# Covid statistics: an overview of the 2020 worldwide reports

written by
Andreea ANDRONACHE

Last updated: 27 January 2023

## Introduction

World Health Organization issued several cases of an unknown form of virus on 31 December 2019. From 3 February 2020 until the end of the same month, the world faced a significant raise in the number of patients suffering from a new form of pneumonia.

This report aims to describe the statistical data on the Worldwide Pandemic situation in 2020, analyzing the results of the death rates. Kaggle was the primary source of information.

In the first part of the report, the overall dataset will be analysed considering the worldwide data and then the results over Europe.

## Reported cases

February 2020 signalled a global increase in the number of cases of the COVID-19 virus.

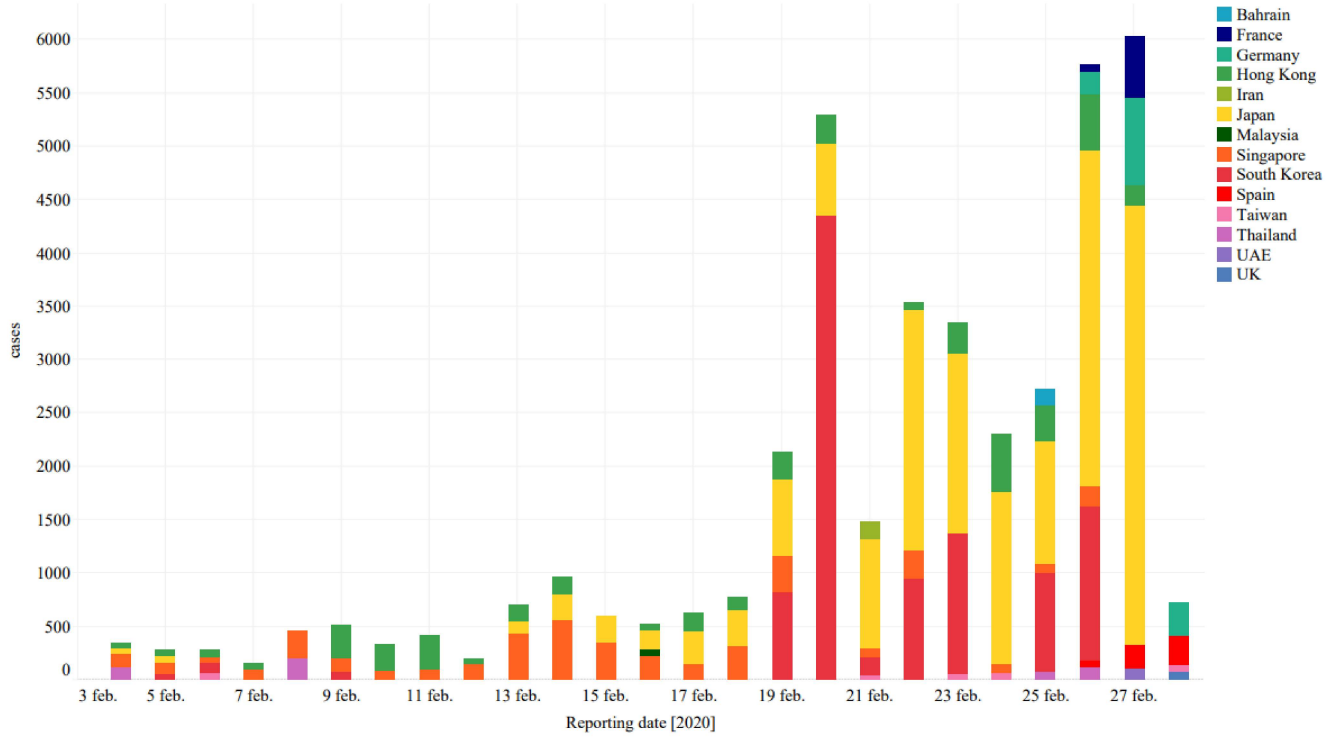


Figure 1: Reporting cases in February 2020.

As can be noticed in the graph of Fig. 1, there appeared a significant increase in the number of cases in *Japan* and *South Korea*. It may be said that Asia presented the highest risk factor regarding the accelerated number of cases per day.

