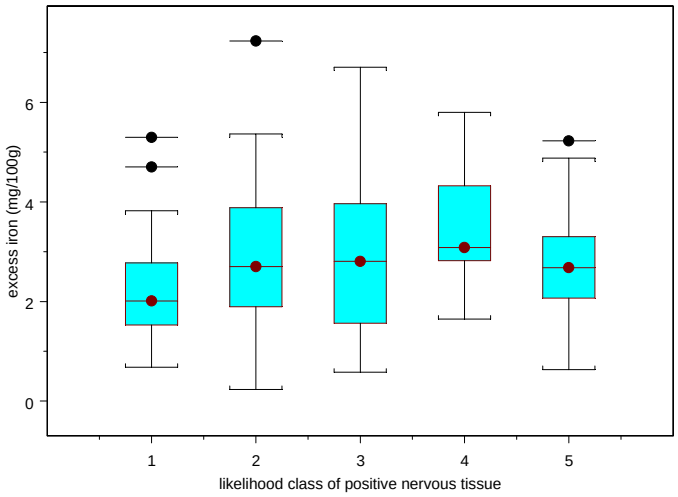
This notion is buttressed somewhat by examining the relationship of excess iron and calcium by the likelihood of UNT tissue classification. In Table 4 is presented the median excess iron and calcium levels over samples within likelihood classes, by machine. Of particular interest are the results for the class 1 samples for the Hydrosep machine: the median excess iron result is 2.13 mg/100g and the median calcium result is 68 mg/100g. These are both the lowest of those presented associated with the Hydrosep machines.

Table 4: Median excess iron and calcium levels (mg/100g) of samples within classes of likelihood of positive UNT result, defined in text.

		UNT Positive					
		No			Yes		
		N	Excess Iron (mg/100g)	Calcium (mg/100g)	N	Excess Iron (mg/100g)	Calcium (mg/100g)
Machine Type	Likelihood +						
H	1	22	2.13	68.0	0	.	.
	2	25	2.71	88.0	4	2.89	82.0
	3	24	3.25	86.0	11	2.24	96.0
	4	7	3.08	97.0	11	3.37	108.0
	5	11	2.66	94.0	18	2.71	103.5
	all	89	2.73	87.0	44	2.98	99.5
P	1	5	1.54	95.0	.	.	.
	2	3	2.01	115.0	1	7.23	123.0
	3	18	2.58	105.0	6	3.44	108.0
	4	1	2.41	87.0	2	2.40	81.0
	all	27	2.32	97.0	9	3.18	100.0

Figure 5 is a boxplot of the excess iron results for different likelihood classes. For the 27 results of class 1, three of them: 3.8, 4.7, and 5.3 mg/100g, exceed 3.1 mg/100g.

Figure 5: Boxplot of excess iron results for samples for different likelihood classes of positive unacceptable nervous tissue.



The relationship of excess iron and calcium with the likelihood of positive tissue samples, within establishments, was examined using Eq. 5. Within establishments but between machines, iron levels on the average did not differ by large amounts, with one exception. Seven establishments used two machines. However, one of these establishments had, relatively, vastly different excess iron measurements. For this establishment, the protein values of the samples processed on the two machines are quite different, with a corresponding difference in the iron levels. The individual results from this establishment are presented below in Table 5. The machine pressures, dwell times, and product types are virtually the same; there were no positive SC tissue findings and only 1 positive DRG tissue finding.

Table 5: Individual results for an establishment (33 on Table 1) with two machines.

Machine	Protein (%)	Iron (mg/100g)	Iron Protein Ratio	Calcium (mg/100g)	Excess Iron (mg/100g)
1	9.60	3.13	0.33	73.00	1.67
1	7.10	2.19	0.31	64.00	1.11
1	9.60	1.69	0.18	65.00	0.23
2	13.60	5.77	0.42	95.00	3.71
2	12.10	6.10	0.50	116.00	4.26
2	10.70	6.99	0.65	102.00	5.37
2	8.40	3.36	0.40	70.00	2.08
2	13.70	6.67	0.49	111.00	4.59

All the above samples, except one (the 4th row), were produced from product for which the type of vertebra used was not specified, and all samples were from product that used bones of an unspecified type (not ribs, pelvis, or flat bones but others). In the following analyses, the two machines for this establishment of Table 5 were considered as separate establishments.

The 6 values of Z (Eq. 5), corresponding to the relations of excess iron and calcium with the occurrences of UNT, SC, or DRG are: Z(exfe, UNT) = 0.30279; Z(exfe, SC) = -0.22916; Z(exfe, DRG) = 0.89923; Z(ca, UNT) = 1.91766; Z(ca, SC) = 0.85937; and Z(ca, DRG) = 1.79846. These results show that, within establishments, calcium was statistically significantly related with the likelihood of unacceptable nervous tissue being present in the samples. The values of excess iron are less so related, which could in part be due to the higher relative error of the estimated values due to sampling and analytical procedures.

Relations of excess iron and calcium with processing parameters

Data analyses were performed for the purpose of exploring relationships between the excess iron and calcium values with machine operating parameters and product type. The

results were equivocal: relationships seen within establishments were not seen between establishments.

Analyses of variances of excess iron, calcium, and protein, assuming machine type as a fixed factor, were performed. There were, in all cases, large and significant (P- value < 0.001) between-establishment effects. Consequently, to determine the significance of effects or relationships, the factor, establishment is considered as a random factor.

The relationship of excess iron and calcium levels with machine type (H or P), pressure, dwell time, and product type was examined. Mean and median establishment - specific values for excess iron, protein, and calcium were computed (where the two machines for the establishment of Table 5 were considered as two establishments). Weighted analyses of variances on the median and mean excess iron and calcium values were performed; with machine type (H or P) as a fixed factor and the weights were the number of samples. The analysis showed an insignificant difference machine type effect, with the possible exception for protein. The differences of the averages of the median excess iron values for the H machines and the P machines is 0.34 mg/100g (= 2.81 for the H machines minus 2.47 for the P machines), with a standard error of about 0.43 and. P-value = 0.44; for calcium, the difference is –11.9 mg/100g, for a P- value of 0.15; but for protein, the difference is -1.10%, with P- value of 0.09.

Analyses of covariance with dwell time, ln(dwell time), and machine pressure as covariates were calculated. Analyses were performed for excess iron, natural log of the excess iron and calcium. For the Hydrosep machines, machine pressure was statistically significant at approximately the 0.12 level for ln(exFe) where the slope for the covariate was negative. For calcium, for the Hydropsep machines, the machine pressure covariate was significant at the 0.17 and the slope of the covariate was positive. These analyses do not demonstrate a strong significant relationship of excess iron and calcium with the machine operating parameters over all the establishments. To explore the relationships further, within and between machines correlations were calculated.

Within machine correlations (Pearson, Spearman, and Kendall) of the excess iron, calcium, and protein, with the machine pressures and dwell times, and type of product, as well as calcium were computed. Table 6 gives the average of the Spearman correlations for the two types of machines.

Table 6: Weighted average, within-machine, Spearman correlations by types of machine (Hydrosep and Protocon) between excess iron, calcium, and protein, with machine dwell times and pressure, as well as calcium. Weights are equal to number of observations. The variable "pbone" means presence of non-vertebrae bones, and is equal to 1 if non-vertebrae bones were used and zero, otherwise. The variable "vet" is equal 1 when a vertebra was recorded indicting no special type, and equal to zero when neck or back vertebra was recorded.

	Machine Type								
	H			P			All		
	mcorr-exfe	mcorr-ca	mcorr-pr	mcorr-exfe	mcorr-ca	mcorr-pr	mcorr-exfe	mcorr-ca	mcorr-pr
Variable									
ca	0.133	1.000	0.072	0.424	1.000	-0.456	0.166	1.000	0.005
dwell	0.204	0.030	-0.124	0.160	-0.462	0.128	0.264	-0.070	-0.101
pbone	0.417	0.111	-0.001	-0.192	-0.192	-0.495	0.266	0.086	-0.135
presspsi	-0.047	0.296	-0.037	-0.170	0.010	0.108	-0.111	0.217	-0.018
vet	0.049	-0.129	-0.103	-0.366	-0.473	0.063	-0.204	-0.352	-0.020

Note that calcium is positively correlated with excess iron, but excess iron is negatively correlated with machine pressure while calcium is positively correlated with machine pressure (as indicated in the analysis of covariance discussed above). Of the 21 Spearman correlations between excess iron and dwell times, 4 are negative (1 of which belonged to a P machine) and 16 are positive, for a two-sided significance (ignoring machine type) of 0.01 for the sign test. Of the 27 Spearman correlations of excess iron and machine pressure, 17 are negative and 9 are positive, for significance of 0.0.17 for the sign test. Of the 27 Spearman correlations of calcium and machine pressure, 6 are negative (4 belonging to the H machines), and 20 are positive, for significance of < 0.01 for the sign test. Of the 21 Spearman correlations between calcium and dwell times, 12 are negative and 8 are positive, for a two-sided significance (ignoring machine type) of 0.5 for the sign test. Of the 14 Spearman correlations of calcium and vertebra type, 8 are negative and 3 are positive, for a significance of 0.23 for the sign test. This analysis

suggests that, within establishments, excess iron is positively correlated with dwell time and calcium is positively correlated with machine pressure.

Across establishments, the same pattern and significance of these correlations are not seen. The median establishment specific excess iron is negatively correlated (Spearman) with dwell times for the H machines (P- value = 0.24), and positively correlated for the P machines (P- value = 0.01); the median establishment specific calcium has nearly zero correlation with mean machine pressure for both P and H machines, a slight positive correlation with vertebra type for the H machines, but negative with the P machines. For the H machines, the Spearman correlation of the median establishment-specific calcium and dwell times was negative (P-value <0.03), but it was positive for the P machines. And the median excess iron is positively correlated with the means of the establishment specific machine pressures for both types of machines.

One further analysis was performed. Mixed linear effect models with the dependent variable of calcium and excess iron, assuming establishments and machines within establishments are random factors, were performed where $\log_{10}$ dwell times for excess iron and machine pressure for calcium were considered as covariates, as above. For calcium, when adding covariates: vet and the interaction of vet and pbone, the test statistic equal to minus 2 times the loglikelihood ratio statistic decreased by about 20, which, based on 3 degrees of freedom, is statistically significant. Table 7 presents the mean calcium and excess iron levels for the types of products and machines by type of product.

Table 7: Mean calcium and excess iron levels for types of machines and product.

