**Outbreak of *Salmonella* Typhimurium in the UK**

Case-control study analysis: Part One

# Pre-module exercise

## Developed by:

Ranya Mulchandani (C2019) and Alicia Barrasa Blanco (Scientific Coordinator)

## April 2021

# Background

The scenario presented in this case study is based on an investigation of a cluster of *Salmonella* Typhimurium cases identified in the UK through the analysis of whole genome sequencing data in 2020.

The information presented in this case study and the associated data files have been deliberately changed to facilitate the acquisition of the learning objectives.

The aim of this case study is to analyse the data collected as part of a foodborne outbreak investigation, using either Stata or R software.

The case study is formed of two parts. The pre-module exercise is the first of the two parts.

# Learning outcomes

By the end of this exercise, participants should be able to:

1. Analyse data from a foodborne outbreak investigation;
2. Explore the potential role of several food vehicles using stratified analysis.

## Prerequisites

Participants are expected to be familiar with data management and basic analysis in either STATA or R.

# Part 1: Outbreak identification and study design

In Week 1 of 2020, 2 cases of genetically identical (and distinct from all other isolates) *Salmonella* Typhimurium were identified in the UK through routine phylogenetic analysis. In March, reporting of cases associated at the 5-SNP level increased, despite a UK-wide national lockdown in response to the COVID-19 pandemic. By 14th April 2020, 37 *Salmonella* Typhimurium cases were associated with this cluster.

On 29th April 2020, an incident management team (IMT) meeting was held to discuss the cases and guide further investigations. Subsequently, two sub-groups of the IMT were formed for overseeing food-chain and epidemiological investigations. The first meeting of the epidemiological sub-group was held on 6th May 2020.

## **Question 1: What information would you request at the epidemiological sub-group meeting?**

At the epidemiological sub-group meeting you are provided with a brief description of cases so far identified as part of this cluster.

As of 6th May 2020, 50 cases had been linked to this cluster, with sample dates from 02 August 2019 to 18 April 2020. There are equal numbers of males and females affected, the median age is 41 and the age range is 2-89 years. Cases have been identified across England (n=41), Wales (n=3), Scotland (n=3) and Northern Ireland (n=3). Additionally, there have been reports of six cases outside of the UK (in two European countries) with isolates genetically related to the outbreak strain. The source of the outbreak is currently unknown.

The following case definitions were agreed upon:

* **Confirmed case:** A laboratory confirmed case of *Salmonella* Typhimurium with PHE SNP address designated as 1.222.503.919.5052.6145.% (t5.6145), resident in the UK.
* **International case:** A laboratory confirmed case of *Salmonella* Typhimurium not resident in the UK and with an isolate within 2 allelic differences by cgMLST from the UK representative strain, or a PHE SNP address designated as 1.222.503.919.5052.6145.% (t5.6145).
* **Probable case:** A laboratory confirmed case of *Salmonella* Typhimurium resident in the UK and with an isolate within 2 allelic differences by cgMLST from the UK representative strain awaiting PHE SNP-typing results.

## **Question 2: What investigations would you recommend at this stage?**

Case interviews with an exploratory (hypothesis-generating) questionnaire were initiated by the national gastrointestinal infections team. A further epidemiological sub-group meeting was held on 26th May 2020. The results from the exploratory questionnaires are discussed at this meeting. Interviews had been held with thirteen of the cases. The food types reported by the thirteen cases are presented in Table 1.

**Table 1: Summary of food products reports from exploratory/semi-exploratory interviews**

|  |  |  |  |
| --- | --- | --- | --- |
| Food type | Number of cases (n=13) | Proportion of cases (%) | Further detail |
| Chicken | 12 | 92% | Fresh unprocessed supermarket chicken (9) |
| Salad/vegetables | 11 | 85% | Lettuce (7)  Peppers (7)  Tomatoes (6)  Peas (6)  Onions (6) |
| Nuts | 10 | 77% | Nut bars (9)  Almonds (7)  Brazil nuts (6)  Nut bars from Producer A (5) |
| Fish/seafood | 9 | 69% | Prawns (5)  Breaded fish (3) |
| Beef | 8 | 62% | Minced beef (6) |
| Eggs | 8 | 62% |  |

## **Question 3: What further investigations would you recommend considering the new information?**

Several food items were identified as potential vehicles of infection. The primary hypothesis was the consumption of nuts and nut bars, as the proportion of cases reported to have consumed these products was higher than would be expected (routine food exposure data had been reviewed elsewhere). Secondary hypotheses include chicken and fish with sauce and salads. A decision was made to proceed with a case-control study to test these hypotheses using an analytical epidemiological approach.

