Raw Beef Sampling Data –

Data Documentation

Overview

The data covered by this documentation are the sampling results of FSIS' raw beef products, including:

- routine microbiological sampling of raw beef trim products
- routine microbiological sampling of raw beef components products
- routine microbiological sampling of raw ground beef products
- follow-up microbiological sampling of raw beef products.

This dataset is now available in an open, non-proprietary, publicly accessible CSV format instead of XLSX, consistent with requirements set by the Foundations for Evidence-Based Policymaking Act of 2018. Additional information can be found on the FSIS Laboratory Sampling Data web page.

Data postings are part of the Agency's efforts to prevent pathogens from entering the food supply throughout the farm-to-fork continuum. Posting these datasets may help industry identify repetitive subtypes and implement control measures. It may allow researchers to identify trends to address basic research questions or to develop new diagnostics or therapies such as vaccines.

These datasets are posted for informational purposes only and are not sufficient to determine if there is an association between multiple samples. Therefore, this data should not be used to identify foodborne illness outbreaks, associate samples with foodborne illness outbreaks or determine whether two or more samples are causally related.

The FSIS number is a unique identifier for retrieving whole genome sequence data from the National Center for Biotechnology and Information (NCBI) Pathogen Detection Isolates Browser. The allele codes included in this dataset provide a convenient naming method for reporting Whole Genome Sequencing (WGS) data. Because allele codes can change over time as more WGS data becomes available, a date stamp allows the data to be used in reports.

FSIS, the Centers for Disease Control and Prevention (CDC) and other public health partners monitor WGS information gathered from samples in real time and use sophisticated epidemiological tools to identify whether the cumulative findings might signal a foodborne illness outbreak. Outbreaks identified through this process are announced to the public through CDC's website and FSIS' Outbreak Response Page. The FSIS investigative process is described in FSIS Directive 8080.3.

Two datasets will be provided: archived and current. The archived dataset will provide data starting from October 1, 2013, up to the final day of the previous fiscal year (FY) of the report execution date (e.g., the data posted in April 2023 was for the time period of October 1, 2013 through September 30, 2022). The archived dataset will be updated annually. The current dataset will provide data starting from the first day of the FY following the final date available in the archived dataset through the end of the previous fiscal quarter of the report execution date (e.g., data posted in April 2023 was for the time period of October 1, 2022 through December 31, 2022). The current dataset will be updated quarterly.

Each row in these datasets represents one sample collected and sent to an FSIS laboratory for analysis.

Beginning February 1, 2023, FSIS Expanded its routine verification testing for six Shiga toxin-producing *E. coli* (STEC) that are adulterants (non-O157 STEC; O26, O45, O103, O111, O121, or O145), in addition to the adulterant *E. coli* O157:H7 in samples of raw ground beef, bench trim, and other raw ground beef components collected at official establishments. In this documentation, the STEC group columns are displayed as [Serogroup]ColumnName, with the [Serogroup] being one of the seven STEC (O26, O45, O103, O111, O121, O145, O157). Samples received by an FSIS laboratory on or after February 1, 2023 will be tested for the additional STEC serogroups.

Data contained in this dataset on tested product from establishments are not sufficient to determine an association with human illnesses. Further epidemiologic information is needed to determine if there is an association among the non-clinical isolates and human illnesses.

Data Dictionary

- EstablishmentID
    - A unique numeric identifier that is used to identify an establishment across data tables in the FSIS databases.
- EstablishmentNumber
    - A letter/number combination uniquely identifying each establishment.
- EstablishmentName
    - The name of an establishment on its FSIS grant of inspection.
- State
    - The state where the establishment is located.
- ProjectCode
    - A short name given to easily identify an FSIS sampling project.
    - Projects in this dataset
        - MT43 - Risk-based Sampling of Raw Ground Beef or Veal Products
        - MT44 – Follow-up sampling of raw ground beef product in response to a MT43 or Agricultural Marketing Service (AMS) positive result in raw ground beef product at Federal establishments
        - MT44T – Follow-up sampling of raw ground beef, trim, or other component outside of projects MT44, MT53, and MT52 collected by IPP at Federally inspected establishments.

