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

Case-control study analysis: Part One

# Pre-module exercise

# STATA HelpGuide

## 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. This Stata guide accompanies the full pre-course exercise word document and provide STATA help for section 2 only.

# 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 STATA.

# Part 2: Descriptive and univariate analysis

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

## Understanding your data

Import using the csv dataset into STATA using the import command and make sure you know what your working directory is.

import delimited "C:\YourWorkingDirectory\salm.csv", clear

To see the information, contain on the data set use the commands describe and codebook

## Manging your data

To create new variables required for the analysis - primarily a variable for age group - and record gender use the commands generate and replace. To create labels for the variable content use the commands label define and label values.

generate agegr = .

replace agegr = 0 if age <35

replace agegr = 1 if age >=35

replace agegr = 2 if age >=45

replace agegr = 3 if age >=55

replace agegr = 4 if age >=65

replace agegr = . if age == .

label define agegrLab 0 "0-34" 1 "35-44" 2 "45-54" 3 "55-64" 4 "64-more"

label values agegr genderLab

label define genderLab 0 "Males" 1 "Females"

label values gender genderLab

## Describing your data

To describe the main characteristic of the study population and reproduce table 3 use the commandstabulate*,* andsummarize.

summarize age, d

bysort case: sum age, d

tabulate gender case, col

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

To identify calculate the food-specific attack rates and reproduce tables 4 and 5 you need to perform the univariate analysis using the command cc or cctable (from the cctable ado.file). Here help for cctable is shown

Food-specific attack rates for food groups

cctable case grp\_takeaway grp\_chicken grp\_nutseed grp\_vegetables grp\_fruit grp\_cereal grp\_meatfish

Cases Controls

----------------------------------------------------------------------------------

Exposure Total Exposed % Total Exposed % Odds Ratio p

----------------------------------------------------------------------------------

grp\_nutseed 26 23 88.46 113 56 49.56 7.80 [2.15-42.33] 0.000

grp\_cereal 26 26 100.00 113 86 76.11 . [2.08-.] 0.005

grp\_meatfish 26 26 100.00 113 103 91.15 . [0.64-.] 0.115

grp\_fruit 26 26 100.00 113 107 94.69 . [0.36-.] 0.230

grp\_vegetabl 26 26 100.00 113 107 94.69 . [0.36-.] 0.230

grp\_chicken 26 18 69.23 113 83 73.45 0.81 [0.30-2.40] 0.663

grp\_takeaway 26 9 34.62 113 35 30.97 1.12 [0.40-2.97] 0.807

----------------------------------------------------------------------------------

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

Food-specific attack rates for individual food items

cctable case frozenchick filletchick cabbage carrot macadamia brazil hazelnut cashew pecan peanut pistachio nut\_bar\_a nut\_bar\_b

Cases Controls

----------------------------------------------------------------------------------

Exposure Total Exposed % Total Exposed % Odds Ratio p

----------------------------------------------------------------------------------

nut\_bar\_a 26 11 42.31 113 4 3.54 19.98 [4.94-93.75] 0.000

brazil 26 14 53.85 113 11 9.73 10.82 [3.58-32.65] 0.000

carrot 26 17 65.38 113 20 17.70 8.78 [3.12-25.34] 0.000

cabbage 26 10 38.46 113 8 7.08 8.20 [2.45-27.40] 0.000

hazelnut 26 11 42.31 113 13 11.50 5.64 [1.89-16.44] 0.000

nut\_bar\_b 26 5 19.23 113 2 1.77 13.21 [1.95-143.21] 0.000

cashew 26 12 46.15 113 23 20.35 3.35 [1.23-8.99] 0.006

pecan 26 5 19.23 113 6 5.31 4.25 [0.92-18.23] 0.018

peanut 26 13 50.00 113 31 27.43 2.65 [1.00-6.92] 0.026

macadamia 26 3 11.54 113 3 2.65 4.78 [0.59-37.43] 0.044

pistachio 26 4 15.38 113 7 6.19 2.75 [0.54-11.85] 0.118

frozenchick 26 14 53.85 113 54 47.79 1.27 [0.50-3.30] 0.577

filletchick 26 6 23.08 113 27 23.89 0.96 [0.28-2.80] 0.930

----------------------------------------------------------------------------------

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

To fill in table 6 you need to perform stratified analysis using either the cc command with the option ,by(thirdvariable) or using the ccinter command (from the ccinter ado.file) also with the option ,by(thirdvariable).

