Lower Bound for the Number of Asymptomatics in Infected by COVID-19

V. Nardozza*

03/04/2020†

Abstract

We propose a method for evaluating the number of asymptomatics in a COVID-19 Outbreak. The method will give only a lower bound for the real number.

Key Words: quantitative virology.

1 Introduction

With the COVID-19 crisis ongoing, a popular topic is the possibility that many of the people infected may be asymptomatic. We wander if it is possible to evaluate the number of asymptomatic from available data and we propose a method.

The task is performed more for fan then with a real scientific intent and we wonder if our method gives some real results. We leave to anyone with a more solid scientific background of the one we have to give an answer.

The evaluation is based on official data from the Italian COVID-19 Outbreak of the beginning of 2020.

2 The Method

The infection of COVID-19 have proven to have a mortality that increases with age and pre-existing health conditions and maybe the latter is actually a consequence of the former since probability of health conditions increases with age. Moreover it affects more men than women, once again probably because the health conditions that increase the risk are more common in men.

Base on the above consideration it makes sense to divide the population based on age and gender. Suppose we divide the population in k groups composed by $n_1, \dots n_k$ individuals and such that $\sum_{i=1}^{k} = N$ where N is the total number of people.

Let s_i be the people that have been tested positive to the virus and that at the same time have symptoms. We do not have this data although this data are available to the authorities end research institutions which may therefore

*Electronic Engineer (MSc). Lancashire, UK. mailto: vinardo@nardozza.eu

†Posted at: https://vixra.org - Biology - Current version: v1 - April 2020

perform a more accurate calculation. We will use instead for s_i simply the number or reported infected. On one side many people have been tested without symptoms and on the other side many with symptoms have not been tested and therefore we hope this two effects somehow compensate.

Let r_i be the real but unknown number of people that have been infected. We suppose that in any group the people showing symptoms are fraction η_i of the real infected:

$$\begin{array}{llll} s_i = \eta_i r_i & \text{by setting} & \gamma_i = \frac{1}{\eta_i} & \text{we have} \\ \gamma_i s_i = r_i & \Rightarrow & \gamma_i s_i - r_i = 0 & \end{array} \tag{1}$$

Moreover, let ϵ_i be the fraction of each group that has been infected. We set $\epsilon_i = \alpha_i \epsilon$ where ϵ is fixed and α_i depends from the group. we have:

$$r_i = \epsilon\, \alpha_i n_i \tag{2}$$

Finally we have the equation:

$$\sum_{i=1}^{k} \epsilon_i\, n_i = R \quad \Rightarrow \quad \sum_{i=1}^{k} \epsilon\, \alpha_i \eta_i n_i = S \tag{3}$$

where R is the total number of infected and S is the total number or official infected (i.e. Symptomatics given the reasoning above).

We make now the critical assumption that every individuals in the population has the same probability to get infected. The general impression is that young people do not get the virus. However, this is due to the fact that they do not show any symptom. There is no biological reason to believe that they are immune. We do recognise that some categories go to work and have more social contact then other and this can be taken into account with the α_i. For example adults in working age, even in a lockdown situation (i.e. social distancing measures in place), take the virus home to younger and orderly people but they are usually more exposed. In a lockdown status we suggest $\alpha_i = 1$ for young and elderly people and $\alpha_i = 1.1$ for all others.

If we put eq. [1] and [2] together we get:

$$\left\{ \begin{array}{rcl} \gamma_i s_i - r_i & = & 0 \\ r_i - \epsilon\, \alpha_i n_i & = & 0 \\ \gamma_i & > & 1 \end{array} \right. \tag{4}$$

where we have not used Eq. [3] because it is not independent from the other two. In Eq. [4] we have $2k$ linear equations in the $2k + 1$ unknowns r_i, γ_i and ϵ. To get a solution we have to make an assumption on one of the γ_i to get an additional equation. By solving the above equations we get the answer to our original question.

3 A Real Example

In this section we will apply the method to a real example which is the Italian COVID-19 Outbreak of the beginning of 2020.

The following table reports the available data for the above outbreak at 10/03/20. Although we are writing more then two weeks after the validity of above data set, (i.e. 03/04/20) the data reported below are the only available to us. However, although it should be better to use more updated data (at the time of writing total cases have reached as much as 10 times the cases the available data refers to), for the scope of this article they are good enough. Moreover, since there is a discrepancy between cases in men and women, group should be also divided by gender. Once again we have not the data to do it.

The table contains information of the composition of the groups, data on the number of infected per group (r_i, total cases including deaths), the α_i chosen for each group and the composition of the population at that date (n_i) expressed as a percentage of the whole population.

Group	Age	Symptomatics (s_i)	α_i	$Pop.(n_i[\%])$
1	0-9	34	1	8.434
2	10-19	85	1	9.558
3	20-29	297	1	10.274
4	30-39	473	1.1	11.720
5	40-49	897	1.1	15.313
6	50-59	1470	1.1	15.494
7	60-69	1530	1	17.156
8	70-79	1945	1	9,878
9	> 80	1968	1	7.174
Tot.	N/A	8699	N/A	100

Table 1 : Data for Italian outbreak at 10/03/2020

It is not necessary to use the real values for the n_i since, as long as they are proportional to the real one, Eq. [4] still works and the unknown ϵ will absorb the proportional constant. This is why we use percentages instead.

To solve Eq. [4] we need an additional assumption since, as said before, there are more unknowns then equations. We will assume that $\gamma_9 = 1.2$. With the values in Tab. 1 and the above assumption we find:

Age	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	> 80
γ_i	81.66	37.02	11.39	8.97	6.18	3.82	2.62	1.67	1.20

Table 1 : Values of η_i and $\gamma_i = \frac{1}{\eta_i}$

We can eventually evaluate the ratio A_r between Symptomatics and total infected people:

$$A_r = \frac{\sum_{i=1}^{k} s_i}{\sum_{i=1}^{k} r_i} = \frac{\sum_{i=1}^{k} s_i}{\sum_{i=1}^{k} \gamma_i s_i} = 0.253 \tag{5}$$

The above estimation is a lower bound because it is based on the assumption that the total number of the infected, aged over 80, is only 1.2 times the number of the ones that are symptomatic.

However, if our assumptions are correct, for each person showing symptoms in the entire population there are at least 3 that are asymptomatic.