		Machine Type						All		
		H			P					
		Ca		Exfe	Ca		Exfe	Ca		Exfe
		N	Mean	Mean	N	Mean	Mean	N	Mean	Mean
Vertebrae	Type bones									
Only Neck	No rib etc.	13	80.2	2.30	14	111.8	2.66	27	96.6	2.48
	Some rib, or etc.	3	95.0	2.54	5	66.6	1.30	8	77.3	1.76
	All	16	83.0	2.35	19	99.9	2.28	35	92.2	2.31
Any	Type bones									
	No rib etc.	31	91.1	2.68	9	102.8	2.56	40	93.8	2.65
	Some rib, or etc.	86	90.2	3.09	9	94.4	2.79	95	90.6	3.06
	All	117	90.5	2.98	18	98.6	2.68	135	91.5	2.94
All	Type bones									
	No rib etc.	44	87.9	2.57	23	108.3	2.62	67	94.9	2.58
	Some rib, or etc.	89	90.4	3.07	14	84.5	2.26	103	89.6	2.96
	All	133	89.6	2.90	37	99.3	2.48	170	91.7	2.81

As is evident, there does not appear to be a consistent pattern; the statistical significance of the interaction seen in the mixed linear effect model does not translate to a practical significance. Thus, no conclusion or statement is being made concerning the possible effects of these factors on calcium or excess iron levels.

In conclusion, there seems to be evidence that would suggest, at least as a hypothesis, that, given everything else being equal, calcium is an increasing function of machine pressure, and excess iron is an increasing function of dwell time. This is based on comparison within establishments. However, across establishments, these relationships do not hold. The uncontrolled nature of this study and the high variability of the results preclude developing estimates of functions that can be used to predict relationships; further, more controlled studies, are needed.

Comparison of pre- and post-desinewing.

There are 135 pairs of matched samples collected from pre- and post-desinewing. The desinewing operation remove cartilage and bone material from the product, which would include iron-depleted protein and calcium, thus it would be expected that calcium and protein levels would decrease, the iron to protein ratio and excess iron measure would increase. In addition, the desinewing operation might make the distribution of UNT tissues more homogenous throughout the finished AMR product, thus making UNT tissues more likely to detect.

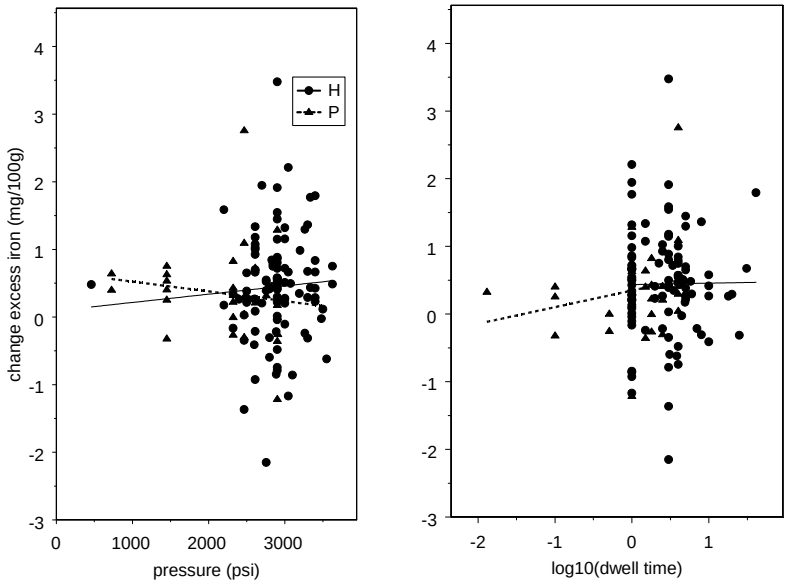
In expectation, the above relationships were valid. However, measurement error and other factors created a significant number of comparisons that were in the opposite direction than expected. Of particular importance in the relationship of protein, pre- and post - desinewing, since the protein that is removed is iron-depleted, thus, causing, if all things remained equal, an increase in the iron to protein ratio from that that would be seen if the desinewing operation was not performed. For protein, 30% of the results had the pre-desinewing protein result lower than the post-desinewing result. The analytical standard deviation for protein is a function of the true level, p, $sd(p) = 0.03p^{0.65}$ (Price, et al, 1994), so that, for example, the standard deviation of measured values on a sample with a true protein level of 16% would be 0.17%. The difference of 2 independent results thus would be 0.25%. Hence, a positive difference greater than 0.50%, would, assuming normality of errors, with 97.5% confidence, represent paired samples with the true pre-desinewing protein value lower than that of the post-desinewing value. Ten percent of the samples had values had pre-desinewing protein measured values more than 0.5% less than that of the paired post-desinewing values. These “unexpected” results could be due to sampling error: the samples themselves might not have been good “representative” samples of the product.

The means of the protein values pre- and post-desinewing for all the paired samples are 15.1% and 14.7%, respectively, for an average difference of 0.42%. The Pearson and Spearman correlations of the differences versus the pre-desinewing protein values were significant at significance level of 0.075 and 0.20, respectively, whereas, the correlations for the logarithmic transformed values were significant at the 0.81 and 0.37 levels. Thus to characterize the differences, the ratio of the post- to the pre-desinewing is considered

There was not a statistically significant machine effect, when assuming establishment is a random factor, though there is a significant establishment effect (P- value < 0.01). For the H machines, the geometric mean of the ratios of the post- to pre- desinewing protein values is 96.5%, while that for the P machines is 98.5%. Over all results, the geometric mean of the ratios is 96.9%. Thus, on average, through the desinewing step, the protein content was reduced by 3%.

Figure 7 contains scatterplots of the change of excess iron versus the machine pressure and $\log_{10}$ of the dwell time. The plots indicate little correlation of the change of excess iron with the machine pressure and dwell time.

Figure 7: Scatterplot of the change of excess iron versus the machine pressure and $\log_{10}$ of the dwell time, and linear regression lines, by type of machine.



The correlation between the change in levels of excess iron and the excess levels in pre-desinewing product within establishments is negative. On average, the excess iron level for the post-desinewing product was larger by 0.43 mg/100g than that of the pre-desinewing product. The iron levels increased by about the same amount: from an average of 4.76 mg/100g to an average of 5.20 mg/100g. Within-establishment Spearman correlations of the changes in excess iron and the excess iron levels in the pre-desinewing product were almost all negative (P-value < 0.001 for the sign test). However, the median establishment specific changes in excess iron were not negatively correlated with the median establishment - specific excess iron for the pre-desinewing product.

ANOVA models for change of excess iron were performed, with pre-desinewing excess iron as the covariate, machine type as a fixed factor and establishment as a random factor and included as an interaction effect with the covariate. The machine type effect was significant at about the 0.05 significance level, and the "average" slope of the covariate of pre-desinewing excess iron levels was negative and significant at better than the 0.01 level. Based on these models, the average standard error of predicted increase in excess iron from the pre- to post-desinewed product is about 0.44, comparable to the mean increase of 0.43 stated above. Covariates of dwell times and machine pressure were added to the model, but there was no significant improvement. Thus, while there does appear to be a relationship of the change in the excess iron levels with that of the excess iron levels in the pre-desinewed product, the relative error of the prediction of the amount of the decrease is large.

Comparison of unacceptable nervous tissue pre- and post-desinewing.

There were 134 matched samples for which analyses of unacceptable nervous tissue were made. Of these 134 samples, 94 of them were negative for unacceptable nervous tissue. However, of these 94, 13%, or 12 of them were positive for the matched pre-desinewing sample; the rate applies to both types of machines: 3 of 14 samples from the Protocon machines were positive and 9 of 70 samples from the Hydrosep machines were positive. While the number of samples is too small to discern any pattern, one establishment had 3 of them. For this establishment 81% of the samples tested were positive. For the SC tissue, 7.4% of the 94 samples were positive, and for the DRG 5.7% of the samples were positive.

Analysis of ELISA procedure for determining the presence of UN tissue.

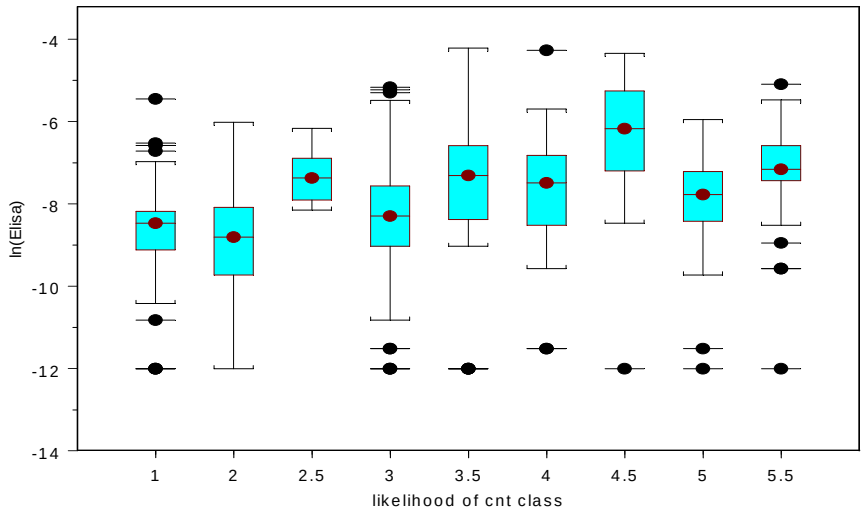
A short explanation of the ELISA (enzyme-linked immunosorbent assay) procedure is presented before the analysis of the results. The ELISA is a sandwich immunoassay that utilizes two antibodies to glial fibrillary acidic protein (GFAP) [one antibody is bound to the bottom of the wells, which entraps the GFAP from the sample, a second antibody conjugated to peroxidase is then used to detect the bound GFAP]. Samples for analyses were formed by first taking enough material from the 1 pound of product that was sent to the laboratory to form four blocks (1/2 x 3/4 x 1 inch each) of tissue. These were placed in a small plastic bag and mixed as thoroughly as possible by manipulating the bag. This comminuted product was than sampled by inserting a cotton swab 3 times into the product at different sites. Excess material is removed from the swab. Supposedly the swab will entrap, on average, about 50 mg of a meat sample. The swab is then inserted into 1 ml of sample diluent, agitated to dislodge the meat sample (20 times), and 50 ul of the 1000 ul of diluent/meat sample is added to one well. The controls were freeze-dried bovine brain at four standard dilutions of "risk material"-- GFAP-- in buffer.

