SAMPLING ISSUES IN THE SCIENCE-BASED APPROACH TO MICROBIOLOGICAL TESTING IN MEAT INSPECTION PROGRAMS

Ronald S. Fecso
National Agricultural Statistics Service, USDA
3251 Old Lee Hwy., Room 305, Fairfax VA 22030-1501

Key Words: Food safety; *E. coli*; benchmark surveys; control charts

1. INTRODUCTION

Frequent media reports of illnesses caused by microbial pathogens in food increase concern by consumers and regulators. Buzby and Roberts (1995 and 1996) provide background on the origins and costs of these illnesses. They report that a 1991 survey of primary household meal planners or preparers found that 43 percent reported bacteria and parasites in food to be the food safety issue of most concern. Further, they report that these pathogens cause between 6.5 million and 33 million cases of human illness and up to 9,000 deaths each year in the United States with cost estimates exceeding $5.6 billion annually. In response to consumer perception and demands, government scrutiny of the food supply increasingly focuses on efforts to restrict microbial pathogens in the food supply.

While people can be exposed to pathogens from a variety of sources, foods, especially meat and poultry, are the primary sources of illness from pathogens such as *Listeria monocytogenes*, *E. Coli* O157:H7, *Salmonella*, and *Campylobacter jejuni*. For most pathogens, the process begins with the introduction of the pathogen in the intestinal tract of some animal before arriving at the slaughter plant. The raw meat contamination occurs during the stages from slaughter where the intestinal tract may be punctured, through skinning or defeathering where feces on the hides or feathers of animals can cause contamination. The initial contamination of one carcass can spread to other carcasses at many stages of the packing line, for example, at chilling. Sampling issues related to measurement and control of the slaughter related sources of contamination are the focus of this paper.

USDA's Food Safety and Inspection Service (FSIS) generally has responsibility for regulations regarding raw meat and poultry safety. The traditional FSIS inspection program for meat and poultry used organoleptic (sight, touch, and smell) properties to identify and prescribe corrective actions for problems with, predominantly fecal, contamination at slaughter plants. The inability of the traditional inspection process to prevent the outbreaks of food borne illness prompted food scientists to recommend that FSIS require all slaughter and processing establishments to adopt a system of process control--known as HACCP (Hazard Analysis and Critical Control Point)--to prevent food safety hazards. (To conserve limited space, there will be no detailed description of HACCP requirements. Readers may find details on FSIS, GPO and other web sites.) FSIS proposed eliminating the traditional inspections and creating regulations based on HACCP principles that place a new emphasis on controlling microbial pathogens.

This regulation was the most far-reaching and complex to affect the meat and poultry industry in several decades. The HACCP regulation emphasizes the use of scientific methods especially microbial science and process control for monitoring the presence and reducing the incidence of harmful bacteria. To verify that HACCP systems are effective in reducing contamination with harmful bacteria, FSIS has set pathogen reduction performance standards for Salmonella that slaughter plants and plants that produce raw, ground meat and poultry have to meet. In addition, slaughter plants are required to conduct microbial testing for generic E. coli to verify that their process control systems are working as intended to prevent fecal contamination, the primary avenue of contamination for harmful bacteria. The regulation, as initially proposed incorporated (1) sampling of carcasses at the slaughter plants to obtain microbiological tests, (2) a control charting methodology to evaluate sample results, and (3) benchmark surveys to calibrate the control charts.

FSIS used statistical analysis in the development of the HACCP regulation, but the statistical basis of process control performance measures and microbiological testing lacked outside peer review procedures for the statistical aspects of the proposal. This paper highlights some problematic statistical issues identified by the author while serving on an Office of Risk Assessment and Cost-Benefit Analysis panel reviewing the HACCP proposal and how these issues affect the scientific arguments for the regulation. A more detailed paper is planned.

3. THE "ARGUMENT"

To cover the statistical issues of this complex scientific and political problem in a few pages, the basic "scientific argument" upon which the proposed regulations for testing for generic *E. Coli* rest is abstracted. (Similar problems arise in the proposed *Salmonella* testing proposals, but they are not discussed here.)