On 12th June 2020, a case-control study was initiated. A study specific questionnaire was designed to test the hypotheses, and included questions on demographic information and food exposures. Food exposures in the 7-day exposure period were examined in two ways: firstly, an overall exposure to a food group (e.g. any vegetable) and secondly, exposures to specific items within the food group (e.g. carrots).

Controls were recruited via a market research panel. Controls were asked for their food exposures between 18 and 25 June 2020, while cases were asked about their exposures in the 7-days prior to symptom onset. Cases were eligible for inclusion if they met the outbreak case definition for a confirmed case, had a date of symptom onset after 01 April 2020 and had not done a hypothesis-generating questionnaire. Case questionnaires were conducted on the phone, whereas questionnaires were sent out via the market research panel for controls.

## **Question 4: What would you plan of analysis be?**

## **Plan of Analysis:**

* **Perform data cleaning**
* For each variable, look at the range, unexpected and missing values
* Correct data using the original forms used if needed
* **Describe each variable**
* For each variable, describe frequency distributions including missing values and, if needed, means, median, modes, quartiles, SD, outliers
* Make appropriate histograms and box plots
* Choose relevant characteristics to describe the population
* **Identify the outbreak vehicle if any**
* Chose the appropriate measure of association
* Chose the appropriate statistical tests and appropriate level of confidence
* Calculate food-specific attack rates
* Look at the proportions of cases exposed
* Calculate the percentages of cases exposed to each exposure
* Search for any dose response if appropriate
* Interpret the results
* **Perform a stratified analysis**
* Identify the variables that are potential effect modifiers (EM) and confounders
* Design appropriate stratification tables
* Stratify on each level taken by the EM and confounders
* Compute appropriate measurements to identify confounding and effect modification
* Apply appropriate statistical tests
* Interpret the results
* **Perform a multivariable analysis**
* This will be discussed during the module (part two), so hold your horses.

# Part 2: Descriptive and univariate analysis

A dataset containing all the information for analysis can be found in **salmv1.4.csv**. For convenience, the following table summarises the variables it contains:

**Table 2: Summary table of variables collected in the study dataset.**

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Type | Code | Definition |
| case | Binary | 1 = case, 0 = control | Case or control |
| gender | Binary | 1 = male, 0 = female |  |
| age | Numeric |  | Age in years |
| grp\_takeaway | Binary | 1 = yes, 0 = no | Consumption of takeaway (any) |
| grp\_chicken | Binary | 1 = yes, 0 = no | Consumption of chicken (any) |
| grp\_nut­seed | Binary | 1 = yes, 0 = no | Consumption of nuts or seeds (any) |
| grp\_vegetables | Binary | 1 = yes, 0 = no | Consumption of vegetables (any) |
| grp\_fruit | Binary | 1 = yes, 0 = no | Consumption of fruit (any) |
| grp\_cereal | Binary | 1 = yes, 0 = no | Consumption of products containing cereal (any) |
| grp\_meatfish | Binary | 1 = yes, 0 = no | Consumption of meat or fish (any) |
| frozenchick | Binary | 1 = yes, 0 = no | Consumption of fresh or frozen chicken |
| filletchick | Binary | 1 = yes, 0 = no | Consumption of chicken fillets |
| cabbage | Binary | 1 = yes, 0 = no | Consumption of raw cabbage (e.g. in salads) |
| carrot | Binary | 1 = yes, 0 = no | Consumption of raw carrots |
| dressings | Binary | 1 = yes, 0 = no | Consumption of dressings/dips or sauces (e.g. tahini, Caesar, tartar, hummus) |
| macadamia | Binary | 1 = yes, 0 = no | Consumption of macadamia nuts |
| brazil | Binary | 1 = yes, 0 = no | Consumption of brazil nuts |
| hazelnut | Binary | 1 = yes, 0 = no | Consumption of hazelnuts |
| cashew | Binary | 1 = yes, 0 = no | Consumption of cashews |
| pecan | Binary | 1 = yes, 0 = no | Consumption of pecans |
| peanut | Binary | 1 = yes, 0 = no | Consumption of peanuts |
| pistachio | Binary | 1 = yes, 0 = no | Consumption of pistachios |
| almond | Binary | 1 = yes, 0 = no | Consumption of almonds |
| nut\_bar\_a | Binary | 1 = yes, 0 = no | Consumption at least one of Brand A nut bars (each bar contained different combinations of fruit and nuts) |
| nut\_bar\_b | Binary | 1 = yes, 0 = no | Consumption of at least one of Brand B nut bars (each bar contained different combinations of fruit and nuts) |