There were 295 ELISA measurements. Of the 295 ELISA measurements, 207 were of final AMR products collected after the desinewer process. The reported results of 207 final AMR products are presented in the Appendix: Table A.1. The reported results were transformed by the natural logarithm, because doing so provides a better resolution of the

data, and because the standard deviations of the logarithmic transforms of the results would be more homogeneous than those of the untransformed results, since the repeatability standard deviation of optical density (OD) responses is often an increasing function of the expected OD. The natural log of the ELISA results reported as zero were assigned a value of –12, since the lowest non-zero ELISA transformed result was –11.5. For the remainder of the report, ELISA results will sometimes refer to the logarithmic transformed results; from the context it should be clear which is meant. Analysis of variances and graphical examination revealed no, or very little, significant differences between the non-zero ELISA transformed responses between the pre- and post-desinewed samples. Thus, this designation will be ignored, at least initially, in the subsequent analyses.

Discrepancies between results of the ELISA tests and the direct UNT tissue determination could be caused in part by sampling and measurement errors: matched or paired samples may actually be different with respect to the presence of UNT, or there may be amounts of UNT that are below the (direct) method's sensitivity contributing to a false negative result for the direct method. Even if the tests for UNT were negative, the corresponding matched sample might contain sufficient amounts of UNT to cause an OD ELISA response, or vice versa. Thus, the comparison of the ELISA results with the detection of UNT in samples should account for the likelihood that the sample would contain UNT, even if it were not found by the direct method. As described above, 5 classes of samples were defined, depending upon the establishment's percentage of samples for which CNS tissue was detected. The class 1 samples are most likely to represent samples that are truly negative. Figure 8 provides boxplots of the ELISA results by likelihood classes of UNT.

Figure 8: Boxplots for natural log of ELISA results, by classes defining the likelihood of UNT being present in sample. Class 1 is the lowest likelihood, and class 5.5 is the highest likelihood. The classes that are whole integers consist of matched samples that were negative for UNT.



If the cutoff values were such that there would be a 25% false positive rate (including all results), then 12 of 55 (22%) samples from class 4.5 and 5.5 would be classified as negative. The cutoff value determined using the 25% criterion is –8.10 on the natural log scale or 0.000304 in the common number scale. Table 8 presents the fraction of ELISA positive samples, using the 25% false positive rate cutoff value.

Table 8: Fraction of ELISA positive samples, using the 25% false positive rate cutoff value, by class of likelihood of positive UNT sample and pre- or post-desinewing.

		Pre or Post Desinewing			
		A		B	
Likelihood UNT +	Direct Method CNS	#	Fraction +	#	Fraction +
1	Negative	12	0.250	32	0.250
	Positive	.	.	1	0.000
2	Negative	14	0.143	38	0.289
	Positive	0	.	8	0.625
3	Negative	23	0.391	45	0.489
	Positive	9	0.444	23	0.565
4	Negative	11	0.818	7	0.714
	Positive	3	0.667	17	0.824
5	Negative	8	0.625	8	0.625
	Positive	7	0.714	28	0.786
All	Negative	68	0.412	130	0.392
	Positive	19	0.579	77	0.701

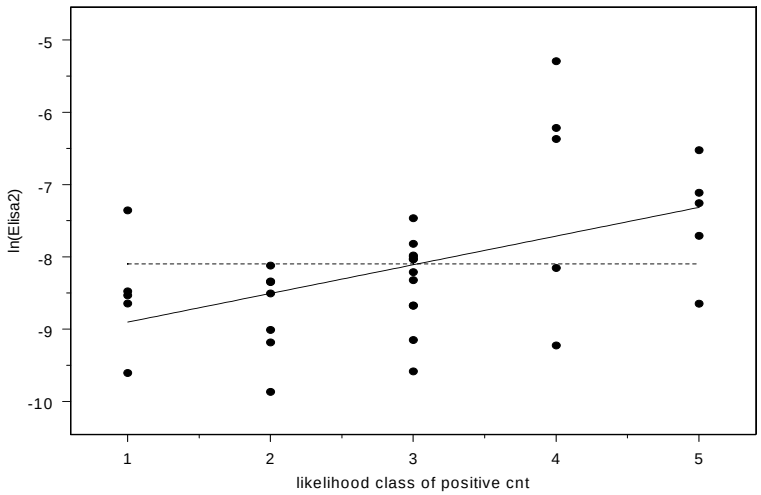
The above statistics in Table 8 may not "represent" the performance characteristics of the ELISA procedure; establishment effects might be important, as the following example shows. There were two establishments for which more than 10 ELISA analyses were performed on post-desinewing samples. Both establishments were classified as high likelihood CNS positive; used the Hydrosep machines; had high machine pressures, averaging over 3000 psi; used additional bones (not identified); used non-specified vertebra; and had samples with an average of the calcium levels equal to about 100 mg/100g. The only differences noted were with the machine dwell times, and protein and iron levels, where one of the establishments had higher average dwell times (3.9 s to 1 s); higher protein (14.6% to 12.6%); higher iron (5.34 mg/100g to 2.87 mg/100g), and consequently, higher excess iron (3.13 mg/100g to 1.96 mg/100g). These results are of particular interest, and thus are presented in Table 9.

The ELISA did well for establishment 4, in which all ELISA tested samples were positive; however, the ELISA did not do so well for establishment 2, in which only 3 of the 11 samples tested were evaluated as positive and 5 samples had false negative ELISA results relative to positive results for the direct method. Figure 9 gives the establishment-specific mean values of ln(ELISA) versus the likelihood of positive CNS class, together with a regression line (to depict the relationship) and the derived cutoff value (dotted line) for distinguishing a positive result, based on the 25% rule developed above.

Table 9: Results from two establishments, for which ELISA responses were different. All samples represent post-desinewed product.

Establishment id	Calcium (mg/100g)	Excess iron (mg/100g)	Machine dwell time(s)	Positive for CNS direct method	ELISA result	Positive for ELISA CNS (>0.000304)
4	105.00	2.57	1.0	yes	0.00078	yes
4	92.00	1.74	1.0	no	.	.
4	92.00	1.68	1.0	no	0.00058	yes
4	82.00	2.07	1.0	yes	0.00073	yes
4	94.00	2.17	.	no	0.00074	yes
4	94.00	1.25	1.0	yes	0.00075	yes
4	111.00	2.68	1.0	yes	0.00123	yes
4	127.00	1.55	1.0	yes	0.00421	yes
4	.	.	.	yes	0.00167	yes
4	.	.	1.0	yes	0.00615	yes
4	.	.	1.0	yes	0.00260	yes
4	.	.	1.0	yes	0.00414	yes
4	.	.	1.0	yes	0.00093	yes
2	97.00	3.46	1.0	yes	.	.
2	95.00	3.30	2.5	no	.	.
2	105.00	2.57	1.5	yes	.	.
2	97.00	2.89	4.4	no	.	.
2	85.00	2.96	4.6	no	0.00106	yes
2	95.00	3.30	5.0	yes	0.00006	no
2	117.00	2.59	4.2	no	0.00014	no
2	98.00	2.84	2.3	no	0.00023	no
2	110.00	3.52	4.3	yes	0.00024	no
2	94.00	2.54	4.4	yes	0.00071	yes
2	95.00	4.88	5.5	yes	0.00205	yes
2	83.00	2.66	4.3	no	0.00022	no
2	.	.	5.2	yes	0.00007	no
2	.	.	5.2	yes	0.00000	no
2	.	.	4.3	yes	0.00007	no

Figure 9: Establishment-specific mean values of ln(ELISA) on post-desinewed samples versus the likelihood of positive CNS class, together with a regression line. The derived cutoff value for distinguishing positive samples is the dotted line.



The establishment effect is clearly seen.

Some of the disagreements between the ELISA and the histologic (immunohistochemical) methods might be due to the non-homogenous AMR products; the ELISA tests a relatively small volume of AMR product compared to the histologic (immunohistochemical) method. Since the AMR product is by nature a heterogeneous substance, the distribtuion of the CNS tissue within the product might be heterogeneous. Thus analyzing a sample that consists of (too) small amount of product might result in a false negative finding for the ELISA method, whereas the matched sample analyzed by the histologic method would be positive, or vice-versa. This latter scenario is thought to be less likely because of the differences in the amounts of sample that are analyzed by the two different methods. Thus, materials in the sample that are not CNS tissue could cause the false ELISA results. However, regarding the whether the ELISA false negatives associated with the product from certain establishments, for example establishment 2 in Table 9, at this time it is not known is due to low densities of CNS for the AMR product of the establishments, or some other material within the AMR product of the establishments that has a deleterious effect of the ELISA sensitivity.

Verification (follow-up) samples

From twenty-three establishments, 68 verification samples were collected and 67 analyzed for the presence of SC tissue (one sample could not be tested because of putrefaction of the tissue). Most of these samples were collected after the survey was

completed. These sample results were not included in the analyses presented above. One establishment had 19 verification samples, one had 12, and the other establishment had no more than 3. Table 10 presents summaries results of the analyses matched by establishment, with results from the survey. As can be seen, the percentage of positive results for the two establishments with 12 and 19 results are similar, but for the remainder, the percentage for the follow-up samples (28%) is less than that of the survey samples (37%). It can be seen that the establishments with the smaller fraction of positive samples for the survey had generally the smaller fraction of positive verification samples.

Table 10. Summary of fraction of positive results for SC tissue for follow-survey up verification samples and survey samples by establishment.

Observation	Number samples survey	Fraction positive survey	Number samples follow-up	Fraction positive follow-up
1	6	0.000	1	0.000
2	13	0.000	1	0.000
3	7	0.143	1	0.000
4	6	0.167	1	0.000
5	6	0.167	2	0.000
6	12	0.167	2	0.000
7	10	0.200	3	0.000
8	5	0.200	1	0.000
9	7	0.286	1	0.000
10	6	0.333	19	0.316
11	6	0.333	1	1.000
12	6	0.333	1	0.000
13	6	0.333	1	0.000
14	15	0.333	2	0.000
15	7	0.429	12	0.500
16	6	0.667	1	0.000
17	6	0.667	3	1.000
18	6	0.667	2	0.000
19	6	0.667	2	0.500
20	4	0.750	3	0.667
21	13	0.769	3	0.000
22	6	0.833	3	0.667
23	7	0.857	1	1.000

Table 11 presents by the time the samples were analyzed: periods 1 and 2 correspond to the first two-thirds and the last third of the survey, and period 3 refers to the follow-up samples that were collected after the completion of the survey. The results for the establishments with 12 and 19 results are presently separately; for the other establishments the results are grouped together and assigned the value of 1 for the variable group in Table 11. As seen in Table 11, the percentage of positive samples collected during the survey periods is not significantly different, but the percentage of the follow-up samples is lower. This pattern holds when considering the establishments' percentages of positive samples; Table 12 presents the fractions of positive samples for the group 1 samples, defined above, where the samples are further divided by whether or not the fraction of SC samples from the survey are greater than 1/3.