    - MT44T_C – Follow-up *E. coli* Sampling of Raw Ground Beef, Trimmings or Components
      - NOTE: MT44T_C replaced MT44T in February 2023
    - MT52 – Follow- up sampling at suppliers of beef manufacturing trimmings or other components from originating slaughter suppliers, in response to a MT43, MT65, or AMS ground beef positive result.
    - MT52_C – Follow-up Sampling of Suppliers of Raw Ground Beef Trim or Components
      - NOTE: MT52_C replaced MT52 in February 2023
    - MT53 – Follow-up sampling of trim or other components at the establishment that produced product in response to a MT60, MT65, MT64, or AMS trim testing positive.
    - MT53_C – Follow-up Sampling of Beef Manufacturing Trim or Other Raw Ground Beef or Beef Patty Components
      - NOTE: MT53_C replaced MT53 in February 2023
    - MT54 – *E. coli* O157:H7 Sampling of raw ground beef or beef patty components (other than trim)
    - MT55 – *E. coli* O157:H7 Sampling of bench trim intended for ground beef or other non-intact product
    - MT60 – Sampling of beef manufacturing trimmings
    - MT60_C – Sampling of beef manufacturing trimmings
      - NOTE: MT60_C replaced MT60 in February 2023
    - MT64 – Sampling of raw ground beef or beef patty components (other than trim).
      - NOTE: MT64 replaced MT54 in September 2015
    - MT65 – Sampling of bench trim for further use in any raw, non-intact beef products.
      - NOTE: MT65 replaced MT55 in September 2015
    - MT65_C – Sampling of bench trim for further use in any raw, non-intact beef products
      - NOTE: MT65_C replaced MT65 in February 2023
- ProjectName
  - The name of the FSIS sampling project.
- FormID
  - The form number used to uniquely identify a specific sample.
- CollectionDate
  - The date the FSIS inspector collected the sample at the FSIS-regulated establishment.
- SampleSource
  - The type of product collected in the sample.

- SalmonellaSpAnalysis
    - The result of the analysis for *Salmonella* species in the sample.
        - Negative = *Salmonella* was not found in the sample.
        - Positive = *Salmonella* was found in the sample.
- SalmonellaSerotype
    - The name of the distinct variation of the tested species of bacteria. A list of the serotypes that are more commonly associated with human illness can be found on the Centers for Disease Control and Prevention (CDC) web site on their National *Salmonella* Surveillance web page.
- SalmonellaPFGEPrimaryPattern
    - The specific primary pattern identified from Pulsed-Field Gel Electrophoresis (PFGE), which is the laboratory technique used to produce a deoxyribonucleic acid (DNA) fingerprint for a group of the same type of bacteria. FSIS discontinued PFGE characterization in March 2019 and replaced the technique with the more modern whole genome sequencing (WGS) characterization.
- SalmonellaPFGESecondaryPattern
    - The specific secondary pattern identified from Pulsed-Field Gel Electrophoresis (PFGE), which is the laboratory technique used to produce a deoxyribonucleic acid (DNA) fingerprint for a group of the same type of bacteria. FSIS discontinued PFGE characterization in March 2019 and replaced the technique with the more modern whole genome sequencing (WGS) characterization.
- SalmonellaAlleleCode
    - Definition: A code assigned by CDC-PulseNet based on the number of differences in pre-defined genes in the WGS data. The allele code also includes the date when FSIS retrieved the allele code from PulseNet. It is possible for PulseNet to adjust the allele code after it was retrieved. The data format is:
      allele code[space][pipe][space]mm/dd/yyyy (e.g., SALM1.0 – 1.2.3.4.5.6 | 01/01/2021). When PulseNet is unable to assign an allele code, the entry will be: Allele Code Ineligible[space][pipe][space]mm/dd/yyyy (e.g., Allele Code Ineligible | 01/01/2021).
- SalmonellaFSISNumber
    - A unique identifier for retrieving Whole Genome Sequencing (WGS) data for a *Salmonella* isolate from the National Center for Biotechnology and Information (NCBI) Pathogen Detection Isolates Browser. NCBI developed the Browser to help users learn about the sequences they contribute. NCBI has provided a video introduction to this browser, and this document contains a table outlining information available in NCBI'S Pathogen Detection Isolates Brower for additional reference.
- SalmonellaAMRProfileAST
    - The antimicrobial resistance profile of the antimicrobial drugs phenotypically tested to which isolates are found to be resistant using the National Antimicrobial Resistance Monitoring System (NARMS) panel 5. The Food and Drug Administration (FDA) in its Guidance 152 classified antimicrobial drugs based on importance of the drug to human

medicine. Isolates displaying resistance to multiple antimicrobial drugs tested by the NARMS testing panel are classified according to the antimicrobial drug(s) with the highest classification of risk. A resistance profile that is "pan-susceptible" means that the isolate is not resistant to any of the antimicrobial drugs tested. See the FDA Antimicrobial drug classification table in this document.

