

OUTBREAK INVESTIGATION MODULE

**From data analysis to communication of findings**

# CASE STUDY (R)

**An outbreak of**

***Campylobacter* jejuni in Greece**

**6-10 December 2021,**

**virtual**

Case study: Version 12.0 – December 2021

Authors: Ioannis Karagiannis

Based on an outbreak investigation conducted by I. Karagiannis, T. Sideroglou *et al*.

Copyright and License

**DISCLAIMER**

The information presented in this exercise and the associated data files have been deliberately changed so as to facilitate the acquisition of the learning objectives (among others: descriptive epidemiology, univariate analysis, stratified analysis, effect modification, confounding) for EPIET/UK FETP/PAE/EUPHEM/MediPIET fellows. The background story, data, results, conclusions and recommendations deriving from this exercise do not necessarily match the ones suggested by the investigators at the time of the outbreak investigation.

### Source:

This case study was first designed and written by Ioannis Karagiannis for the training needs of EPIET, PAE, NorFETP and Austrian FETP fellows of Cohort 16 (2010).

### Revisions:

December 2011: Major expansion of background and rationale; addition of preliminary questions; addition of explanation of variables; added help for tasks of descriptive analysis; expansion of the help in the univariable analysis; major expansion of the help provided for the stratified analysis.

November 2012: Breakdown of background to more questions to facilitate learning; addition of a table and a map for attack rates by municipality and by age group; renamed variable “gender” to “sex” to indicate biological sex; added IDs to dataset; creation and addition of variable “well” to teach confounding; minor changes in the phrasing of the tasks throughout; addition of two-by-two tables for univariate analysis.

November 2013: Minor clarifications in the background; change of the wording from “case-cohort” to “case-control” throughout; clarifications in the help provided throughout.

December 2014: None.

December 2015: Minor clarifications in the background; addition of expected learning outcomes; addition of loops in Stata; addition of an information bubble on user-written commands; minor stylistic improvements throughout.

December 2016: Removal of three two-by-two tables; addition of answers on the presence of effect modification/confounding in Table 6; correction of typos.

December 2021: minor revision; correction of typos

### You are free:

* **to Share** — to copy and distribute the work
* **to Remix** — to adapt and build upon the material

### Under the following conditions:

* **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). The best way to do this is to keep as it is the list of contributors: sources, authors and reviewers.
* **Share Alike** — If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one. Your changes must be documented. Under that condition, you are allowed to add your name to the list of contributors.
* **Notification** — If you use the work in the manner specified by the author or licensor, notify Ioannis Karagiannis ([ioannis.karagiannis@phe.gov.uk](mailto:Ioannis.karagiannis@phe.gov.uk)).
* You cannot sell this work alone but you can use it as part of a teaching.

### With the understanding that:

* **Waiver** — Any of the above conditions can be [**waived**](http://creativecommons.org/licenses/by-sa/3.0/) if you get permission from the copyright holder.
* **Public Domain** — Where the work or any of its elements is in the [**public domain**](http://wiki.creativecommons.org/Public_domain)under applicable law, that status is in no way affected by the license.
* **Other Rights** — In no way are any of the following rights affected by the license:
  + Your fair dealing or [**fair use**](http://wiki.creativecommons.org/Frequently_Asked_Questions#Do_Creative_Commons_licenses_affect_fair_use.2C_fair_dealing_or_other_exceptions_to_copyright.3F) rights, or other applicable copyright exceptions and limitations;
  + The author's [**moral**](http://wiki.creativecommons.org/Frequently_Asked_Questions#I_don.E2.80.99t_like_the_way_a_person_has_used_my_work_in_a_derivative_work_or_included_it_in_a_collective_work.3B_what_can_I_do.3F) rights;
  + Rights other persons may have either in the work itself or in how the work is used, such as [**publicity**](http://wiki.creativecommons.org/Frequently_Asked_Questions#When_are_publicity_rights_relevant.3F) or privacy rights.
* **Notice** — For any reuse or distribution, you must make clear to others the license terms of this work by keeping together this work and the current license.