## **Question 5: What are the main characteristics of the study population?**

There is a total of 26 cases and 113 controls in the dataset who completed questionnaires. Age and gender are summarised in Table 3.

**Table 3: Summary characteristics of study population**

|  |  |  |
| --- | --- | --- |
|  | Case | Control |
| Count | 26 | 113 |
| Age |  |  |
| Median | 57 | 52 |
| Mean | 53 | 50 |
| Interquartile range (IQR) | 46 – 70 | 40 – 59 |
| Range | 5 – 76 | 20 – 77 |
| Gender |  |  |
| Female | 15 (58%) | 56 (50%) |
| Male | 11 (42%) | 57 (50%) |

## **Question 6: What are the food-specific attack rates for the food group variables? What is the appropriate measure of association in this study?**

In a case-control study, the appropriate measure of association is an odds ratio. Food-specific attack rates and odds ratios for the food groups are available in Table 4.

**Table 4: Food-specific attack rates for food group exposures**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Case | Control | OR (95% CI) | p-value |
| Grouped exposure categories |  |  |  |  |
| Nuts & seeds | 23 (88%) | 56 (50%) | 7.8 (2.2 – 27.5) | <0.001 |
| Cereal & flour containing products | 26 (100%) | 86 (76%) | Undefined (not estimable) | 0.005 |
| Meat, fish, eggs | 26 (100%) | 103 (91%) | Undefined (not estimable) | 0.115 |
| Salads & vegetables | 26 (100%) | 107 (95%) | Undefined (not estimable) | 0.230 |
| Fruit | 26 (100%) | 107 (95%) | Undefined (not estimable) | 0.230 |
| Chicken | 18 (69%) | 83 (73%) | 0.8 (0.3 – 2.1) | 0.667 |
| Takeaway | 9 (35%) | 35 (31%) | 1.1 (0.5 – 2.8) | 0.807 |

Following review of the food group variables, consumption of nuts and seeds appeared to be associated with illness (OR = 7.8, 95% CI 2.2 – 27.5, p-value <0.001). It was not possible to calculate ORs for salads, fruit, meat or cereal products due to the universal exposure of this items among cases.

## **Question 7: What would you next steps of analysis be?**

Each specific food item exposure was further analysed (additional variables were available in the original dataset, however for simplicity, only a sub-set have been made available in this exercise); results can be found in Table 5.