Table 11: Fraction of positive SC tissue samples by period of time. Group 1 consist of results from all establishments except the two establishments with 12 and 19 follow-up samples, which are assigned a group value of 12 and 19 respectively.

		Period							
		1		2		3		All	
		N	Frac + SC	N	Frac + SC	N	Frac + SC	N	Frac + SC
Sample Group	Type								
1	Follow-up	11	0.364	4	0.500	21	0.190	36	0.278
	Survey	103	0.311	56	0.518	.	.	159	0.384
12	Follow-up	5	0.200	1	0.000	6	0.833	12	0.500
	Survey	4	0.500	3	0.333	.	.	7	0.429
19	Follow-up	.	.	.	.	19	0.316	19	0.316
	Survey	4	0.250	2	0.500	.	.	6	0.333

Table 12: Fraction of positive SC tissue samples by period of time for Group 1 samples, consisting of results from all establishments except the two establishments with 12 and 19 follow-up samples. Samples are classified as to whether or not the sample's establishment had more then 1/3 of the survey samples with detected SC.

		Type			
		Follow-up		Survey	
		N	Frac + SC	N	Frac + SC
HighPos	Period				
0	1	4	0.000	68	0.132
	2	1	0.000	37	0.324
	3	13	0.077	.	.
1	1	7	0.571	35	0.657
	2	3	0.667	19	0.895
	3	8	0.375	.	.
All		36	0.278	159	0.384

Reference:

Price, Cindy G.; Webb, Neil B.; Smith, Wertice J.; Marks, Harry M.; Yoffe Aaron M. 1994, "Comparison of Mercury and Copper based catalysts in the Kjeldahl determination of nitrogen in meat and meat products: collaborative study", J. of AOAC International, vol. 77, 6: p. 1542-1556.

Appendix:

Table A.1. Advanced Meat Recovery (AMR) Products Survey of 2002:
Enzyme-Linked Immunosorbent Assay (ELISA)
Results of 207 final (post desinewer) AMR Product Samples for the Presence of Central Nervous System (CNS) Tissue
Compared to the FSIS Direct Method

(*The establishment ids were randomly assigned.)

Establishment id*	Direct Method** Positive for CNS result	ELISA Method*** result	ELISA Positive for CNS (>0.000304)
1	no	0.00570	yes
1	no	0.00034	yes
1	no	0.00025	no
1	no	0.00072	yes
1	yes	0.00012	no
1	yes	0.00083	yes
2	no	0.00106	yes
2	yes	0.00006	no
2	no	0.00014	no
2	no	0.00023	no
2	yes	0.00024	no
2	yes	0.00071	yes
2	yes	0.00205	yes
2	no	0.00022	no
2	yes	0.00007	no
2	yes	0.00000	no
2	yes	0.00007	no
3	yes	0.00037	yes
3	no	0.00030	no
3	no	0.00025	no
3	no	0.00012	no
3	no	0.00023	no
4	yes	0.00078	yes
4	no	0.00058	yes

4	yes	0.00073	yes
4	no	0.00074	yes
4	yes	0.00075	yes
4	yes	0.00123	yes
4	yes	0.00421	yes
4	yes	0.00167	yes
4	yes	0.00615	yes
4	yes	0.00260	yes
4	yes	0.00414	yes
4	yes	0.00093	yes
5	yes	0.00000	no
5	no	0.00078	yes
5	no	0.00037	yes
5	no	0.00016	no
5	no	0.00050	yes
6	no	0.00001	no
6	no	0.00048	yes
6	no	0.00034	yes
6	yes	0.00102	yes
6	no	0.00020	no
7	no	0.00022	no
7	no	0.00017	no
7	no	0.00028	no
7	yes	0.00176	yes
7	no	0.00025	no
8	yes	0.00013	no
8	no	0.00006	no
8	no	0.00000	no
8	no	0.00019	no
8	no	0.00004	no
9	no	0.00093	yes
9	yes	0.00032	yes
9	no	0.00042	yes
9	yes	0.00214	yes
9	no	0.00041	yes
9	yes	0.00081	yes
9	no	0.00000	no
9	yes	0.00023	no
10	yes	0.00121	yes
10	yes	0.00094	yes
10	no	0.00048	yes
10	yes	0.00079	yes
11	no	0.00022	no
11	no	0.00003	no
11	yes	0.00139	yes

11	yes	0.00138	yes
12	no	0.00042	yes
12	yes	0.00075	yes
12	no	0.00113	yes
12	yes	0.00652	yes
12	yes	0.00627	yes
13	no	0.00084	yes
13	yes	0.00209	yes
13	yes	0.00271	yes
13	yes	0.00231	yes
13	yes	0.00523	yes
13	no	0.00109	yes
14	no	0.00050	yes
14	yes	0.00011	no
14	no	0.00014	no
14	no	0.00011	no
15	yes	0.00029	no
15	no	0.00011	no
15	no	0.00004	no
15	no	0.00008	no
15	no	0.00026	no
16	no	0.00076	yes
16	no	0.00023	no
16	no	0.00012	no
16	no	0.00015	no
16	no	0.00024	no
16	no	0.00025	no
16	no	0.00139	yes
16	no	0.00025	no
16	no	0.00000	no
16	no	0.00021	no
17	no	0.00013	no
17	no	0.00017	no
17	no	0.00008	no
17	no	0.00000	no
17	no	0.00015	no
17	no	0.00024	no
17	no	0.00003	no
17	no	0.00023	no
17	no	0.00000	no
17	yes	0.00011	no
18	no	0.00051	yes
18	yes	0.00206	yes
18	no	0.00052	yes

18	no	0.00002	no
19	yes	0.00119	yes
19	yes	0.00062	yes
19	no	0.00001	no
19	no	0.00020	no
19	yes	0.00130	yes
20	no	0.00012	no
20	no	0.00033	yes
20	no	0.00006	no
20	yes	0.00027	no
20	no	0.00054	yes
20	yes	0.00113	yes
21	yes	0.00020	no
21	yes	0.00021	no
21	yes	0.00036	yes
21	yes	0.00000	no
22	yes	0.00000	no
22	yes	0.00067	yes
22	no	0.00005	no
22	no	0.00004	no
22	no	0.00023	no
22	yes	0.00030	no
22	no	0.00026	no
23	no	0.00090	yes
23	no	0.00001	no
23	no	0.00031	yes
23	no	0.00416	yes
23	no	0.00036	yes
23	no	0.00086	yes
23	no	0.00032	yes
23	no	0.00030	no
23	yes	0.00000	no
23	yes	0.00033	yes
24	yes	0.00639	yes
24	no	0.00168	yes
24	yes	0.00449	yes
24	yes	0.01302	yes
25	yes	0.00087	yes
25	no	0.00537	yes
25	no	0.00033	yes
25	yes	0.00109	yes
25	no	0.00000	no
26	no	0.00071	yes
26	no	0.00031	yes
26	no	0.00005	no

26	no	0.00006	no
26	no	0.00021	no
26	no	0.00011	no
26	no	0.00007	no
26	no	0.00012	no
26	no	0.00016	no
26	no	0.00000	no
27	yes	0.00018	no
27	no	0.00010	no
27	yes	0.00020	no
27	no	0.00000	no
28	no	0.00431	yes
28	no	0.00007	no
28	no	0.00004	no
28	no	0.00002	no
28	no	0.00069	yes
29	yes	0.00178	yes
29	yes	0.00105	yes
29	no	0.00095	yes
29	yes	0.00048	yes
29	yes	0.00138	yes
29	yes	0.00000	no
29	yes	0.00050	yes
30	yes	0.00033	yes
30	no	0.00056	yes
30	no	0.00112	yes
30	no	0.00037	yes
30	no	0.00037	yes
30	no	0.00000	no
31	yes	0.00084	yes
31	yes	0.00059	yes
32	no	0.00019	no
32	no	0.00022	no
33	no	0.00094	yes
33	no	0.00033	yes
33	no	0.00036	yes
33	no	0.00147	yes
34	no	0.00244	yes
34	no	0.00038	yes
34	no	0.00018	no
34	no	0.00010	no
34	no	0.00088	yes
34	yes	0.00029	no
34	yes	0.00063	yes
34	yes	0.00211	yes

34	no	0.00000	no
34	no	0.00015	no

Footnotes:

*The establishment ids were randomly assigned.

**FSIS Method: For detecting CNS (spinal cord) tissue, FSIS used a method that is a modification of the method published in the *Journal of Food Protection*. Vol. 63, No. 8, 2000, Pages 1107-1112. Titled "An Evaluation of Methods for the Detection of Spinal Cord in Product Derived from Advanced Meat Recovery Systems" by Lynda Collins Kelley, et al. The published method uses four different stains to examine AMR product: (hematoxylin and eosin (HE)) - the routine stain used for all histologic examinations; glial fibrillary acidic protein - GFAP; neurofilament; and synaptophysin - three immunohistochemical stains). Four slides are examined for each stain, resulting in the examination of a total of 16 slides from each AMR sample. The modified procedure is identical to the published procedure with the exception that no sample is stained for either neurofilament or synaptophysin. The modified method used for the survey would not create false positives, i.e., a sample positive for central nervous system (CNS) tissue using the modified method would be positive using the published method.

***Enzyme-linked Immunosorbent Assay (ELISA): ELISA tests were completed on the AMR product samples using the ELISA for GFAP as a screen for CNS tissue.

Source: FSIS Eastern Laboratory, Athens, Georgia
AMR Products Survey 2002

Attachment:

Derivation of excess iron amount for distinguishing hand-deboned meat from meat produced by Advanced Recovery Systems.

Based on results for hand-deboned meat obtained in the 1996 FSIS AMR survey

3/11/03

Determining the maximum acceptable level of excess iron in meat products produced by advanced meat recovery systems

Introduction

As discussed in the FSIS response to comments presented in the preamble to Docket 96-027P, FSIS is using an excess iron measurement for evaluating process control because this measure is associated with bone marrow (i.e., hematopoietic cells) in the product. The assumption is that there is a significant probability that more than negligible amounts of bone marrow (i.e., hematopoietic cells) would be present in product with elevated excess iron measurements. If an obtained excess iron measurement is larger than a statistically defined amount, then the obtained measurement is considered elevated.

The objective of the rule, stated in the preamble of the 1998 proposal is to "provide clear standards…. that include adequate markers for bone-related components (levels consistent with defects anticipated when meat is separated by bone by hand)." This objective is interpreted to mean that for advanced meat recovery (AMR) product to be labeled meat, the excess iron measured levels should be no higher than worst case levels expected (or anticipated) for meat derived from hand deboning when produced under acceptable manufacturing practices. Thus if a product produced by advanced recovery systems has excess iron measured levels greater than these worst case excess iron measured levels, then there is a significant probability that the high iron levels in the product are due to the incorporation of more than negligible bone marrow (i.e., hematopoietic cells) into the product.