**This licence is based on** [**http://creativecommons.org/licenses/by-sa/3.0/**](http://creativecommons.org/licenses/by-sa/3.0/)

**Part I: Objectives**

At the end of the session, participants should be able to:

* describe the steps followed in an outbreak investigation
* undertake descriptive epidemiological investigations in R
* describe an outbreak by time, place and person
* carry out univariate analysis for a case-control study in R
* compute and interpret unadjusted odds ratios in an outbreak investigation

## Introduction

**Scenario**: On 3 June 2009, a local Public Health office in Greece informed the national Centre for Diseases Control and Prevention (HCDCP) in Athens about an unusual increase in *C. jejuni* cases among children residing in rural areas around the town. The local hospital informed HCDCP that 31 stool samples among children <15 years of age had been tested positive for *C. jejuni* since 29 May 2009.

**Question 1:** What do you know about *Campylobacter*? Is it known to be causing outbreaks? What information would you need to be sure whether *this* is an outbreak?

You check the notified *Campylobacter* cases in the statutory notification system and see that you have never in the last years seen such a high number of *Campylobacter* notifications; in fact, there have never been more than 10 notifications per month. No major changes in laboratory practices or the personnel working at the lab have been reported, nor has the mandatory notification system undergone any changes recently. There is some more information on *Campylobacter* in the Appendix (page [19](#_bookmark10)).

**Question 2**: How do you proceed to investigate the outbreak? What may have caused it?

You are also reminded that an outbreak of *Salmonella* Typhimurium had been identified in the same area five years earlier. You look at the old outbreak report and find out that tap water had been suspected as the vehicle of the outbreak back then. Given the fact that there is already a sharp decline in cases, you fear that Campylobacter may already be absent from the source(s) of the outbreak. Since environmental investigations are likely to show no proof of *Campylobacter* presence, you need epidemiological evidence for the source of the outbreak.

## Background: Descriptive epidemiology

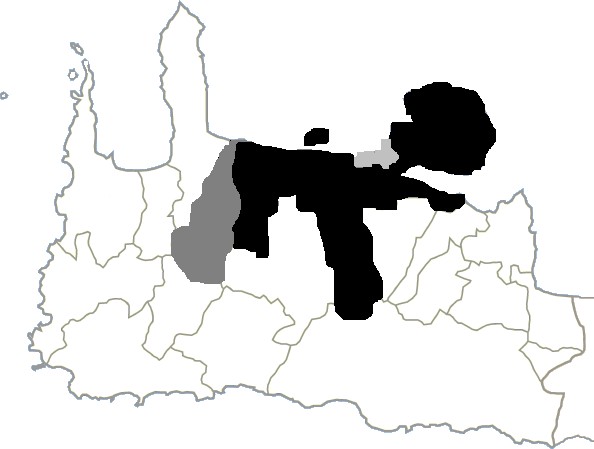
The outbreak appears to have stopped by the time you go to the outbreak area. You decide to first describe the cases by time, place and person. For this reason, you ask the hospital to inform you about all people with a laboratory diagnosis of *Campylobacter* since 1 May 2009. A few hours later, you receive information on all these cases. Based on the information, you first describe all cases by place, time and person.

NB The results are presented in the following table and figures as background information. If you want to practice calculating attack rates (ARs) from the raw data, please use the Excel file (**campy.xls**). There is no help provided for this task here, but you can have a look at the script of the ‘Copenhagen Case study’ where we calculated ARs as well. You will create the epicurve yourself in task 5.

**Table 1:** Attack rates of notified *Campylobacter* cases, May-June 2009, by municipality of residence.

|  |  |  |  |
| --- | --- | --- | --- |
| **Municipality** | **Population** | **Cases** | **AR (per 10,000)** |
| Town municipality | 55,838 | 8 | 1.44 |
| Rural municipality 1 | 10,321 | 19 | 18.4 |
| Rural municipality 2 | 10,586 | 11 | 10.4 |
| Rural municipality 3 | 7,463 | 6 | 8.04 |
| Rural municipality 4 | 7,301 | 6 | 8.22 |
| Rural municipality 5 | 6,313 | 6 | 9.50 |
| Rural municipality 6 | 2,932 | 12 | 40.9 |
| Rural municipality 7 | 3,296 | 6 | 18.2 |

town area



0 cases/10,000 population 0.01-9.99/10,000 population 10.0-19.9/10,000 population

≥20.0/10,000 population

**Figure 1**: Attack rates of *C. jejuni* by municipality, May-June 2009

**Table 2**: Attack rates of notified *Campylobacter* cases in the affected municipalities by age (filled in).

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **Population** | **Cases** | **AR (per 10,000)** |
| <12 months (0 years) | 1,050 | 22 | 209.5 |
| 1-4 years | 5,252 | 42 | 80.0 |
| 5-14 years | 9,721 | 10 | 10.3 |
|  |  |  |  |



**Figure 2**: Epicurve of notified *C. jejuni* cases, Town X, May-June 2009

**Question 3**: Based on these descriptive results by time, place and person, is there any more information you would like to have? How do the descriptive analysis results inform your hypothesis generation?