- SalmonellaAMRProfileWGS
    - The antimicrobial resistance profile that was predicted based on the resistance genotype determined by NCBI using AMRfinder Plus, NCBI's reference gene catalog and empirical data. Genes with partial coverage or mistranslations were not considered when predicting phenotype, and only drugs on the NARMS testing panel are considered for the profile. Additionally, the prediction is based on the presence/absence of genes and does not account for gene expression.
- STECAnalysis
    - The result of the analysis for *STEC* in the sample.
        - Negative = *STEC* was not found in the sample
        - Positive = *STEC* was found in the sample.
- [Serogroup]Result
    - The result of the analysis for *E. coli* OXXX in the sample.
        - Negative = *E. coli* OXXX was not found in the sample
        - Positive = *E. coli* OXXX was found in the sample.
            - All samples with a positive *E. coli* OXXX result were either prevented from going into commerce due to test and hold or were recalled.
- [Serogroup]PFGEPrimaryPattern
    - The specific primary pattern identified from Pulsed-Field Gel Electrophoresis (PFGE), which is the laboratory technique used to produce a deoxyribonucleic acid (DNA) fingerprint for a group of the same type of bacteria. FSIS discontinued PFGE characterization in March 2019 and replaced the technique with the more modern whole genome sequencing (WGS) characterization.
- [Serogroup]PFGESecondaryPattern
    - The specific secondary pattern identified from Pulsed-Field Gel Electrophoresis (PFGE), which is the laboratory technique used to produce a deoxyribonucleic acid (DNA) fingerprint for a group of the same type of bacteria. FSIS discontinued PFGE characterization in March 2019 and replaced the technique with the more modern whole genome sequencing (WGS) characterization.
- [Serogroup]AlleleCode
    - A code assigned by CDC-PulseNet based on the number of differences in pre-defined genes in the WGS data. The allele code also includes the date when FSIS retrieved the allele code from PulseNet. It is possible for PulseNet to adjust the allele code after it was retrieved. The data format is:

allele code[space][pipe][space]mm/dd/yyyy (e.g., EC1.0 – 1.2.3.4.5.6 | 01/01/2021). When PulseNet is unable to assign an allele code, the entry will be: Allele Code Ineligible[space][pipe][space]mm/dd/yyyy (e.g., Allele Code Ineligible | 01/01/2021).

- [Serogroup]FSISNumber
    - A unique identifier for retrieving WGS data for a *E. coli* isolate from the NCBI Pathogen Detection Isolates Browser. NCBI developed the Browser to help users learn about the sequences they contribute. NCBI has provided a video introduction to this browser, and this document contains a table outlining information available in NCBI'S Pathogen Detection Isolates Brower for additional reference.
    - this document contains a table outlining information available in NCBI'S Pathogen Detection Isolates Brower for additional reference.
- [Serogroup]AMRProfileAST
    - The antimicrobial resistance profile of the antimicrobial drugs phenotypically tested to which isolates are found to be resistant using the NARMS panel 5. FDA in its Guidance 152 classified antimicrobial drugs based on importance of the drug to human medicine. Isolates displaying resistance to multiple antimicrobial drugs tested by the NARMS testing panel are classified according to the antimicrobial drug(s) with the highest classification of risk. A resistance profile that is "pan-susceptible" means that the isolate is not resistant to any of the antimicrobial drugs tested. See the FDA Antimicrobial drug classification table in this document.
- [Serogroup]AMRProfileWGS
    - The antimicrobial resistance profile that was predicted based on the resistance genotype determined by NCBI using AMRfinder Plus, NCBI's reference gene catalog and empirical data. Genes with partial coverage or mistranslations were not considered when predicting phenotype, and only drugs on the NARMS testing panel are considered for the profile. Additionally, the prediction is based on the presence/absence of genes and does not account for gene expression.

Relationship to Other Data

This data can be combined with other FSIS datasets using the EstablishmentID variable.

Notes and Limitations

Information about FSIS sampling laboratories and procedures can be found on the FSIS website on the Laboratories & Procedures web page and the Microbiology Laboratory Guidebook (MLG) web page.

Prior to February 1, 2023, samples that have a project code of MT43, MT54, MT55, MT64, and MT65 were only tested for *E. coli* O157:H7. Therefore samples received prior to February 1, 2023 will have NULL values in the fields for [O26, O45, O103, O111, O121, O145]Result, PFGEPrimaryPattern, PFGESecondaryPrimaryPattern, AlleleCode, FSISNumber, AMRProfileAST, and AMRProfileWGS if the test was not performed.

FSIS switched to the Cloth Sampling Method for the following project codes in February 2023: MT44T, MT52, MT53, MT60, MT65. Those project codes were retired and replaced with the following project codes: MT44T_C, MT52_C, MT53_C, MT60_C, and MT65_C. For more information, please see FSIS Directive 10010.1.

NULL values indicate that the specific variable is not available for that record.

FSIS added *Salmonella* analysis to the MT55 and MT60 sampling projects in July 2014.