It is generally agreed that most problem pathogens result from fecal contamination at some point and almost all fecal matter contains generic *E. Coli*. Therefore, a

fecal contaminated carcass has generic *E. Coli* and a carcass contaminated with a problem pathogen contains generic *E. Coli*. Thus, a carcass "lacking" (or at least acceptably low in) generic *E. Coli* indicates "good' (at least relative to the acceptable standard) slaughter practice resulting in a lacking (or at least an acceptably low level) of fecal contamination and problematic pathogens.

It is relatively easy to test for generic *E. Coli* compared to the problematic organisms. Thus, FSIS believes, requiring testing for generic *E. Coli* to supply test evidence indicative of good production practices would compel plants with poor practices to initiate changes aimed at reducing fecal and thus pathogen contamination.

To be effective, the statistical methodology used must provide information related to poor or changing production practices. The following sections discuss areas of conflict between the statistical methods originally proposed and the "argument." Where relevant, subsequent changes to the regulation are mentioned.

4. FSIS JUSTIFICATION FOR THE *E. COLI* PROCEDURES

Elder presents arguments for the *E. coli* verification procedure specified in the proposed Pathogen Reduction /HACCP regulation. The preamble to the regulation stated that E. coli testing is a process control verification tool that slaughter plants are being required to use to verify the adequacy of their controls to prevent and reduce fecal contamination. FSIS developed the proposed regulation based on their interpretation of the purpose of verification under HACCP, as defined by the National Advisory Committee on Microbiological Criteria for Foods (NACMCF)(1989).

Elder summarizes NACMCF commentary by stating that "some degree of microbiological end-product testing is a recognized component of HACCP verification, though by no means the only component." He reports that HACCP documents contain no further guidance concerning specific recommendations for verification sampling. However, he cites Prince (1992) placing verification in a broader context. "Verification is different than monitoring. Monitoring is like quality control in that the monitoring step is going on during the process so adjustments can be made in the process before the product leaves the production line. Verification is like quality assurance in that it is a check on the system."

Elder then looks for further guidance on verification testing in the literature of quality assurance/quality control (QA/QC). The glossary of the American Society for Quality Control Statistics Division, 1983 defines quality assurance as "all those planned or systematic actions necessary to provide adequate confidence that a product or service will satisfy given needs." The glossary notes that quality assurance "involves the necessary plans and actions to provide confidence through verifications, audits . . . " In Juran's (1974) chapter on the subject, quality assurance is defined as "the activity of providing, to all concerned, the evidence needed to establish confidence that the quality function is being performed adequately." Noting that this "confidence" in not to be confused with "statistical confidence," which has a special, restricted meaning, the basis for confidence includes:

- a formal plan that spells how fitness for use will be achieved,
- a system of reviews to verify that the plan, if followed, will result in fitness for use, and
- a system of audits to verify that the plan is actually being followed.

Elder states that under HACCP, these three elements correspond to the HACCP plan, validation and verification [as an audit].

Juran discusses specifics of the "system of audits," describing quality audits as an independent review to compare some aspects of quality performance with a standard for that performance, (1) by a company to evaluate its own activities or those of its suppliers and (2) by regulatory agency to carry out its mission to regulate. Audits maybe of quality plans themselves or of the execution of those plans.

For product audit, which is a type of verification sampling, Juran states that when the purpose of the audit is assurance rather than control, "the sample size is arbitrary and is a balance between the cost of large samples and the unreliability of small samples." The product audit results provide a "running scoreboard" of product performance.

Elder concludes from the discussion of quality assurance, "that establishment verification of process control of fecal contamination through *E. coli* testing fits comfortably into the quality assurance framework. The concepts of product audit, volume-based testing, 'running scoreboard', and criteria based on 'marker quality' as determined by a regulatory agency all can be identified in the *E. coli* verification testing procedure."

4.1 FSIS Discussion of Distribution Models for Microbiological Data

Elder notes that the *E. coli* data collected for verification will be variables data, for example, measurements of the number of *E. coli* colony forming units per square centimeter. Discussion of sampling procedures requires the identification of a suitable statistical model for the distribution of the data. Insight on the distribution of microbiological data in general is found in Kilsby (1982). Elder cites Kilsby's description of a variables sampling procedure for lot acceptance,

based on the average and standard deviation of logarithms of counts, that is appropriate for lognormal data.