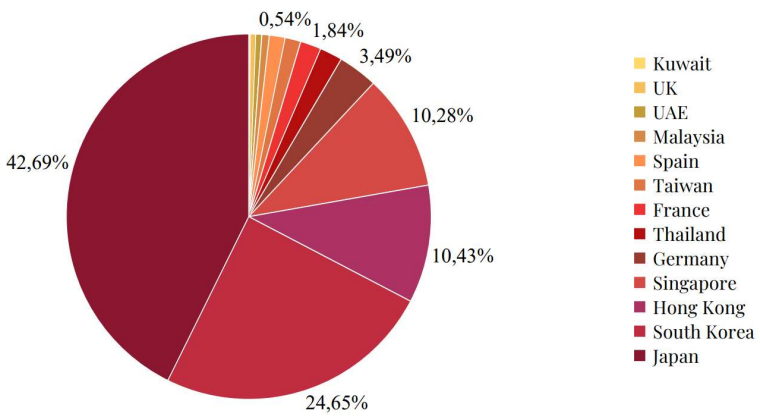


Figure 2: Reporting cases per country.

People claimed the appearance of the symptoms so that they were declared positive with Covid-19 virus.[1] Following the presented information, there are some questions regarding the data:

1. What is the *average age* of the patients?
2. What is the predominant *gender* of the patients who passed out?
3. What are the *symptoms*?

These questions will be approached further.

## The mean age of infected people

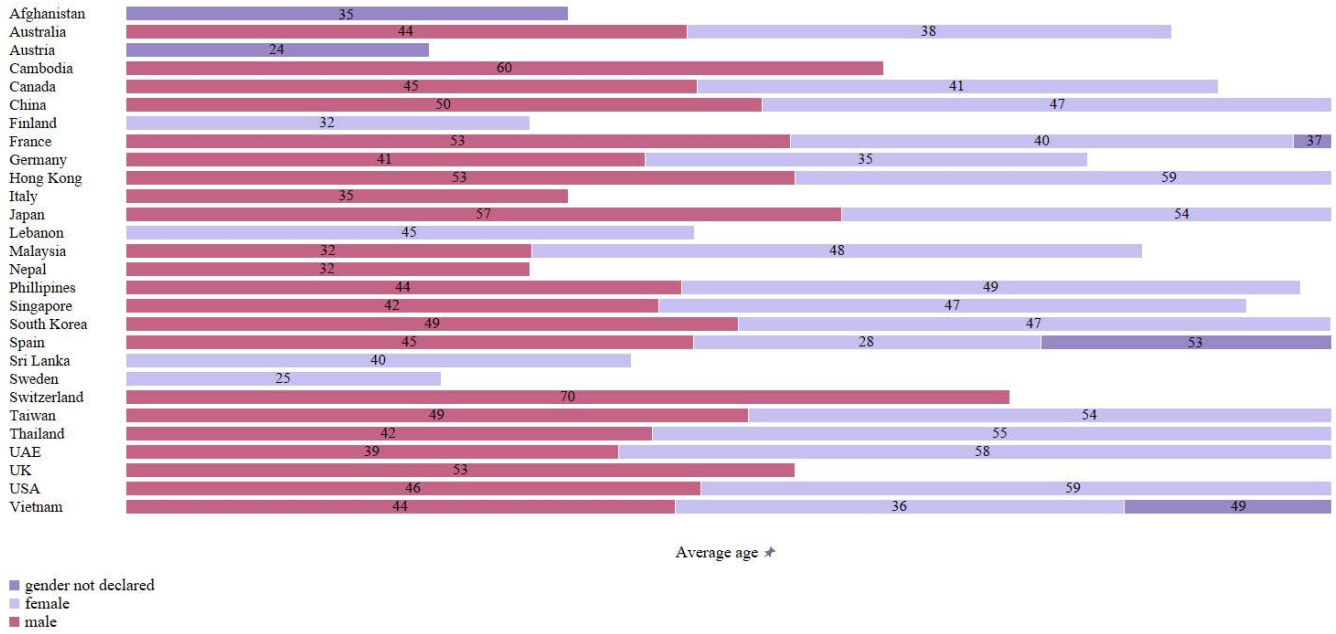


Figure 3: The average age of infected patients per country.

As shown in Fig. 3, the average age of infected patients varies from country to country.