Effect of cabbage consumption stratified by the consumption of Nut Bar Brand A, example of potential confounding

ccinter case cabbage, by(nut\_bar\_a)

Number of obs = 139 , Missing = 0

nut\_bar\_a = Exposed

------------------------------+

cabbage Cases Controls|

------------------------------| Odds Ratio 3.60 [0.19-219.89]

Exposed 6 1 | Attrib.risk.exp 0.72 [-4.24-1.00]

UnExposed 5 3 | Attrib.risk.pop 0.39

------------------------------+

Total 11 4

Exp % 55% 25%

nut\_bar\_a = Unexposed

------------------------------+

cabbage Cases Controls|

------------------------------| Odds Ratio 5.30 [0.96-24.65]

Exposed 4 7 | Attrib.risk.exp 0.81 [-0.04-0.96]

UnExposed 11 102 | Attrib.risk.pop 0.22

------------------------------+

Total 15 109

Exp % 27% 6%

Test of Homogeneity (M-H) : pvalue : 0.7902453

Crude OR for cabbage : 8.20 [2.45-27.40]

MH OR for cabbage adjusted for nut\_bar\_a : 4.71 [1.35-16.44]

Adjusted/crude relative change : -42.64 %

Effect of cashew consumption stratified by the consumption of Nut Bar Brand A, example of a result that’s neither confounding nor effect modification

ccinter case cashew, by(nut\_bar\_a)

Number of obs = 139 , Missing = 0

nut\_bar\_a = Exposed

------------------------------+

cashew Cases Controls|

------------------------------| Odds Ratio 2.50 [0.13-157.06]

Exposed 5 1 | Attrib.risk.exp 0.60 [-6.47-0.99]

UnExposed 6 3 | Attrib.risk.pop 0.27

------------------------------+

Total 11 4

Exp % 45% 25%

nut\_bar\_a = Unexposed

------------------------------+

cashew Cases Controls|

------------------------------| Odds Ratio 3.46 [0.95-12.15]

Exposed 7 22 | Attrib.risk.exp 0.71 [-0.06-0.92]

UnExposed 8 87 | Attrib.risk.pop 0.33

------------------------------+

Total 15 109

Exp % 47% 20%

Test of Homogeneity (M-H) : pvalue : 0.8185857

Crude OR for cashew : 3.35 [1.23-8.99]

MH OR for cashew adjusted for nut\_bar\_a : 3.25 [1.16-9.13]

Adjusted/crude relative change : -3.13 %

Effect of Brazil nuts consumption stratified by the consumption of Nut Bar Brand A, example of effect modification

. ccinter case brazil, by(nut\_bar\_a)

Number of obs = 139 , Missing = 0

nut\_bar\_a = Exposed

------------------------------+

brazil Cases Controls|

------------------------------| Odds Ratio 0.83 [0.04-15.78]

Exposed 5 2 | Attrib.risk.exp 0.17 [-14.78-0.96]

UnExposed 6 2 | Attrib.risk.pop 0.08

------------------------------+

Total 11 4

Exp % 45% 50%

nut\_bar\_a = Unexposed

------------------------------+

brazil Cases Controls|

------------------------------| Odds Ratio 16.67 [4.07-69.06]

Exposed 9 9 | Attrib.risk.exp 0.94 [0.75-0.99]

UnExposed 6 100 | Attrib.risk.pop 0.56

------------------------------+

Total 15 109

Exp % 60% 8%

Test of Homogeneity (M-H) : pvalue : 0.0209197

Crude OR for brazil : 10.82 [3.58-32.65]

MH OR for brazil adjusted for nut\_bar\_a : 6.41 [2.15-19.17]

Adjusted/crude relative change : -40.71 %

The results from the stratified analysis will be discussed in the **MVA module** (part two of this case study).

## 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.