**Table 5: Food-specific attack rates for individual-level variables**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Case | Control | OR (95% CI) | p-value |
| Chicken |  |  |  |  |
| Fresh or frozen chicken | 14 (54%) | 54 (48%) | 1.3 (0.5 – 3.3) | 0.577 |
| Chicken fillets | 6 (23%) | 27 (24%) | 1.0 (0.3 – 2.8) | 0.930 |
| Vegetables |  |  |  |  |
| Cabbage | 10 (39%) | 8 (7%) | 8.2 (2.5 – 27.4) | <0.001 |
| Carrots | 17 (65%) | 22 (20%) | 7.8 (2.6 – 22.4) | <0.001 |
| Nuts |  |  |  |  |
| Nut Bar Brand A | 11 (42%) | 4 (4%) | 20.0 (4.9 – 93.8) | <0.001 |
| Brazil nuts | 14 (54%) | 11 (10%) | 10.8 (3.6 – 32.7) | <0.001 |
| Cashews | 12 (46%) | 15 (13%) | 5.6 (1.9 – 15.8) | <0.001 |
| Hazelnuts | 11 (43%) | 13 (12%) | 5.6 (1.9 – 16.4) | <0.001 |
| Nut Bar Brand B | 5 (19%) | 2 (2%) | 13.2 (2.0 – 143.2) | <0.001 |
| Peanuts | 13 (50%) | 22 (19%) | 4.1 (1.5 – 11.2) | 0.001 |
| Pecans | 5 (19%) | 6 (5%) | 4.3 (0.9 – 18.2) | 0.018 |
| Macadamia | 3 (12%) | 3 (3%) | 4.8 (0.6 – 37.4) | 0.044 |
| Pistachios | 4 (15%) | 7 (6%) | 2.8 (0.5 – 11.9) | 0.118 |

## **Question 8: What is your interpretation of findings from the univariate analysis?**

In the univariate analysis among the individual nut variables, brazil nuts, cashews, hazelnuts and peanuts were all statistically significantly more likely to have been consumed by cases than controls (p<0.001) in addition to two brands of nuts bars (Brand A – OR: 20.0, 95% CI 4.9 – 93.8 and Brand B – OR: 13.2, 95% CI 2.0 – 143.2). It was discovered that both Brand A and B were produced by the same manufacturer!

Additionally, in the univariate analysis among the chicken and vegetable variables, consumption of cabbage and carrots appear to also be statistically significantly more likely to have been consumed by cases than controls (p<0.001).

## **Question 9: What would your next steps be? How would you further assess potential vehicle(s) of infection?**

For your convenience, you can use the table below which will help you to identify if Nut bar Brand A is a potential confounder and/or effect modifier.

**Table 6: Stratified analysis**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | OR stratified by Nut Bar Brand A [95% CI] | | M-H Test for homogeneity (p-value) | Crude OR  [95% CI] | Adjusted (M-H) OR  [95% CI] | % change | Is Nut Bar Brand A an effect modifier or a confounder? |
|  | **Yes** | **No** |  |  |  |  |  |
| Fresh or frozen chicken |  |  |  |  |  |  |  |
| Chicken fillets |  |  |  |  |  |  |  |
| Cabbage |  |  |  |  |  |  |  |
| Carrots |  |  |  |  |  |  |  |
| Brazil nuts |  |  |  |  |  |  |  |
| Peanuts |  |  |  |  |  |  |  |
| Hazelnuts |  |  |  |  |  |  |  |
| Cashews |  |  |  |  |  |  |  |
| Nut Bar Brand B |  |  |  |  |  |  |  |
| Pecans |  |  |  |  |  |  |  |
| Macadamia |  |  |  |  |  |  |  |
| Pistachios |  |  |  |  |  |  |  |

The results from the stratified analysis will be discussed in the **MVA module.**

## About Public Health England

Public Health England exists to protect and improve the nation’s health and wellbeing, and reduce health inequalities. We do this through world-class science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. We are an executive agency of the Department of Health, and are a distinct delivery organisation with operational autonomy to advise and support government, local authorities and the NHS in a professionally independent manner.

Public Health England  
Wellington House

133-155 Waterloo Road  
London SE1 8UG  
Tel: 020 7654 8000  
[www.gov.uk/phe](http://www.gov.uk/phe)   
Twitter: [@PHE\_uk](https://twitter.com/PHE_uk)

Facebook: [www.facebook.com/PublicHealthEngland](http://www.facebook.com/PublicHealthEngland)

© Crown copyright 2020  
You may re-use this information (excluding logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit [OGL](https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/) or email [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk). Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