Two different water supply systems exist in the area; one for the town municipality and one for the seven adjacent rural municipalities. The 2004 *Salmonella* outbreak had affected the exact same rural municipalities, leaving the capital municipality unaffected.

You conduct a case-control study in order to identify the mode and vehicle of transmission: you define *cases* as laboratory-confirmed *Campylobacter* cases diagnosed at the local hospital in May-June 2009 and choose *controls* from the population registry of the affected municipalities, frequency matching for sex and age by one-year intervals (6-month intervals for cases aged below 1 year). In other words, you want to achieve the same age distribution among controls like among cases.

For the rest of this exercise, you will be presented and working with a set of the variables used in the outbreak investigation. All variables presented here have to do with water consumption. The file **campy.dta** contains information on all cases, as well as the selected controls.

## Analytical epidemiology

The rest of the case study will be dealing with the analytical epidemiology as part of the outbreak investigation. This consists of univariate and stratified analysis of the case-control study. From here on, help for the tasks (R codes) can be found in the file (**Campy\_R\_Guide\_2021**):

The dataset **campy.dta** contains twenty variables. The first few ones contain demographic information of participants (cases and controls) and the next ones provide information on different exposures directly or indirectly linked to tap water consumption.

For your convenience, here’s a list of the variables in the dataset:

* + id: ID of the questionnaire
  + case: case or control
  + datesym: date of symptom onset (if applicable or available)
  + age: age of the participant in years
  + sex: sex of the participant
  + supply: water supply system (1: rural area, 0: town area)
  + tap: do you drink tap water?
  + bottled: do you drink bottled water?
  + filter: do you use a filter for tap water at home?
  + well: do you have a well with water in your property?
  + pacifier1: do you use a pacifier?
  + pacifier2: is the child’s pacifier washed with tap water? (if applicable)
  + dishwasher: do you use the dishwasher for children’s utensils? (if applicable)
  + microwave1: do you use the microwave to prepare the child’s milk?
  + microwave2: do you use the microwave to prepare the child’s food?
  + breastfeeding: does the child breastfeed?
  + concentrated: does the child drink concentrated milk?
  + powder: does the child drink powder milk?
  + freshmilk: does the child drink fresh milk?
  + dilutetap: is the child’s milk diluted with unboiled tap water?

**Task 2**: How many observations does your dataset have? How many cases and how many controls does it contain?

**Task 3**: Explore each one of the variables. What information do they contain? Are they categorical or continuous variables?

**Task 4**: Can you think of any variables you could generate based on the ones you already have?

**Task 5**: Perform a descriptive analysis:

* What is the age distribution among cases and controls?
* Do cases and controls differ in terms of age and gender distribution?
  + Would you use a parametric or a non-parametric test to answer this question? Why?
* Draw an epidemic curve.

## 

## Univariate analysis

Now that you have already explored the data and performed some descriptive analysis, you may be wondering what really lies in the dataset.

**Task 6**: Conduct the univariate analysis; fill in Tables 3a-3c for these three example exposures with your results. Then, fill in the summary table 4 on the next page.

* Keeping in mind that we are talking about a dichotomous outcome (case) and dichotomous exposures, what statistical test could we use to see whether there is a statistical association?
* Which measure of association is appropriate for this study?
* Which commands can be used to calculate the measure of the association?
* How many cases and controls live in each water-supply zone?

**Table 3a**: Two-by-two table for sex

|  |  |  |
| --- | --- | --- |
| **Exposure** | **Cases** | **Controls** |
| **Sex: male** |  |  |
| **Sex: female** |  |  |

### OR: 95% CI: \_ p-value: \_

**Table 3b**: Two-by-two table for water supply system

|  |  |  |
| --- | --- | --- |
| **Exposure** | **Cases** | **Controls** |
| **Water supply system: rural** |  |  |
| **Water supply system: town** |  |  |

### OR: 95% CI: \_ p-value: \_

**Table 3c**: Two-by-two table for having a well in one’s property

|  |  |  |
| --- | --- | --- |
| **Exposure** | **Cases** | **Controls** |
| **Well in the property** |  |  |
| **No well in the property** |  |  |

### OR: 95% CI: \_ p-value: \_

Instead of constructing a two-by-two table for each one of the exposures in the case study, one can also produce a summary table with a slightly different structure (here: Table 4).