One should recognize that the process of creating the "batch" to measure impacts the distribution. Also, the methods are for "lot acceptance" sampling. FSIS makes a leap from Kilsby's application that focuses on ground and mixed products in a lot acceptance situation to whole carcass methods for process verification without much justification.

To illustrate how microbiological data are more complicated, Kilsby and Pugh (1981) note that sometimes "results are censored; i.e., numbers are too low to be detected by the analytical method employed." Further, the data may have true negatives, which cannot be distinguished from false negatives (censored positive observations). Statisticians refer to one such distribution as a censored delta-lognormal distribution. Elder cites Aitchison and Brown, and Crow and Shimizu for detailed discussion of the lognormal model and reference to the delta-lognormal distribution, which is a mixture of true zeros and lognormally distributed positive values. He notes that neither discusses the censored delta-lognormal distribution.

Elder discusses available data. "The data relevant to choosing a distributional model for E. coli verification testing depends on which purpose of verification testing is being addressed. Data from individual establishments' in-control slaughter processes are needed to evaluate distributional models for verification from the individual establishment perspective. Such data were not available to FSIS during the development of the regulation: FSIS baseline studies were not designed to provide data on individual establishments; and establishments did not volunteer data. However, overall distributions of baseline results give a qualitative idea of what sort of distributions may be encountered in individual establishments, and these distributions are directly useful for the 'market based standards' discussed by Juran."

Based on the national distribution for broilers from the FSIS steer/heifer (1994) and broiler (not released at the time of the panel study) baseline studies, it is plausible that results from a particular broiler slaughter process may be lognormally distributed, but it is not a plausible model for all steer/heifer slaughter processes. The data also suggests that a single distributional model that could be applied to both the broiler and steer/heifer slaughter classes would have to be more complex. In the case of steer/heifer data, a nominally quantitative test, a high degree of censoring or a high percentage of true negatives render it nearly qualitative.

Elder states that, "for the sake of uniformity and simplicity, it is desirable to have a single distributional model and a single verification procedure derived from it that can be applied to all slaughter classes. The model that appears likely to be appropriate for variables data is the censored delta-lognormal model (of which the lognormal model is a special case). Unfortunately, estimating parameters of a censored distribution can be a complex undertaking (Crow and Shimizu, 1988) that would be impracticable in the setting addressed by the regulation. It was the prospect of using such a model that precluded FSIS from selecting traditional Shewhart 'X-bar and R charts' or other variables control charts for the *E. coli* verification procedure."

FSIS Discussion of Selecting a Sampling Procedure

Elder reports that FSIS has selected a three-class attributes sampling plan and adapted it to verification testing by applying it to a moving window of test results. Three-class attributes sampling are defined by four parameters: a boundary (**m**) defining acceptable counts of the organisms of concern; a higher boundary (**M**) defining unacceptable counts; the number (**n**) of sample units to consider; and the number (**c**) of sample units permitted to exceed **m** in the number of units (**n**) considered. No sample result is permitted to exceed the higher limit, **M**.

For the procedure chosen by FSIS the process performance level is the fraction of tests for which (1) the current test does not exceed **M** and (2) not more than 3 of the last 13 tests exceed **m**. The expected performance is found using the trinomial distribution (Bray, 1973), assuming that the moving count of results above **m** starts at zero and runs without resetting thereafter. The selection of **m** and **M** as approximate 80th and 98th percentiles of baseline results for each slaughter class (type of animal) implies that approximately 18 percent of results are marginal for a process meeting baseline performance.

The three-class attributes plan can be applied in the following manner. As each new result is obtained, compare it to the upper limit, **M**. A result above **M** is unacceptable and is an indication of possible process problems that should be investigated. In addition, keep a moving count (moving sum) of the number of results above **m** in the last n tests. A moving count of more than **c** is another indication of possible process problems that should be investigated.

SAMPLING CARCASSES FOR MICROBIAL TESTING

The effectiveness of the proposed microbiological and statistical testing program depends on interrelationships among the carcass sampling method, the distribution of fecal contamination on the carcass, and the laboratory methods. Carcass sampling methods differed for meat and poultry based primarily on carcass size issues and not on supporting statistics. Where statistical analyses were presented to support the proposals, they were not adequate, consisting primarily of descriptive univariate presentations.