When a sample screens positive for *Salmonella*, there normally is only one isolate (i.e., subtype determined using serology [serotype], PFGE, antibiotic resistance, allele code, or whole genome sequence) derived from laboratory confirmation procedures. Prior to streaking on agar, enrichment broths are tested using rapid screening technology, and if they do not test negative on that analysis, they are then struck to agar plates. During the confirmation process, lab staff will pick up to three typical isolated colonies before reporting as Negative. Generally, only one confirmed positive colony will be carried on to characterization analysis. On very rare occasions, more than one typical colony may be carried on for characterization analysis. In such circumstances, the multiple isolate data (e.g., *Salmonella* serotype) are separated by a semicolon.

Confirmed positive bacterial isolates can become unsuitable for additional testing during shipping, transfer, or storage. This result is final as further characterization cannot be performed. When this occurs, the isolate characterization result fields (e.g., serotype, antimicrobial resistance, allele code) will be populated with "Final result – characterization not available" and the FSIS Number will be reported as "FSISNOTSEQ".

WGS data must be interpreted within the context of how it will be used, e.g., to detect outbreaks or contamination events. Additional corroborating information, including case-patient food exposure and product distribution records, may be necessary to properly interpret the WGS data.

Prior Analysis

Prior analysis using this data can be found on the FSIS website, such as the Microbiological Testing Program for *Escherichia coli* O157:H7 and non-O157 Shiga toxin-producing *Escherichia coli* (STEC) web page.

FDA's Antimicrobial Drug Classification According to Their Importance to Human Medicine

Antimicrobial Class	Antimicrobial Drug	Abbreviation	FDA Classification
1st Generation Cephalosporins (Cephems)	Cephalothin (Cefazolin)	CEP	Important
3rd Generation Cephalosporins (Cephems)	Ceftiofur	TIO	Critically Important
	Ceftriaxone	AXO	Critically Important
Aminoglycosides	Amikacin	AMI	Highly Important
	Apramycin	APR	Highly Important[1]
	Gentamicin	GEN	Highly Important
	Kanamycin	KAN	Highly Important
	Streptomycin	STR	Highly Important
B-Lactam/B-Lactamase Inhibitor Combinations	Amoxicillin - Clavulanic Acid (Amoxicillin)	AUG	Highly Important
Carbapenems	Imipenem	---	Highly Important
Carboxypenicillins	Ticarcillin	TIC	Highly Important
Cephamycins (Cephems)	Cefoxitin	FOX	Important
Fluoroquinolones	Ciprofloxacin	CIP	Critically Important
Folate Pathway Inhibitors	Sulfamethoxazole (1998-2003)	SMX	Not Classified
	Sulfisoxazole (2004-2009)	FIS	Not Classified
	Trimethoprim-Sulfamethoxazole	COT	Critically Important
Macrolides	Azithromycin	AZI	Critically Important
	Erythromycin	ERY	Critically Important
Phenicols	Chloramphenicol	CHL	Highly Important
	Florfenicol	FFN	Highly Important[1]
Quinolones	Nalidixic Acid	NAL	Important
Ketolides	Telithromycin	TEL	Not Classified
Lincosamides	Clindamycin	CLI	Highly Important
Penicillins	Ampicillin	AMP	Highly Important
Tetracyclines	Tetracyclines	TET	Highly Important

Critically Important (C): Antimicrobial drugs which meet BOTH criteria 1 and 2 in Appendix A of the FDA Guidance for Industry #152 are considered critically important to human medical therapy.

Highly Important (H): Antimicrobial drugs which meet EITHER criteria 1 or 2 in Appendix A of the FDA Guidance for Industry #152 are considered highly important to human medical therapy.

Important (I): Antimicrobial drugs which meet EITHER criterion 3 and/or 4 and/or 5 in Appendix A of the FDA Guidance for Industry #152 are considered important to human medical therapy.

Not Classified (NC): Antimicrobial drugs which are not given a classification in FDA's Guidance for Industry #152 (dated October 23, 2003).

[1] Where noted, FSIS has classified drugs approved for animal use only using the same classification that FDA has designated for drugs in the same antimicrobial class that are approved for human use.

Information Available in NCBI's Pathogen Detection Isolates Browser

Single Nucleotide Polymorphism (SNP) cluster	A SNP cluster is a group of isolates whose genomes are closely related. In the Pathogen Detection Browser this element will contain a link that opens a new page with information about closely related sequences in the database.
Min-same or Min-diff	Minimum SNP distance from the query isolate to one of the same or a different isolation type. Isolation types are clinical (including human or animal) or environmental (including food).
BioSample	Further information (metadata) pertaining to the sample from which the sequence was isolated.
Assembly	Technical information pertaining to the sequence.
AMR Genotype	Information pertaining to antimicrobial resistant (AMR) genes found in the isolate sequence. Additional information about each AMR gene in this field is provided by a Reference Gene Catalog. Note: Empty cells do not necessarily indicate a lack of AMR genes.