The first conclusion of these results is given by the mean age of infected persons, that is *49*. But where are registered the limit values of ages? The answer is provided in the table below:

| Metric | Female | Male | Not Available |
|---|---|---|---|
| Minimum age | 25 | 32 | *24* |
| Countries | Sweden | Malaysia, Nepal | Austria |
| Maximum age | 59 | *70* | 53 |
| Countries | Hong Kong, USA | Switzerland | Spain |

Table 1: Minimum, maximum declared age per country.

[1] The dataset displays the worldwide pandemic-reported cases. Thus, some data is missing/ not declared. The overall results are indicative.

## The prevailing gender of patients

It is important to take into account the affected gender. After consulting the dataset, it was stated that the majority of death cases were represented by men.

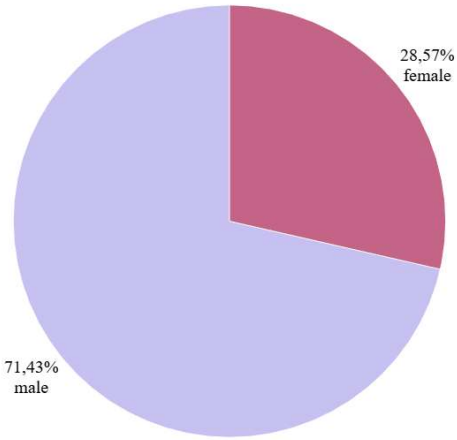


Figure 4: Death rate.

Looking at the pie chart, there shows a tendency to rely on the result and build the second conclusion of the report: *men* are exposed to a higher chance of dying due to the virus. However, it should be adequate to use statistics to prove the previous hypothesis.

Thus:

- Female gender has more chances to survive?

For further computations, let's assume that gender has no effect.

Two parameters are worth considering: the *confidence interval* and the *p-value* [2].
The first step is to see if the null hypothesis can be rejected or not.

Consider 0.99 as the confidence interval, meaning that one can be 99% sure that the true parameter value lies within the sample.

Let's see the mean values[3]:

- for women: 3.7%
- for men: 8.5%

The 99 % confidence explains that men have from 0.8% to 8.8% chances of passing away. The p-value is 0.002105.
Let's take a closer look at the definition of this parameter:

> *The probability value is a number describing how likely it is that your data would have occurred by random chance (i.e. that the null hypothesis is true).*[4]

If the p-value < 0.05, it means that the null hypothesis is *rejected*, and the result is statistically significant. Therefore, the death rate of the male gender was more expected than of the opposite gender.

[2] Probability value
[3] All results were computed with the aid of the R programming language.
[4] What a p-value tells you about statistical significance

Even if the men were exposed to a higher risk of death according to global data, Europe found the opposite:

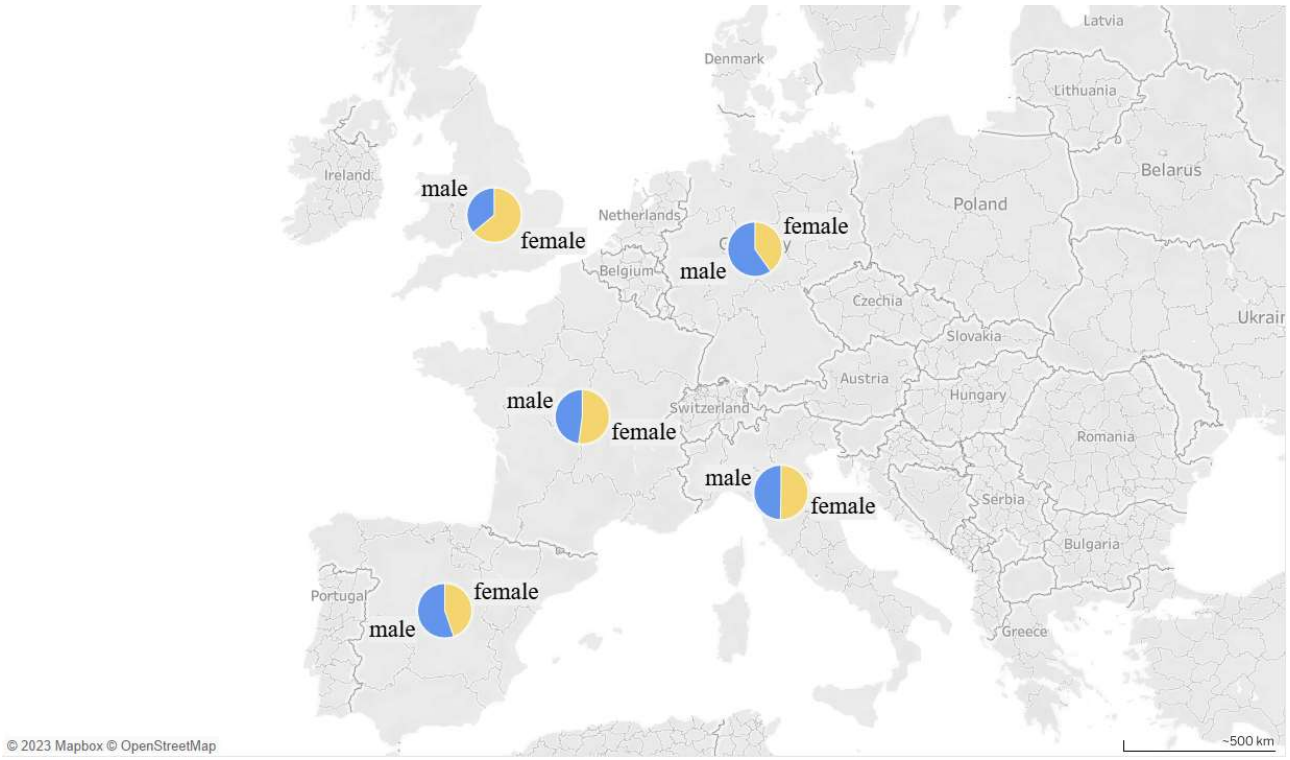


Figure 5: Reported death rate in Europe considering the gender of the patients.

*Note: Every pie chart depicts the percentage of women\man who passed away. The overall results are indicative.

Going back to the age-related topic, another question shows up:

- People who died were *older* than the others who survived?

Firstly, let's check how many patients survived. Look at the diagram below:

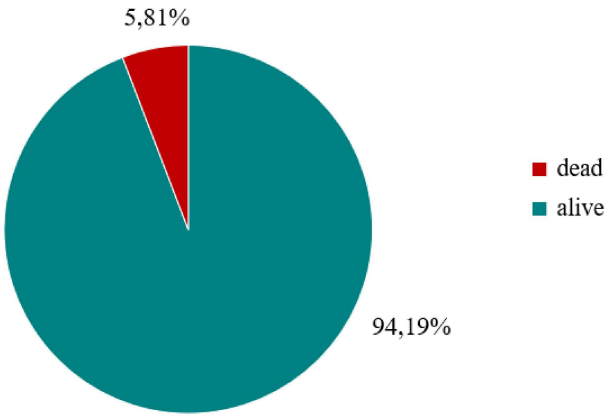


Figure 6: Reported cases.

Secondly, if computing the mean age of deceased, respectively alive persons, the following values will be obtained:

| Average age | |
|---|---|
| Alive | Dead |
| ~48 | ~64 |

*Note: Tilde symbol(~) represents the rounded age. The obtained values using Rstudio are 48.07229, respectively 68.58621.

The null hypothesis is the contradiction of the main topic of this section, thus, let's suppose that the persons who combat the virus were younger.

The confidence interval is 0.99, so the age values have 15.5% to 25.5% chances corresponding to the older population.

P-value < 2.2e-16, almost equal to 0. Again, the p-value is statistically significant. The null hypothesis is thus *rejected* and another key aspect is deduced: *the older reported persons died instead of the younger ones*.

## The symptoms of the patients

The last aspect is given by the health status of the reported population. The majority of patients manifested only one indication of illness.