Even in designed studies conducted outside the agency (e.g., Agricultural Research Service, 1996), FSIS prescribed simplistic analyses that limited the information to be gleaned from the study. The choice of one versus three site methods is an example where one site methods were proposed as if there was no difference, that is, there were no statistical tests presented. Later, data from the above study suggested that there was a substantial and significant difference, a more than threefold increase in percent positive CFU/cm**2>=5 when the area swabbed is tripled.

Further, a review of the available FSIS studies made it clear that the probability that *E. Coli* was found in a sample was directly related to the percent of surface area sampled and the clustering or uniformity of the distribution of fecal matter on the carcass. For example, poultry had nearly 100 percent positive test findings in baseline surveys. This is a reasonable result because the entire carcass is sampled and the slaughter processing tends to result in fecal contamination on all carcasses and more uniformly on each bird. In contrast, beef, where only a small area is sampled and fecal contamination would tend to be concentrated in spots on the carcass, had more than 90 percent negative findings in baseline studies. Yet, it is likely that at least some *E. Coli* is present somewhere on most beef carcasses. Hogs, where the slaughter process tends to distribute fecal contamination over more of the carcasses and more of each carcass's surface but only a small area is sampled, had fewer negative findings.

This "meta analysis" provided evidence that for large carcasses, a small sample site has large sampling error. That is, sampling a single anatomical site, as FSIS at one point recommended for beef carcass inspections, has too low a probability of finding E. Coli even when it is on a carcass.

The next section discusses how the low probability of detection affects the measure's usefulness in control charts.

CONTROL CHARTING METHODOLOGY

The documentation on the rational for the chosen control methodology cites well-known references for process control and some basic applications of control procedures in the food industry. The testing strategy and the statistical methodology for the program were proposed in materials provided by FSIS including working drafts of the regulation, the working draft document "Moving Sum Procedures for Microbial Testing in Meat and Poultry Establishments" and pages 486 to 504 from "Best Management Practices for Salmonella Risk Reduction in Broilers" (Holder, 1993).

The combination of the proposed methods is not a straightforward application of textbook methodology. The theory for calculation of the action limits for the moving sum control chart needs more thorough definition and discussion. The methodology is similar to an "np" chart, but the sampling window is far too small for textbook application where np>5 is required. Thus, three-sigma calculations are not appropriate due to small sampling "windows." (Ryan, 1989) The approach to calculating the action limit to achieve the desired detection characteristics is through binomial probability calculations. Such methodology is appropriate, as opposed to the three-sigma approaches found in the textbook standard methods. Yet, other control chart methodologies could have been candidates, for example, ANOM for attribute data or demerit charts, but there is no discussion of these.

A major inadequacy is the lack of discussion of the impact of measurement error on the process control methodology. McCaslin and Gruska (1976) and others have found that measurement error can be a serious problem in process control. Here, the measurement error of concern is the random selection process on the carcass. Carcass mapping has found that pathogens are not uniformly dispersed on the carcass, but are found in clusters at random. Thus, detection on a carcass suffers from a false negative rate dependant on the carcass distribution. Baseline data analysis suggested that false negatives may far exceed true positives. While the difference in protocol between the baseline data and the operational procedures may not completely invalidate the control process, it would alter the meaning of baseline data and the lower rate of false negatives would affect control operating characteristics. The problem is not as severe in poultry where distributions are more uniform on the carcass.

Understanding Variation

Data collection activities need to have a focused purpose, a decision to be made. Yet, FSIS often expressed a "feeling" that by the mere institution of regulatory measures, food safety benefits would ensue because plants would be compelled to begin microbial testing.

Since decisions in this case are related to the goal of reduced pathogens, one must assess how the collected data will lead to decisions that result in improvement. For example, will the variations outside the control limits for beef be predominantly random or will they lead to identifiable "fixes." Industry data cast some doubt. In one analysis, the Center for Science in the Public Interest (CSPI) made the assessment that the percentages of contaminated carcasses needs to be looked at, not just sample sites. They appropriately focus on the need to address slaughter plant cross contamination in the assessment of the usefulness of the testing. The comparison of existing data across industry reflects differences primarily due to differing slaughter processes

and thus different measurement errors, raising concern about the standardized "one-size-fits-all" methodology.