A statistical criterion for determining that a specified product (lot) was produced under acceptable manufacturing practice is derived by considering the distribution of an appropriate product characteristic (such as excess iron) when the product is produced under acceptable manufacturing practices and choosing a percentile, p, of this distribution as a demarcation value, D(p). Thus, if, for product produced in a lot, the measured characteristic is greater than D(p), it is assumed that the lot was not produced under acceptable manufacturing practice. The confidence that this is a true assumption and thus a correct decision is greater than p, or, in other words, there is less than a (1-p) probability that the product actually was produced under acceptable manufacturing practice even though the decision was made that it was not so produced. The choice of p is often based on an assessment of the relative costs and risks associated with incorrect decisions, and, lacking some compelling reason, is often set between 95% to 99.9%. The choice of 99.9% would correspond to approximately 3 standard deviation units when the distribution is symmetric and normal. Using 3 standard deviation units is a common choice in quality control when there is desire to be highly confident that a decision to reject a product as being produced under good manufacturing practice is a

correct decision. For the regulation, a choice of 99.9% confidence or 3 standard deviations units above a specified target is used for determining all tolerances.

To determine the distribution of excess iron measurements in hand-deboned meat product, the measurement error due to repeatability will be accounted for. Information from USDA's Agriculture Research Service (ARS) is used to establish a repeatability standard deviation of 0.16 mg/100g for a single iron determination. The data and the results of statistical analysis of the data are presented as an attachment to this report. The repeatability standard deviation of protein is set equal to $0.03x^{0.64}$ where x is the % protein content obtained using the Kjeldahl procedure with mercury catalyst (Price, Cindy G.; Webb, Neil B.; Smith, Wertice J.; Marks, Harry M.; Yoffe Aron M. 1994, "Comparison of Mercury and Copper based catalysts in the Kjeldahl determination of nitrogen in meat and meat products: collaborative study", J. of AOAC International, vol. 77, 6: p. 1542-1556).

Maximum mean level (MML) for a lot

The term "lot" in this setting is used to represent product produced by advanced recovery systems that has been processed uniformly. It is assumed that the starting materials used, the calibrations of the machinery, and other processing parameters that affect the composition of the product would be as uniform as possible. Thus, a lot does not necessarily represent the product produced in a day. Within a lot, the excess iron measurements for different samples of the product would be different due to unavoidable differences in ratios of iron to protein in different animals and analytical variations in the measured iron and protein levels in samples. However, the lot would have a mean excess iron level, which would reflect processing control and would provide, therefore, an appropriate measure for evaluation. In accordance with the above objective of determining the excess iron measured limits of product produced by advanced recovery systems that can not be labeled meat, the first step is to determine the maximum mean level, MML, of excess iron for a lot. From the discussion in the previous section, the MML is equal to 3 times the "between lot" standard deviation of the excess iron for meat derived by hand deboning, prior to the bone-in material being processed by the recovery system. Once this level is determined, then compliance criteria, based on chemical analysis of samples, are developed which take into consideration the between sample and analytical measurement variability. In particular, the criterion for an individual sample, based on duplicate analyses (for both protein and iron) is derived.

In order to derive the excess iron MML for a "lot" and a criterion for an individual sample, the 1996 FSIS AMR neckbone survey results for the meat derived from hand deboning will be used. From each of two establishments, 27 samples of meat derived from hand deboning were collected on various days of production with 3 samples a day (Table 1). The FSIS procedure to measure iron employed a hydraulic wet acid digestion procedure. However, another method, performed by ARS scientists, which uses a dry ash procedure for digestion, obtained iron results approximately double those originally obtained by FSIS. Furthermore, the results obtained by the ARS dry ash procedure were more consistent with levels reported in the HNS Handbook 8 levels for

hand-deboned meat. Consequently, the excess iron values will be calculated using iron results obtained by the ARS dry ash procedure. For samples for which there were not ARS dry - ash procedure results, the FSIS results were multiplied by 2.12 (which was the average ratio of the dry - ash procedure results to the FSIS results). For the 54 samples of hand–deboned product, 45 of them were analyzed by the ARS dry – ash procedure. Table 2 provides a comparison of all the FSIS and ARS obtained results.

In actuality, the meat derived from hand deboning in the survey might have been heterogeneous, so that within a day there might be more than one "lot" of homogeneous product. In an analysis of variance, the day within an establishment effect had a significance level (p- value) of 0.15 (based on 16 degrees of freedom). An examination of the data did not reveal any particular result or set of results that could be classified as an outlier. This suggests that the between day variability compared to the within day variability was not relatively large and that a single day might consist of more than one lot. Thus, it would be expected that the actual between lot variance might be larger than the measured between day variance. Since the between sample (within day) variance is considerably larger than the repeatability variance, it is possible that a sample "represents" a lot. The "truth" may actually be between the two extremes, identified here, of a day representing a lot or a sample representing a lot. Thus a "compromise" calculation is used for determining the between lot variance component. Specifically, the between lot variance component is set equal to "p" percent of the within day variance component plus the between day variance component, and the within lot variance is set equal to 1-p%/100 of the measured within day variance component. For the regulatory derived criteria, p was set equal to 50%, so that the between lot variance is assumed to equal the sum of the between day variance plus ½ of the between sample/within day variance, and the within lot variance is assumed to be equal to ½ the between sample/within day variance.

Excess iron for hand–deboned product, ExFe, is computed as iron minus 0.138 times the percentage protein (Fe – 0.138protein). The factor 0.138 is the ratio of the average iron to average protein of the hand deboned neckbone product from the FSIS survey, so that, for this product, the mean of the excess iron results is 0.00 mg/100g. This factor is also equal to the average iron to protein ratios of the samples. As stated above, the repeatability standard deviation for the iron measurements is assumed to be equal to 0.16 mg/100g, which was derived from information obtained from ARS (see attachment), and the repeatability of protein measurements is equal to $0.03x^{0.64}$ where x is the percent protein in the sample. Using these values, from the formula for excess iron, ExFe = Fe - 0.138x, the average repeatability variance from the hand deboned samples was calculated to be equal to 0.0265. An analysis of variance (AOV) of the sample excess iron results is presented in Table 3.

The between day/establishment variance, from Table 3, is estimated to be 0.13408, and the between sample/within day variance, after accounting for the measurement variance is estimated to be 0.2875. Thus, the between lot variance, assuming p= ½, is 0.13408 + ½ 0.2875 = 0.2778, so that between lot standard deviation is 0.5271 mg/100g. The maximum mean excess iron for a lot (MML) is 3 times the between lot standard deviation = 3(0.5271) = 1.5813 mg/100g.

Determining Tolerance for Compliance Purposes

FSIS may take samples to evaluate whether or not establishments are producing product produced by advanced recovery systems with "lot" averages greater than the MML, 1.5813 mg/100g. The amount of product in a sample is assumed the same as the sample amounts of the 1996 FSIS survey of product derived from advanced recovery systems.

FSIS recognizes that the ARS process removes connective tissue that contains "little or no iron." FSIS believes the effect of this removal is not large. Connective tissues can be removed pre- or post-desinewing. The amount that is removed during the pre-desinewing stage of processing depends on the machine pressure applied when separating the meat from the bone; the higher the pressure, the more connective tissue is removed. From the FSIS 1996 survey, it seems that the average difference in protein between pre- and post-desinewing product was 0.5 percent, based on a post-desinewing product average protein of about 16.5 percent. Therefore, as a percentage of protein, the amount of protein associated with connective tissue removed during this step averaged about 3 percent and does not represent a large proportion of the protein that is in the product.

In addition, during the ARS processing, some unbound water is removed which would result in the removal of some water-soluble protein and dissolved solids. A possible consequence therefore is that some water-soluble proteins are removed and most of the bound iron will remain in the product, thus, resulting in a higher iron to protein ratio in the ARS product.

Because of these two reasons for the possible increase in iron to protein ratio of ARS product, for this final rule, FSIS is incorporating 10 percent factor to adjust the protein when calculating levels of excessive iron in ARS product. Thus, in the calculations of excess iron, the measured protein will be multiplied by 1.10.

A general sampling plan is to take n samples throughout the lot, composite them, and perform n_r analytical measurements. If Fe_i and pr_i represent the i^{th} iron and protein results, respectively, then the adjusted excess iron, aExFe, result for an n – sample composite is

$$aExFe = \sum Fe_i / n_r - (0.138)(1.10)\sum pr_i / n_r$$

The expected variance of the adjusted excess iron estimator, aExFe, for n – sample composites obtain by such a sampling plan is:

$$\text{var}(aExFe) = \sigma_s^2 / n + \sigma_r^2 / n_r$$

where σ^2_r is the repeatability variance of adjusted excess iron measurements and σ^2_s is the between sample/within lot variance of the adjusted excess iron.

The above identified within lot variance component is determined using the results obtained from the 1996 FSIS AMR neckbone survey on the hand deboned product, using the ARS dry-ash iron results, as explained above. FSIS is using the results from

this product rather than the AMR product because FSIS considers that the AMR product was not produced in accordance with the FSIS requirements for meat, and thus, can not, justifiably, be used for determining the within lot standard deviation for product produced by advanced recovery systems that is comparable to meat. It might be that, under good manufacturing practices, product produced by advanced recovery systems would be more homogeneous than its counterpart meat derived from hand deboning, so that the within lot variance for such produced AMR product would be smaller than the within lot variances derived here.

In order to select a specific sampling plan (that is, the number of samples for a lot) producer and consumer risks (probabilities of the lot passing the test) must be selected. The MML represents the maximum mean level for a lot that does not result in a non-compliance determination, thus if a lot had a mean equal to the MML then there should be a high probability that this lot would not fail and pass the sampling plan. As discussed above, for determining tolerances, FSIS is selecting 3 standard deviations above the mean, so that if a result on a n – sample composite is obtained that exceeds the demarcation value, then there would be approximately 99.9% confidence that the mean for the "lot" exceeds the MML, and thus the product within the lot would not be considered comparable to meat. To compute a consumer risk, the probabilities of passing a lot with mean excess iron level that is equal to 2 times the MML are determined.

FSIS laboratories would analyze a compliance sample at least in duplicate (see attachment). Thus it is assumed that n – sample composites are analyzed in duplicate, so that $n_r = 2$. Because of the factor 1.10, the variance components for this estimator will be different from those given in Table 3. The analysis of variance for the meat derived from hand deboning was repeated using the above formula. Presented in Table 4, are the derived variance components for the above estimator of the adjusted excess iron statistic for the hand-deboned product.