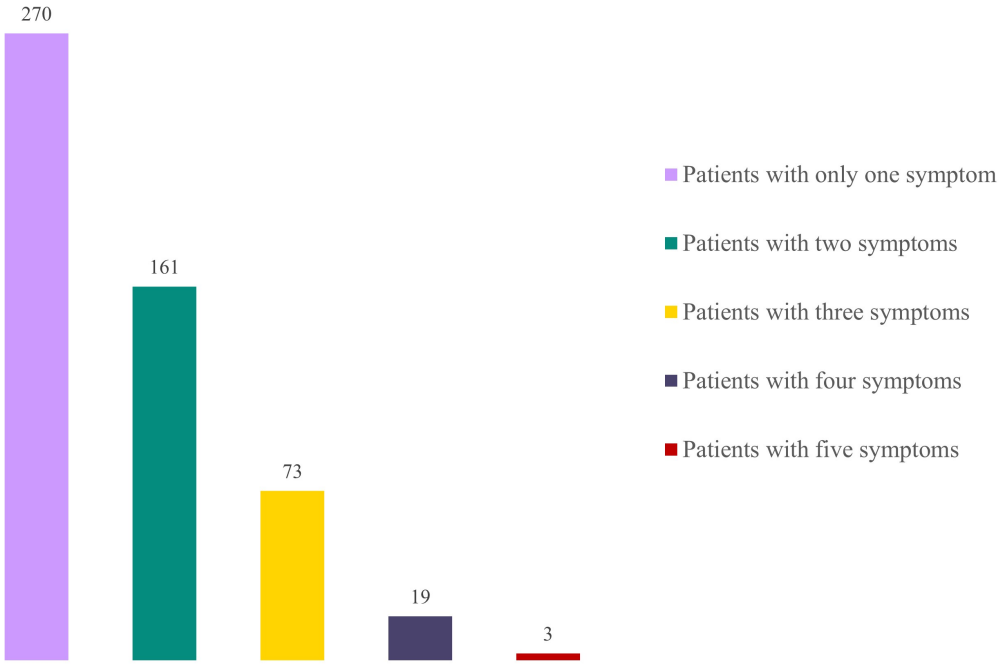


Figure 7: Number of infected persons with multiple symptoms.

Still, what is the most common virus evidence? Checking the diagram below, *fever* and *cough* are the "top players".

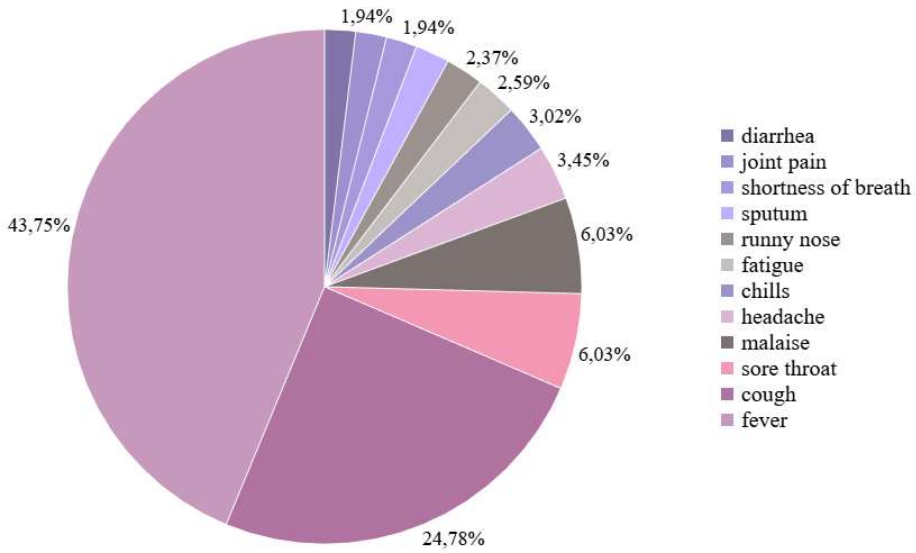


Figure 8: Most common symptoms encountered by the patients.

But did they make victims among the reported cases?

After some calculations, the fever affected only 1.97% of the sample and the cough 3.41%. Drawing a conclusion is pretty inadequate in this case by few given data.


Figure 9: Least common symptoms encountered by the patients.

The picture above comes to show the least common symptoms of the patients, a fact that cannot be omitted.

## Concluding remarks

The report focused on the dataset deemed by the World Health Organization considering Covid-19 the main factor in the disturbance of the worldwide general state.

Thus, the dataset was analyzed, and the resolution is similar to the experience during the pandemic:

- The older population was the principal target of the infection.
- Most of the cases showed up in Asia and the general symptoms were associated with fever, respectively cough.
- Men were affected in a substantial number than women were.

However, the study highlighted the cases from *February 2020*, so all the results give an overall perspective for the beginning of the pandemic.

Tools that gave me aid to make this report + their usage:

| | |
|---|---|
| Excel | Sorting, cleaning, vizualization, computations |
| LaTeX | Writing the report |
| PosgreSQL | Computations, final checks, sorting |
| Rstudio | Statistics, sorting, computations |
| Tableau | Vizualization |