**Table 4**: Summary results of the univariate analysis.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exposure** | **Cases** | | **Controls** | | **OR** | **95% C.I.** |
| **N** | **%** | **N** | **%** |
| Rural water supply system (ref: town area) |  |  |  |  |  |  |
| Consumption of tap water |  |  |  |  |  |  |
| Consumption of bottled water |  |  |  |  |  |  |
| Using a filter for tap water |  |  |  |  |  |  |
| Having a well in one’s property |  |  |  |  |  |  |
| Using a pacifier |  |  |  |  |  |  |
| Washing the pacifier with tap water |  |  |  |  |  |  |
| Use of a dishwasher for children’s utensils |  |  |  |  |  |  |
| Use of microwave to prepare milk |  |  |  |  |  |  |
| Breastfeeding |  |  |  |  |  |  |
| Concentrated milk consumption |  |  |  |  |  |  |
| Powder milk consumption |  |  |  |  |  |  |
| Fresh milk consumption |  |  |  |  |  |  |
| Diluting milk with unboiled tap water |  |  |  |  |  |  |
| Any milk needing to be diluted with water |  |  |  |  |  |  |

**Question 4:** Which exposures seem to be important to you? Why?

## Part II: Objectives

At the end of the session, participants should be able to:

* perform stratified analysis in an outbreak investigation
* identify and interpret effect modification and confounding
* compute and interpret stratum-specific and Mantel-Haenszel odds ratios
* summarise the results of an outbreak investigation from all performed analyses

## Stratified analysis

We have seen so far that some exposures appear to be statistically significantly associated with being a case. Some other exposures do not appear to be associated with disease outcome.

Being a field epidemiologist, you decide not to stop your analysis yet. You think your results are very interesting and, after a discussion with your colleagues, you go further and perform a stratified analysis.

**Question 5**: Discuss what stratified analysis can offer. Which exposures would you stratify by and why?

**Task 7**: Stratify by water supply zone (the variable supply, where 1 stands for the rural area and 0 for the town area), calculate the odds ratios again and fill in Table 5 (next page). Identify effect modification or confounding and put an **X** in the final two columns of the table to show which phenomenon is present.

**Question 6**: After all the analyses you have carried out, can you argue that tap water consumption was the cause of this outbreak?

**Question 7**: How would you summarise your results? What could you recommend?

**Table 5**: Univariate and stratified analysis for the *C. jejuni* outbreak in Greece, May-June 2009.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Exposure** | **OR (95% CI), crude** | **OR (95% CI), rural** | **OR (95% CI), town** | **Effect modification** | **Confounding** |
| Consumption of tap water | 2.84 (1.16-7.94) | 4.30 (1.61-13.3) | 0.71 (0.06-38.5) |  |  |
| Consumption of bottled water | 0.23 (0.10-0.48) | 0.15 (0.06-0.37) | 1.02 (0.14-5.91) |  |  |
| Using a filter for tap water | 0.00 (0.00-0.23) | 0.00 (0.00-0.26) | 0.00 (0.00-2.30) |  |  |
| Having a well in one’s property | 3.82 (1.65-9.87) | 1.28 (0.18-14.5) | 1.92 (0.16-13.4) |  |  |
| Using a pacifier | 0.88 (0.47-1.56) | 0.72 (0.37-1.42) | 2.79 (0.44-30.2) |  |  |
| Washing the pacifier with tap water | 3.00 (1.18-8.09) | 3.23 (1.05-11.0) | 1.55 (0.17-13.6) |  |  |
| Use of a dishwasher for children’s utensils | 0.21 (0.09-0.48) | 0.16 (0.05-0.47) | 0.15 (0.02-1.03) |  |  |
| Use of microwave to prepare milk | 1.79 (0.58-6.22) | 5.67 (1.09-55.2) | 0.21 (0.02-1.98) |  |  |
| Use of microwave to prepare food | 0.75 (0.17-3.84) | 0.89 (0.20-4.64) | N/C |  |  |
| Breastfeeding | 0.52 (0.12-1.71) | 0.52 (0.12-1.71) | 0 (0-14.1) |  |  |
| Concentrated milk consumption | 1.91 (1.04-3.50) | 2.68 (1.34-5.37) | 0.80 (0.13-4.79) |  |  |
| Powder milk | 1.02 (0.50-2.02) | 0.95 (0.43-2.06) | 1.17 (0.10-7.69) |  |  |
| Fresh milk | 0.53 (0.25-1.09) | 0.53 (0.23-1.21) | 0.67 (0.06-4.27) |  |  |
| Diluting milk with unboiled tap water | 1.46 (0.71-3.01) | 3.44 (1.34-9.17) | 0.43 (0.04-3.46) |  |  |
| Any milk that needs to be diluted with water | 2.50 (1.19-5.54) | 3.48 (1.52-8.52) | 0.86 (0.13-9.77) |  |  |