The protein values do not affect by much the standard deviation, so that it can be assumed that the repeatability variance is 0.0267. The between sample/within lot variance, σ_s^2, as discussed above, is equal to (1-p) times the between sample/within day, where p = ½. Thus, σ_s^2 = ½ 0.2905 = 0.1453, and the expected variance for the adjusted excess iron results for n – sample composites is therefore, Var(aExFe) = 0.1453/n + 0.0267/2. The square root of this quantity is the expected standard deviation.

Table 5 provides demarcation values for determining that a lot has mean excess iron greater than the MML, 1.5813 mg/100g, for different numbers of samples taken from the lot, assuming duplicate measurements on the composite of the samples. An individual sample is when n=1 so that the individual sample limit that is specified in the regulation is 2.776 mg/100g. For purposes of the regulation (for recalling the demarcation value), this is adjusted to 2.800 mg/100g, so that if an obtained sample result (based on the average of duplicate analyses of iron and protein) is greater than or equal to 2.800 mg/100g then the product produced by advanced recovery systems can not be labeled meat (see conclusion section, below).

If a different percentage, p, than 50% of the within day variance component is added to the between day variance component, then different answers are obtained for the

individual sample demarcation value and for MML. Figure 1 is a plot of the MML and the individual sample limit. The percentage that gives the maximum individual sample demarcation value is 30% and the maximum value is 2.8048 mg/ 100g, with a MML equal to 1.400 mg/100g. The minimum possible derived individual sample limit, obtained when p=100%, is 2.2942 mg/ 100g, with the maximum possible derived MML equal to 1.948 mg/100g.

FSIS Survey of product produced by advanced recovery systems.

Presented in Table 6 are the establishment means of adjusted excess iron for product produced by advanced recovery systems, after the 10% adjustment, aExFe = Fe – (0.138)(1.10)(protein), and the percentage of samples that were greater than or equal to the derived individual sample limit, 2.800 mg/100g. All the establishment means were greater than the MML of 1.5813 mg/100g. Also included in Table 6 is the establishment means of excess iron (not adjusted) of the meat derived from hand deboning. The highest individual excess iron sample result for the hand-deboned meat was 1.76 mg/100g. For product produced by advanced recovery systems, 62% of the samples had adjusted excess iron results that were greater than or equal to 2.800 mg/100g.

Conclusion:

If a mean of results from duplicate analyses on a sample is greater than or equal to 2.800 mg/100g then it is assumed that there is product that is not meat, because of the incorporation of more than a negligible amount of bone marrow (i.e., hematopoietic cells). The question that needs to be answered is to what product (the lot) does this conclusion apply. In answering this question, it is assumed that contiguous product is in the same lot. Since an establishment is required to have documentation that its production process is in control, it is assumed that a non-compliant finding is a result of a failure or a deficiency in the process control. A consequence is that all product that is produced before or after the non-compliant sample might also have been produced when the process was not in control and thus should or could not be labeled meat. One way of showing that product is not from the same "lot" is to examine the records of values of processing parameters that affect the composition of the product produced by advanced recovery systems or other analytical results from samples of product produced in different parts of the day or on different days and to determine if there are reasons to identify different "lots" which would not have non- complying product (mean levels of excess iron less than 1.58 mg/100g).

Figure 1: Plot of derived maximum mean excess iron for lot (MML) and individual sample excess iron limit as function of percentage, where within lot variance equals the sum of the between day variance plus given percentage of within day variance and within lot variance equals 100-percentage of within day variance.

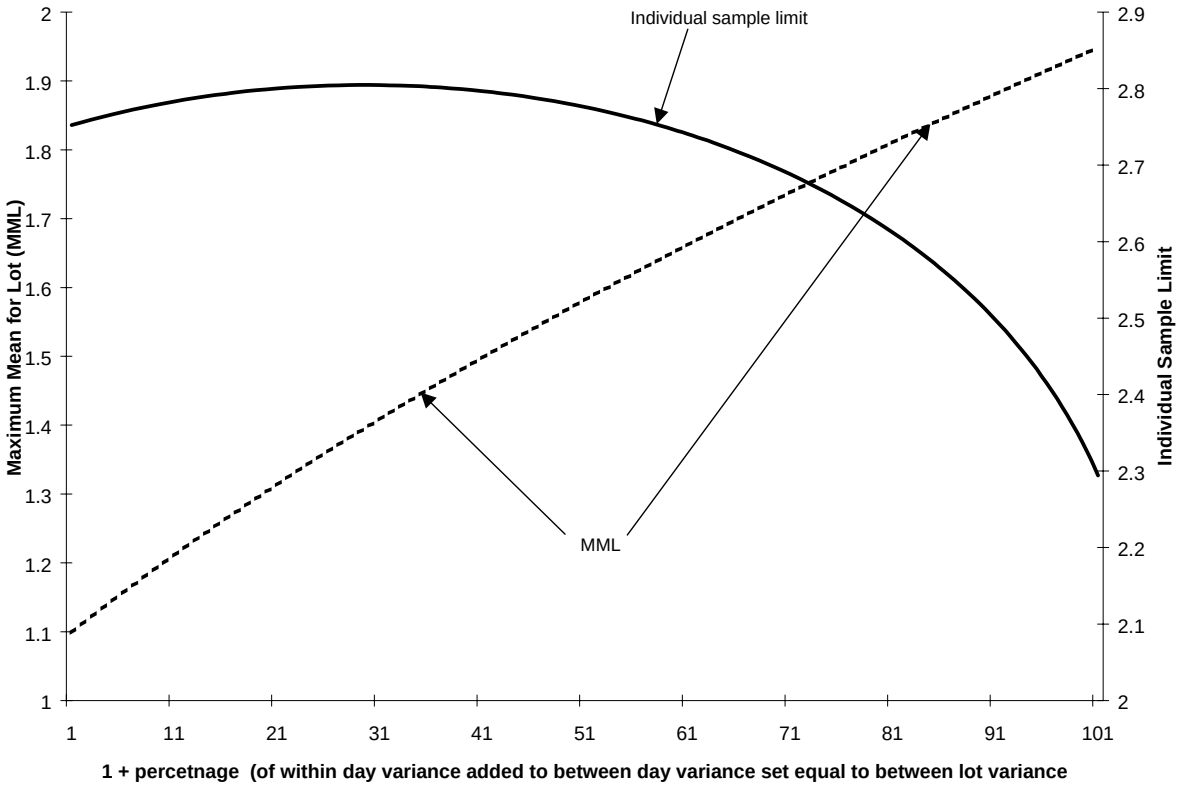


Table 1: Iron and Protein measurements obtained for hand deboned product.
Obtained from 1996 FSIS survey.

OBS	Est. code	Date of sampling code	Iron dry-ash mg/100g	Protein(%)	Excess iron measure mg/100g
1	5	1	2.26	19.13	-0.38
2	5	1	2.36	19.59	-0.34
3	5	1	2.46	19.32	-0.21
4	5	2	2.74	19.56	0.04
5	5	2	2.60	20.24	-0.19
6	5	2	3.29	19.61	0.59
7	5	3	2.54	21.90	-0.48
8	5	3	2.47	21.38	-0.48
9	5	3	4.01	22.65	0.88
10	5	4	2.50	21.25	-0.43
11	5	4	2.62	21.81	-0.39
12	5	4	2.47	22.72	-0.66
13	5	5	3.04	19.91	0.29
14	5	5	3.22	19.50	0.53
15	5	5	3.79	19.33	1.13
16	5	6	3.22	21.76	0.22
17	5	6	3.44	22.15	0.39
18	5	6	4.14	20.85	1.26
19	5	7	3.22	23.05	0.04
20	5	7	2.99	23.35	-0.23
21	5	7	4.33	21.96	1.30
22	5	8	3.03	23.30	-0.19
23	5	8	2.92	22.95	-0.25
24	5	8	4.55	23.25	1.34
25	5	9	3.18	23.25	-0.03
26	5	9	4.93	23.00	1.76
27	5	9	3.52	22.00	0.48
28	8	1	2.70	20.88	-0.19
29	8	1	2.64	21.54	-0.33
30	8	1	2.43	21.49	-0.54
31	8	2	2.88	21.86	-0.14
32	8	2	2.84	21.07	-0.07
33	8	2	2.88	22.34	-0.21
34	8	3	2.58	21.56	-0.40
35	8	3	2.73	21.91	-0.29
36	8	3	2.62	21.02	-0.28
37	8	4	2.56	21.64	-0.42
38	8	4	2.72	22.92	-0.45
39	8	4	2.48	22.14	-0.57
40	8	5	3.01	22.80	-0.14
41	8	5	3.71	21.09	0.80

Table 1 (cont): Iron and Protein measurements obtained for hand deboned product.

OBS	Est. code	Date of sampling code	Iron dry-ash mg/100g	Protein(%)	Excess iron measure mg/100g
42	8	5	3.35	21.44	0.39
43	8	6	2.45	22.02	-0.59
44	8	6	2.14	20.75	-0.72
45	8	6	1.67	21.79	-1.34
46	8	7	2.98	20.90	0.09
47	8	7	2.31	23.91	-0.99
48	8	7	3.66	22.10	0.61
49	8	8	3.35	22.90	0.19
50	8	8	3.23	23.20	0.03
51	8	8	3.40	22.61	0.28
52	8	9	2.37	20.42	-0.45
53	8	9	1.94	21.84	-1.07
54	8	9	3.82	21.91	0.80

Table 2: Comparison of ARS Dry Ash and FSIS Wet Acid digestion results
Units mg/100g (AMR= product from advanced recovery systems

OBS	Type of Product	FSIS Wet Acid	ARS Dry Ash	Ratio Dry Ash to Wet Acid
1	AMR	2.36	5.08	2.15
2	AMR	3.19	5.90	1.85
3	AMR	2.46	7.03	2.86
4	AMR	1.83	4.13	2.26
5	AMR	2.69	4.13	1.54
6	AMR	2.81	5.15	1.83
7	AMR	2.49	4.80	1.93
8	AMR	1.79	5.18	2.89
9	AMR	2.23	5.59	2.51
10	AMR	2.41	5.97	2.48
11	AMR	7.91	8.32	1.05
12	AMR	4.88	7.02	1.44
13	AMR	2.39	5.56	2.33
14	AMR	2.94	5.23	1.78
15	AMR	2.57	4.99	1.94
16	AMR	2.64	5.08	1.92
17	AMR	2.72	4.97	1.83
18	AMR	2.81	4.88	1.74
19	AMR	2.82	5.09	1.80
20	AMR	2.04	5.26	2.58
21	AMR	3.59	5.36	1.49
22	AMR	3.56	5.19	1.46
23	AMR	2.90	6.17	2.13
24	AMR	3.03	5.43	1.79
25	AMR	2.52	3.53	1.40
26	AMR	3.51	5.61	1.60
27	AMR	3.26	5.33	1.63
28	AMR	3.32	5.03	1.52
29	AMR	2.68	4.76	1.78
30	AMR	3.17	5.42	1.71
31	AMR	3.90	6.05	1.55
32	AMR	2.31	5.00	2.16
33	AMR	2.70	6.43	2.38
34	AMR	1.51	4.93	3.26
35	AMR	2.36	5.37	2.28
36	AMR	2.26	5.07	2.24
37	AMR	1.79	5.26	2.94
38	AMR	1.70	4.44	2.61