BENCHMARK SURVEYS TO CALIBRATE THE CONTROL CHARTS

At issue regarding baseline surveys is their suitability for use in determining benchmarks to begin a control process or to revise control limits over time. First, there is a need to determine what the baseline survey should measure. FSIS acknowledges that there is a lack of scientific evidence to support a choice of the degree of contamination of meat that presents a health hazard. Further, current processing technology makes a zero tolerance to microbial presence infeasible, thus FSIS chose a sampling plan which focuses on identification of the distribution of sample measurements of *E. Coli*. With such a distribution, those processors with repeated evidence of contamination well beyond the capability demonstrated by the distribution would then be identified.

Based on the benchmark surveys, levels for incidence of contamination were set nationally on a product volume basis rather than for each plant. An action limit based on the product specific national benchmark is chosen to identify plants required to take further action. Such a strategy, FSIS proposes, imposes added burden only on the worst cases. The proposed design reflects many compromises including the need to minimize burden (time and expense of sampling) especially for small concerns. But several considerations need more attention including the understanding that there is variation in the distribution of contamination between individual plants and within plants over time, characteristics not evident in simple displays of the distributions. More rapid improvement would be expected with a plant-based capability benchmark because the proposed method encourages the examination of only the worst of processes. FSIS indicated that the cost of setting such levels from a benchmark survey would be prohibitive.

Also, the benchmark survey collection protocol differs from the proposed collection protocol to be carried out by the plant. There is considerable concern that the sensitivities of the protocols are quite different. With differing sensitivities, the number of plants exceeding the control limits may be vastly different from that expected under the statistical assumptions used.

Future Benchmark Surveys

FSIS proposes that baseline surveys be used to measure future changes in the national incidence rate. Apparently, the design of future surveys would be similar to that of the first baseline survey. Implied is the use of these measurements to improve product quality by periodically lowering the national benchmark assuming the measured incidence decreases. For such use of the data, more specific information is needed on sampling error of the measured incidence level for individual pathogens and the methodology for estimating them. (Also, more discussion is needed on the control of nonsampling errors in the baseline survey.) If overall sample size is the same as the first baseline survey, the subsampling rates for many of the pathogens are likely to be too low to measure real change over time. In short, more justification is needed to warrant conducting future baseline surveys.

Program performance may be improved using between facility differences rather than product incidence rates. Instead of the baseline survey approach, FSIS could facilitate the formation of a data base essential for understanding the extent and sources of problems, a prerequisite for systemic improvements. Unfortunately, the current proposal does not have provisions for the use of data collected by the plant. Such data would provide a measure of between plant variation and be more useful in setting control limits based on process capability than the baseline survey.

CONCLUSIONS

FSIS discusses the available scientific considerations used to form the basis of the proposed testing and clearly documents what is proposed and available supporting evidence. For poultry, enough statistical evidence is presented to concur with the assessment of the Scientific Review Panel that the requirement of E. Coli testing, while not necessarily reducing illness, should result in reduced presence of fecal contamination if used properly.

The testing of beef carcasses, as proposed, is likely to impact the amount or incidence of fecal contamination directly only in the worst case operations. Marginal gains across the majority of plants in the typical operating range are unlikely to result from the testing. Thus at issue in the proposal for beef is the cost. The extremely poor operations could be identified with much less sampling.

While individual microbial aspects of the plan have been reviewed by the relevant scientific community, validity of the plan as a whole rests upon the proper application of statistical science. There is no evidence that the statistical science has had similar scrutiny outside USDA. For example, at the Technical Conference on *E. Coli* Testing sponsored by FSIS on September 12-13, 1996, there were no conference participants outside of USDA personnel who would regard themselves as experts in the statistical methodology of process control as used for continuous improvement or in survey sampling. While many speakers expressed reservations about the rule, it is doubtful that the interested parties fully assessed the complex statistical issues in this technical conference.

FSIS panels and conferences would benefit by including outside statisticians with the specializations to review the statistical science upon which the program is based, specifically, expertise in statistical process control and sample survey methodology. Even the recently appointed Food Safety Research Working Group, formed

to establish a research agenda that supports the fundamental changes FSIS is making to food safety regulation lacks statistical representation.

ACKNOWLEDGMENTS

This work originated as part of the author's participation on the Review Panel of the Pathogen Reduction: Hazard Analysis and Critical Control Point Systems Regulation. The panel was convened by the Office of Risk Assessment and Cost-Benefit Analysis, Office of the Secretary, USDA. The author thanks the other panel members and ORACBA staff for their subject matter expertise and help in refining the statistical contributions to the panel review.