N/C: not calculable

### Help Task 7

**Table 6**: Univariate and stratified analysis for the *C. jejuni* outbreak in Greece, May-June 2009 (filled in).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Exposure** | **OR (95% CI), crude** | **OR (95% CI), rural** | **OR (95% CI), town** | **Effect modification** | **Confounding** |
| Consumption of tap water | 2.84 (1.16-7.94) | 4.30 (1.61-13.3) | 0.71 (0.06-38.5) | YES | N/A |
| Consumption of bottled water | 0.23 (0.10-0.48) | 0.15 (0.06-0.37) | 1.02 (0.14-5.91) | YES | N/A |
| Using a filter for tap water | 0.00 (0.00-0.23) | 0.00 (0.00-0.26) | 0.00 (0.00-2.30) | YES | N/A |
| Having a well in one’s property | 3.82 (1.65-9.87) | 1.28 (0.18-14.5) | 1.92 (0.16-13.4) | NO | YES |
| Using a pacifier | 0.88 (0.47-1.56) | 0.72 (0.37-1.42) | 2.79 (0.44-30.2) | NO | NO |
| Washing the pacifier with tap water | 3.00 (1.18-8.09) | 3.23 (1.05-11.0) | 1.55 (0.17-13.6) | YES | N/A |
| Use of a dishwasher for children’s utensils | 0.21 (0.09-0.48) | 0.16 (0.05-0.47) | 0.15 (0.02-1.03) | YES | N/A |
| Use of microwave to prepare milk | 1.79 (0.58-6.22) | 5.67 (1.09-55.2) | 0.21 (0.02-1.98) | YES | N/A |
| Use of microwave to prepare food | 0.75 (0.17-3.84) | 0.89 (0.20-4.64) | N/C | N/C | N/C |
| Breastfeeding | 0.52 (0.12-1.71) | 0.52 (0.12-1.71) | 0 (0-14.1) | NO | YES |
| Concentrated milk consumption | 1.91 (1.04-3.50) | 2.68 (1.34-5.37) | 0.80 (0.13-4.79) | YES | NO |
| Powder milk | 1.02 (0.50-2.02) | 0.95 (0.43-2.06) | 1.17 (0.10-7.69) | NO | NO |
| Fresh milk | 0.53 (0.25-1.09) | 0.53 (0.23-1.21) | 0.67 (0.06-4.27) | NO | NO |
| Diluting milk with unboiled tap water | 1.46 (0.71-3.01) | 3.44 (1.34-9.17) | 0.43 (0.04-3.46) | YES | N/A |
| Any milk needing to be diluted with water | 2.50 (1.19-5.54) | 3.48 (1.52-8.52) | 0.86 (0.13-9.77) | YES | N/A |

N/A: not applicable N/C: not calculable

# Campylobacter

## Appendix

*Campylobacter* is a genus of bacteria that are a major cause of gastroenteritis throughout the world. Infection occurs mainly following consumption of contaminated undercooked poultry or contaminated water.

The most common symptoms of *Campylobacter* infection include diarrhoea, abdominal pain, fever, headache, nausea and vomiting. Symptoms usually start 2–5 days after infection, and last for 3–6 days.

Specific treatment is not usually necessary, except to replace electrolytes and water lost through diarrhoea, but antimicrobials may be needed to treat invasive infection and the carriers. Severe complications, such as Guillain-Barré syndrome, may follow *Campylobacter* infection.

More information available on: <http://www.who.int/mediacentre/factsheets/fs255/en/>