39	AMR	2.42	4.37	1.81
40	AMR	1.70	4.89	2.88
41	AMR	2.30	6.20	2.70
42	AMR	2.51	6.31	2.51
43	AMR	2.52	5.61	2.23
44	AMR	2.66	6.17	2.32
45	AMR	3.05	4.72	1.55
46	AMR	3.21	6.11	1.90
47	AMR	2.39	6.10	2.55

Table 2 (cont): Comparison of ARS Dry Ash and FSIS Wet Acid digestion results
Units mg/100g (AMR= product from advanced recovery systems

OBS	Type of Product	FSIS Wet Acid	ARS Dry Ash	Ratio Dry Ash to Wet Acid
48	AMR	2.20	4.99	2.27
49	AMR	1.78	6.11	3.43
50	AMR	2.20	5.85	2.66
51	AMR	2.42	5.91	2.44
52	AMR	2.55	6.38	2.50
53	AMR	3.87	6.78	1.75
54	AMR	1.78	4.86	2.73
55	AMR	1.79	6.74	3.77
56	AMR	2.27	5.95	2.62
57	AMR	4.11	6.43	1.56
58	AMR	2.96	6.19	2.09
59	AMR	2.31	4.99	2.16
60	AMR	2.37	4.72	1.99
61	AMR	2.29	4.67	2.04
62	AMR	1.94	5.42	2.79
63	AMR	1.52	5.43	3.57
64	AMR	2.07	5.62	2.71
65	AMR	3.03	5.63	1.86
66	AMR	2.05	7.47	3.64
67	AMR	2.36	6.58	2.79
68	AMR	3.00	5.30	1.77
69	AMR	2.89	4.85	1.68
70	AMR	3.21	4.36	1.36
71	AMR	2.84	5.44	1.92
72	AMR	2.69	5.89	2.19
73	AMR	2.95	6.09	2.06
74	AMR	3.34	6.33	1.90
75	AMR	3.95	5.91	1.50
76	AMR	4.44	7.52	1.69
77	AMR	3.45	5.37	1.56
78	AMR	3.73	5.47	1.47
79	AMR	3.69	5.94	1.61
80	AMR	2.75	4.89	1.78
81	AMR	2.51	5.69	2.27
82	AMR	2.62	5.65	2.16
83	AMR	2.69	5.93	2.20
84	AMR	2.36	5.07	2.15
85	AMR	1.97	4.42	2.24
86	AMR	2.02	5.40	2.67
87	AMR	2.25	5.91	2.63
88	AMR	3.63	8.95	2.47
89	AMR	3.85	6.98	1.81
90	AMR	4.08	7.13	1.75
91	AMR	2.57	5.70	2.22
92	AMR	3.03	8.19	2.70
93	AMR	1.88	3.26	1.73
94	AMR	2.64	6.52	2.47
95	AMR	2.54	7.27	2.86
96	AMR	3.86	7.01	1.82
97	AMR	3.10	6.39	2.06
98	AMR	3.77	5.71	1.51
99	AMR	3.02	6.80	2.25
100	AMR	2.35	5.16	2.20
101	AMR	2.58	4.86	1.88
102	AMR	2.29	5.02	2.19
103	AMR	2.96	3.87	1.31
104	AMR	3.21	5.91	1.84
105	AMR	1.88	5.93	3.15
106	AMR	2.04	5.48	2.69
107	AMR	2.40	4.14	1.73
108	AMR	2.36	5.11	2.17

109	AMR	3.10	4.99	1.61
110	AMR	3.42	5.83	1.70
111	AMR	2.34	4.86	2.08
112	AMR	3.89	4.91	1.26
113	AMR	2.96	6.05	2.04
114	AMR	3.50	6.43	1.84
115	AMR	3.37	5.38	1.60
116	AMR	1.97	4.96	2.52
117	AMR	3.15	5.60	1.78

Table 2(cont): Comparison of ARS Dry Ash and FSIS Wet Acid digestion results
Units mg/100g (AMR= product from advanced recovery systems

OBS	Type of Product	FSIS Wet Acid	ARS Dry Ash	Ratio Dry Ash to Wet Acid
118	AMR	3.00	6.15	2.05
119	AMR	3.24	5.96	1.84
120	AMR	2.72	5.68	2.09
121	AMR	3.54	6.00	1.69
122	AMR	3.70	6.55	1.77
123	AMR	2.63	4.83	1.84
124	AMR	3.12	5.95	1.91
125	AMR	4.14	8.21	1.98
126	AMR	1.78	4.37	2.46
127	AMR	1.96	5.27	2.69
128	AMR	1.91	3.98	2.08
129	AMR	2.22	5.84	2.63
130	AMR	2.85	5.35	1.88
131	AMR	2.35	6.13	2.61
132	AMR	3.64	7.67	2.11
133	AMR	3.19	7.08	2.22
134	AMR	2.27	6.56	2.89
135	AMR	2.96	5.16	1.74
136	AMR	2.57	6.16	2.40
137	AMR	2.25	6.00	2.67
138	AMR	2.14	6.37	2.98
139	AMR	2.39	6.61	2.77
140	AMR	2.81	6.66	2.37
141	AMR	5.30	6.34	1.20
142	AMR	5.13	6.98	1.36
143	AMR	4.36	5.13	1.18
144	Hand	1.06	2.26	2.13
145	Hand	1.10	2.64	2.40
146	Hand	1.09	2.36	2.17
147	Hand	1.06	2.46	2.32
148	Hand	1.34	2.43	1.81
149	Hand	1.40	2.70	1.93
150	Hand	1.78	3.29	1.85
151	Hand	1.56	2.60	1.67
152	Hand	1.82	2.74	1.50
153	Hand	1.78	2.88	1.62
154	Hand	1.90	2.88	1.51
155	Hand	1.79	2.84	1.58
156	Hand	1.43	2.48	1.74
157	Hand	1.56	2.62	1.68
158	Hand	1.65	2.72	1.65
159	Hand	1.48	2.50	1.69
160	Hand	1.37	2.56	1.87
161	Hand	1.60	2.62	1.64
162	Hand	1.18	2.47	2.09
163	Hand	1.26	2.58	2.05
164	Hand	1.69	1.67	0.99
165	Hand	1.53	2.45	1.60
166	Hand	1.21	2.14	1.77
167	Hand	1.74	3.04	1.75
168	Hand	1.70	3.82	2.25

169	Hand	1.07	1.94	1.81
170	Hand	1.51	2.37	1.57
171	Hand	1.01	3.01	2.98
172	Hand	1.25	2.47	1.98
173	Hand	1.20	4.01	3.34
174	Hand	1.14	2.54	2.23
175	Hand	1.41	3.22	2.28
176	Hand	1.77	4.14	2.34
177	Hand	1.64	3.44	2.10
178	Hand	1.14	2.98	2.61
179	Hand	1.58	3.66	2.31
180	Hand	1.62	2.31	1.43
181	Hand	1.19	3.23	2.71
182	Hand	1.10	3.40	3.09
183	Hand	1.19	3.35	2.82
184	Hand	1.22	3.03	2.48
185	Hand	1.60	4.55	2.84
186	Hand	1.44	2.92	2.03
187	Hand	1.46	4.93	3.38
188	Hand	1.40	4.33	3.09

Table 3: Analysis of variance of excess iron results, ExFe, from 1996 FSIS survey of hand- deboned neckbone samples (2 establishments with 27 observations per establishment). ExFe = iron – 0.138protein

Source of Variation	Variance	Standard deviation
Between establishment	0.08054	0.2838
Between Day within establishment	0.05354	0.2314
Sum: Between establishment/day	0.13408	0.3662
Within day including measurement error	0.3140	0.5603
Measurement error	0.0265	0.1628
Between sample/Within day	0.2875	0.5362
Total variance	0.4481	0.6694

Table 4: Analysis of variance of adjusted excess iron, aExFe, based on results from 1996 FSIS survey of meat derived from hand deboning. aExFe = iron – (0.138)(1.10)protein.

Source of Variation	Variance	Standard Deviation
Between establishment/day	0.1333	0.3651
Within day including measurement error	0.3172	0.5632
error	0.0267	0.1633
Between sample/Within day	0.2905	0.5390

Table 5: Limits for determining that a lot has mean adjusted excess iron, aExFe, greater than 1.58mg/100g for n – sample composites as function of the number of samples, n, per lot, assuming duplicate analysis on the composite of the samples. The aExFe n – sample composite result is equal to the mean of the iron results minus the product of 0.138, 1.10 and the mean protein result. The derived limit is equal to 3 expected standard deviations above the maximum mean for a lot (MML) = 1.5813 mg/100g. Also presented are the probabilities of passing a lot with a true excess iron mean = 3.1626 mg/100g (= 6 between lot standard deviations above zero excess iron).

Number of Samples	Limit	Prob. (%) passing lot mean=3.163
1	2.776	16.5821
2	2.461	0.8345
3	2.327	0.0385
4	2.250	0.0021
5	2.199	0.0001

Table 6: Summary of excess iron results from 1996 FSIS neckbone survey.

The hand–deboned excess iron results are computed as: iron –0.138protein,

The product produced by advanced recovery systems (AMRS) excess iron results are computed as: iron – (0.138)(1.10)protein.

establishment code	number of samples	mean excess iron	percent samples > 2.776
8 hand	27	0.221	0.00
9 hand	27	-0.221	0.00
all hand	54	0.000	0.00
1[a] AMRS	27	2.778	40.74
2 AMRS	24	3.950	87.50
3 AMRS	16	3.280	56.25
4 AMRS	27	2.656	33.33
5 AMRS	25	3.443	76.00
6[b] AMRS	25	3.560	84.00
7 AMRS	19	3.065	57.90
ALL AMRS	163	3.235	61.96

[a]) establishment used Protecon machine, while others used Hydrosep machines.

[b]) establishment did not perform desinewing operation.