DISCLAIMER

The information and conclusions found in this paper are the opinions of the author and do not necessarily reflect the position of the Panel, or the U.S. Department of Agriculture or any of its agencies.

REFERENCES

Agricultural Research Service, "Comparative Sampling Study: Excision and Sponging," Draft report, U. S. Department of Agriculture, 1996.

American Meat Institute, "Comments for FSIS *E. coli* Implementation Meeting," Presented at the *E. Coli* Verification Testing Conference, Washington, D.C., September 12-13,1996.

American National Standards Institute (ANSI), ANSI/ASQC Z1, 1-1985, Guide for Quality Control Charts, American Society for Quality Control, Milwaukee, 1985, p.2.

Buzby, J. C. And T. Roberts, "ERS Estimates U.S. Food borne Disease Costs," *Food Review*, USDA, Econ. Res. Serv., Washington, D.C., Vol. 18, Issue 2, May-August 1995, pp. 37.

Buzby, J. C. And T. Roberts, "Microbial Food borne Illness: The Cost of Being Sick and the Benefits of New Prevention Policy," *Choices*, The American Agricultural Economics Assoc.,Ames, Iowa, First Quarter 1996, pp. 14-17.

Center for Science in the Public Intrest, "Limitations of the *E. Coli* Verification Testing Program in Addressing Cross-Contamination, and Other Comments on Criteria, Sampling Frequency, and testing Methodologies," Presented at the *E. Coli* Verification Testing Conference, Washington, D.C., September 12-13,1996.

Crow, E.L. and K. Shimizu (Editors), Lognormal Distributions: Theory and Application, Marcel Dekker, New York, 1988.

Elder, R.S., "Statistical Basis for *E. Coli* Sampling Procedure," draft paper, Food Safety and Inspection Service, U.S.D.A., May 1996.

Food Safety and Inspection Service, "Nationwide Beef Microbiological Data Collection Program: Steers and Heifers," United States Department of Agriculture, Washington, 1994.

Juran, J.M., "Upper Management and Quality," section 21 in Quality Control Handbook, Third Edition, J.M. Juran (Editor-in-Chief), McGraw-Hill Book Company, New York, 1970, p. 569.

Juran, J.M. and F.M. Gryna, Quality Planning and Analysis, McGraw-Hill Book Company, New York, 1970, p. 569.

Kilsby, D.C., "Sampling Schemes and Limits," Chapter 10 in Meat Microbiology, Brown, M.H. (Editor), Applied Science, London, 1982, p. 406.

Kilsby, D.C. and M. E. Pugh, "The Relevance of the Distribution of Micro-organisms Within Batches of Food to the Control of Microbiological Hazards from Foods," Journal of Applied Bacteriology, 51, 1981, pp. 345-354.

National Advisory Committee on Microbiological Criteria for Foods, "HACCP Principles for Food Production," Food Safety and Inspection Service, United States Department of Agriculture, Washington, 1989.

National Advisory Committee on Microbiological Criteria for Foods, "Hazard Analysis and Criteria Control Point System," Food Safety and Inspection Service, United States Department of Agriculture, Washington, 1992.

National Research Council, 1985. *An Evaluation of the Role of Microbiological Criteria for Foods and Food Ingredients*. National Academy Press, Washington, D.C.

Prince, G., "Verification of the HACCP Program," Chapter 10 in HACCP Principles and Applications, Pierson, M.D. and D.A. Corlett (Editors), Van Nostrand Reinhold, New York, 1992, p. 92.

Ryan, T. P., *Statistical Methods for Quality Improvement*. John Wiley and Sons, Inc., New York, 1989.

Scientific Review Panel, "Review of the Pathogen Reduction: Hazard Analysis and Critical Control Point System Regulation," Office of Risk Assessment and Cost-Benefit Analysis, USDA, Washington, D.C., April 4, 1996.

Subcommittee on Microbiological Criteria, Committee on Food Protection, Food and Nutrition Board, National Research Council, An Evaluation of the Role Of Microbiological Criteria for Foods and Food Ingredients, National Academy Press, Washington, 1985, pp. 132-144.