My outbreak report

Helene Langet

2023-12-31

## Overview

The outbreak ran from **date** to **date**. This report covers the period **date** to **date**. Over the studied period, there were **N** cases, including **N** confirmed cases and **N** confirmed deaths.

## Population characteristics

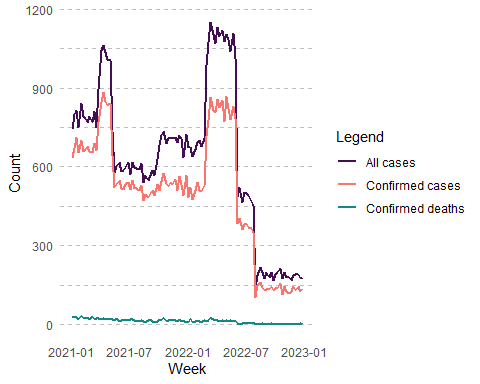
**cross-reference** provides a summary of the demographic characteristics and the outcome proportion for the overall population, while **cross-reference** compares the characteristics of individuals who died versus those who are still alive.

| **Characteristic** | **N = 65,669**1 |
| --- | --- |
| age | 50 (35, 65) |
| sex |  |
| 1 | 33,114 (50%) |
| 2 | 32,555 (50%) |
| bmi | 29 (21, 38) |
| confirmed |  |
| 0 | 13,235 (20%) |
| 1 | 52,434 (80%) |
| death |  |
| 0 | 64,455 (98%) |
| 1 | 1,214 (1.8%) |
| 1Median (IQR); n (%) | |

| **Characteristic** | **Overall**, N = 65,6691 | **0**, N = 64,4551 | **1**, N = 1,2141 |
| --- | --- | --- | --- |
| sex |  |  |  |
| 1 | 33,114 (50%) | 32,504 (50%) | 610 (50%) |
| 2 | 32,555 (50%) | 31,951 (50%) | 604 (50%) |
| age | 50 (35, 65) | 50 (35, 65) | 52 (37, 67) |
| bmi | 29 (21, 38) | 29 (21, 38) | 34 (28, 41) |
| 1n (%); Median (IQR) | | | |

## Outbreak evolution

**cross-reference figure** illustrates the outbreak’s progression, which can be divided into distinct phases.



Weekly count of all cases, confirmed cases and deaths

## Logistic regression model

In this section, we build and evaluate a logistic regression model to understand the relationship between certain variables and the likelihood of death among confirmed cases.

The logistic regression model uses death as the response variable, and bmi and age as predictor variables. It is implemented in R as shown in the code chunk referenced by **cross-reference**.

#Logistic regression model  
model <- glm(death ~ bmi + age,  
 subdf |> dplyr::filter(confirmed == "1"),  
 family = binomial)

The results of the logistic regression model are summarized in the formatted regression table, which is presented in **cross-reference**.

| **Characteristic** | **OR**1 | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| bmi | 1.04 | 1.03, 1.04 | <0.001 |
| age | 1.00 | 1.00, 1.01 | 0.003 |
| 1OR = Odds Ratio, CI = Confidence Interval | | | |

## Appendix

The dataset includes the following variables:

* age = age of the individual (continuous variable) ;
* sex = sex of the individual (binary variable: 1 = male, 2 = female) ;
* bmi = Body Mass Index (BMI) of the individual (continuous variable) ;
* confirmed = flag of confirmed cases (binary variable: 0 = no, 1 = yes)
* death = death outcome (binary variable: 0 = alive, 1 = died)

Data summary

|  |  |
| --- | --- |
| Name | subdf |
| Number of rows | 65669 |
| Number of columns | 8 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| Date | 2 |
| factor | 3 |
| numeric | 3 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: Date**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| date | 0 | 1 | 2021-01-03 | 2022-12-31 | 2021-11-20 | 725 |
| week | 0 | 1 | 2021-01-03 | 2022-12-25 | 2021-11-14 | 104 |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| sex | 0 | 1 | FALSE | 2 | 1: 33114, 2: 32555 |
| confirmed | 0 | 1 | FALSE | 2 | 1: 52434, 0: 13235 |
| death | 0 | 1 | FALSE | 2 | 0: 64455, 1: 1214 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| id | 0 | 1 | 5768.89 | 4029.27 | 1 | 2352 | 4992.00 | 8896.00 | 14993.00 | ▇▆▅▃▂ |
| age | 0 | 1 | 49.96 | 17.58 | 20 | 35 | 50.00 | 65.00 | 80.00 | ▇▇▇▇▇ |
| bmi | 0 | 1 | 30.02 | 11.11 | 10 | 21 | 29.12 | 37.58 | 62.22 | ▆▇▇▃▁ |