Attachment:

Repeatability of iron measurements using the ARS Dry-Ash procedure

Data to determine the repeatability of the ARS Dry- Ash procedure was provided to FSIS by Dr. Bob Windham of ARS. Analyses were conducted on beef samples. For further details contact Dr. Bob Windham. The first data set consists of duplicate results obtained by the same laboratory on 47 samples. The second data set are results from a 3-laboratory, 5-sample collaborative study, where each sample was analyzed in duplicate by each lab.

The results from the 47 samples of the first data set are given in Table 1. Statistical analysis did not indicate a non-zero correlation of the standard deviations and mean levels of the samples, so that it is assumed that the repeatability standard deviation does not depend upon the level of iron in the sample. Figure 1 is a plot of the sample standard deviations versus the sample means for the 47 samples. The line represents a quadratic fit. It can be seen from this graph the 5 data points that have standard deviations greater than 0.5 mg/ 100g. The standard deviations of these 5 data points can be assumed to be outlier standard deviations. This can be seen by computing the ratio of the maximum sample variance to the sum of the sample variances and comparing this ratio to appropriate percentiles of a beta distribution (Hawkins, D. M., 1980, Identification of Outliers, Chapman and Hall, New York, NY, Appendix 9). Specifically, let $v_{(j)}$ be a random variable representing the jth ordered sample variance from k samples. The ratio of the maximum sample variance to the sum of the sample variances,

$$r_k = v_{(k)} / \sum_{j=1}^{k} v_{(j)}$$

is compared to an appropriate percentile of a beta distribution with parameters ½ and (k-1)/2. To determine whether $v_{(k)}$ is an outlier with respect to the set $\{v_{(j)}$, for $j<k\}$, the observed value of the ratio, r_k, is compared to the $1-\alpha/k$ percentile of the beta distribution, where α represents the significance of the statistical test of $v_{(k)}$ being an outlier. For k=43, … 47, the ratios r_k were computed and the corresponding significance levels, α_k, were determined. For k=47, $\alpha_{47} = 0.00029$, so that the highest computed variance can be considered as an outlier. For k=46, $\alpha_{46} = 0.00007$, so that the second highest variance can be considered as an outlier. Also, $\alpha_{45} = 0.00482$, $\alpha_{44} = 0.03097$ and $\alpha_{43} = 0.03577$, so that the five highest variances can be considered as outliers. Assuming that the variances on these five samples are outlier results and thus excluding them from the analysis, the repeatability standard deviation from the remaining 42 samples is estimated (by computing the square root of the mean of sample variances) to be 0.161 mg/100g. If the sample with standard deviation 0.58 mg/100g is included in the calculations, then the estimated repeatability standard deviation is estimated to be 0.182 mg/100g. Further, the distribution of the differences of the duplicate analyses within the 42 samples appeared to

be normally distributed. Thus, percentiles of the measurement distribution can be assumed to be distributed as normal.

The data (Table 2) from the collaborative study (3 labs, 5 samples, measured in duplicate) contained possible two outlier results. The 4.91 mg/100g result obtained by the second lab for the first replicate of the third sample is quite different from the other five results, which range from approximately 8 to 9 mg/100g. Thus, this result was not used in the statistical analysis. In addition, for the fifth sample, the sample standard deviation obtained by the first lab, also appears to be an outlier. This can be seen by examining Table 3, which presents means and standard deviations of the replicate results for a sample. The computed ratio, r_{14}, of the maximum sample variance to the sum of the 14 sample variances (excluding the third sample from the second lab) is 0.723, which has statistical significance of $\alpha = 0.0008$. The estimated standard deviation of repeatability (obtained through an analysis of variance), excluding only the outlier result of 4.91 mg/100g was 0.182 mg/100g. When the results for the fifth sample that were obtained by the first lab are also deleted, the estimated standard deviation of repeatability is 0.100 mg/100g.

For deriving the criteria for excess iron in AMR product that can be labeled meat, the repeatability standard deviation is assumed to be 0.16 mg/100g. This is based on the estimated repeatability standard deviation obtained when deleting the 5 samples with standard deviations greater than or equal to 0.58 mg/100g. Support for the 0.16 mg/100g value is the 0.10 mg/100g estimate of the repeatability standard deviation from the collaborative study when the two outlier results are deleted. Because of the few large differences of duplicate sample results, it is recommended there should be at least duplicate analyses on samples used for compliance purposes.

Figure 1: Plot of within sample standard deviations versus sample mean iron level.

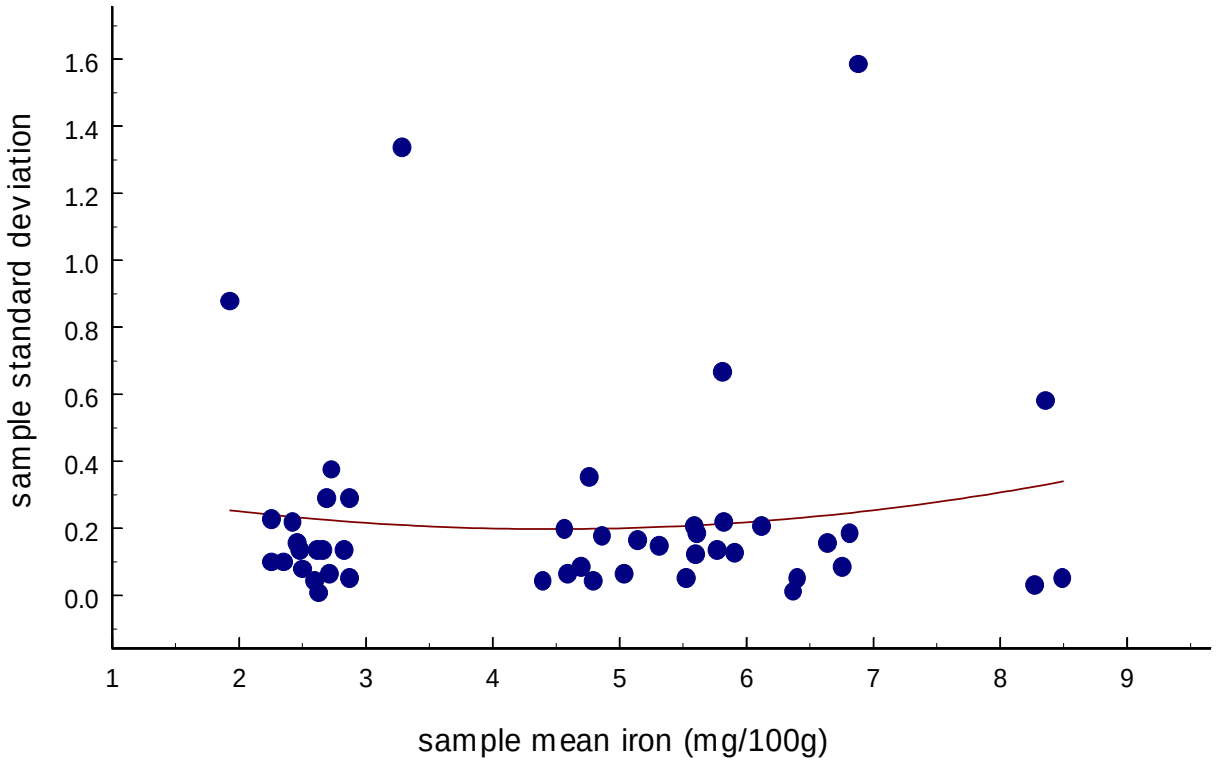


Table 1: Duplicate iron results (mg/100g) from 47 meat samples

sample	replicate 1	replicate 2	mean	standard deviation
1	2.33	2.19	2.26	0.10
2	2.63	2.64	2.64	0.01
3	2.29	2.43	2.36	0.10
4	2.35	2.57	2.46	0.16
5	2.58	2.27	2.43	0.22
6	2.90	2.49	2.70	0.29
7	4.24	2.35	3.30	1.34
8	2.63	2.57	2.60	0.04
9	2.47	3.00	2.74	0.37
10	2.84	2.91	2.88	0.05
11	2.67	3.08	2.88	0.29
12	2.74	2.93	2.84	0.13
13	2.58	2.39	2.49	0.13
14	2.72	2.53	2.63	0.13
15	2.67	2.76	2.72	0.06
16	2.56	2.45	2.51	0.08
17	6.95	6.69	6.82	0.18
18	4.64	4.76	4.70	0.08
19	4.52	5.02	4.77	0.35
20	4.83	4.77	4.80	0.04
21	4.74	4.99	4.87	0.18
22	4.37	4.43	4.40	0.04
23	5.82	6.00	5.91	0.13
24	5.42	5.21	5.32	0.15
25	4.55	4.64	4.60	0.06
26	5.67	5.98	5.83	0.22
27	5.56	5.49	5.53	0.05
28	5.52	5.69	5.61	0.12
29	6.38	6.37	6.38	0.01
30	6.70	6.82	6.76	0.08
31	5.03	5.26	5.15	0.16
32	5.68	5.87	5.78	0.13
33	6.44	6.37	6.41	0.05
34	8.30	8.26	8.28	0.03
35	8.46	8.53	8.50	0.05
36	7.95	8.77	8.36	0.58
37	4.71	4.43	4.57	0.20
38	6.53	6.75	6.64	0.16
39	6.29	5.35	5.82	0.66
40	5.45	5.74	5.60	0.21
41	5.48	5.74	5.61	0.18
42	5.00	5.09	5.05	0.06
43	5.77	8.01	6.89	1.58
44	6.27	5.98	6.13	0.21
45	2.10	2.42	2.26	0.23
46	2.76	2.57	2.67	0.13
47	2.55	1.31	1.93	0.88

Table 2: Results from Collaborative study; data provided by ARS

lab	repli-cation	Sample number 1	2	3	4	5
1	1	5.68	6.44	8.30	8.46	8.77
1	2	5.87	6.37	8.26	8.53	7.95
2	1	4.96	5.92	4.91	7.41	7.90
2	2	4.73	5.96	8.00	7.25	8.05
3	1	5.81	5.21	8.99	8.77	8.88
3	2	5.79	5.53	9.03	8.73	8.84

Table 3: Means and standard deviations for samples

	Laboratory						Pool	
	1		2		3			
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Sample								
1	5.775	0.134	4.845	0.163	5.800	0.014	5.473	0.122
2	6.405	0.049	5.940	0.028	5.370	0.226	5.905	0.135
3	8.280	0.028	8.000	.	9.010	0.028	8.516	0.028
4	8.495	0.049	7.330	0.113	8.750	0.028	8.192	0.073
5	8.360	0.580	7.975	0.106	8.860	0.028	8.398	0.341