RESEARCH ARTICLE | JUNE 07 2024

An innovative approach to estimate infection by COVID-19

Manuela Oliveira ✉; Eugénio Garção; Luís M. Grilo; João T. Mexia

AIP Conf. Proc. 3094, 450005 (2024)
https://doi.org/10.1063/5.0212209

An Innovative Approach to Estimate Infection by COVID-19

Manuela Oliveira[1,2,6, a)] Eugénio Garção[3, b)] Luís M. Grilo[1,4,5,6, c)] João T. Mexia[6, d)]

[1]*Department of Mathematics, University of Évora,*
Colégio Luis António Verney, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal

[2]*CIMA - Center for Research on Mathematics and its Applications, University of Évora, Evora, Portugal*
[3]*Department of Mechatronics Engineering, University of Évora, Evora, Portugal*
[4]*Ci2 - Smart Cities Research Center, Tomar, Portugal*
[5]*CIICESI / ESTG – P. Porto, Porto, Portugal*

[6]*NOVA Math (Center of Mathematics and its Applications), NOVA FCT, NOVA University of Lisbon, Lisbon, Portugal*
a)Corresponding author: mmo@uevora.pt
b)jesg@uevora.pt
c)luis.grilo@uevora.pt
d)jtm@fct.unl.pt

Abstract. Given that individuals in a certain population are different (among other things they have a different immune system), it is possible that some are infected with the known virus COVID-19 and are asymptomatic and therefore not diagnosed with the disease. Thus, estimates of the number of infected and dead with COVID-19 may not correspond to reality. This study seeks to indicate a procedure to estimate the number of individuals in the infected population that are asymptomatic (not diagnosed, but possible transmitters of the disease), based on the number of infected individuals (already diagnosed). We showed how with data available (numbers of symptomatic, symptomatic in the hospital and deceased) on the evolution of the pandemic in five regions of mainland Portugal it is possible to estimate the number of asymptomatic and immune individuals in the population.

INTRODUCTION

Human coronaviruses were discovered in the 1960s (30 decades after its appearance in animals), having taken different mutations since then. According to its type, it causes respiratory, hepatic, gastrointestinal, and neurological symptoms. In the end of 2019, cases of pneumonia have appeared in Wuhan, a city in China's Hubei Province, rapidly spreading across China and to other countries of the world. It was named "severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)" by the International Committee on Taxonomy of Viruses and we commonly know it as COVID-19. Epidemiology focuses on the identification of patterns in disease occurrence in order to provide information that may be useful to help prevent it. The research of how a disease may be transmitted is influential for that identification. In the case of COVID-19 this research should focus on the relations between the disease and the populations of susceptible individuals that might be infected [1].

THE SAMPLE FRAMEWORK

The use of auxiliary information in estimating population mean or total is well known in the field of survey sampling. Various survey sampling schemes such as stratified sampling, cluster sampling is frequently used. Stratified

International Conference of Numerical Analysis and Applied Mathematics
AIP Conf. Proc. 3094, 450005-1–450005-4; https://doi.org/10.1063/5.0212209

sampling has various advantages, in particular in cases where the estimates of the population characteristics are needed not only for the entire population but also for its different subpopulations (e.g. [2, 4]).

In this paper we propose to define a protocol for the estimation of the people infected by COVID-19. Starting from a population where it has been ascertained that individuals are infected (the verified cases), the aim is to estimate the population that is infected but has not yet been diagnosed (the asymptomatic cases). Estimates of the case fatality ratio, and other epidemiological parameters, will likely be lower than current crude mortality estimates once the full spectrum of disease is able to be included in the denominator. In addition, the role of asymptomatic or subclinical infections in human-to-human transmission of COVID-19 virus is not well understood and it is not yet clear whether those who are reported as asymptomatic may be able to transmit the virus to other individuals.

Addressing COVID-19 risk encompasses thus:

1. Locating the individuals infected to help stop the disease transmission;
2. Estimating the number of individuals that are asymptomatic or that cannot be infected in the future (immune) to decide when lightening confinement (deconfinement policy).

If we assume for COVID-19 the evolution diagram (Figure 1).

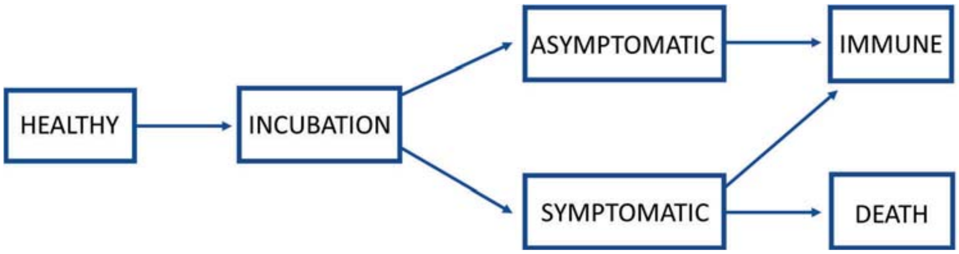

FIGURE 1. Schematic diagram with the COVID-19 evolution.

We easily see that the information collected in connection with testing special classes. Our propose will be to present a stratified sampling scheme to provide the information required for the second aspect.

THE SAMPLING DESIGN

As a first level stratification level we propose to avail ourselves of the usual deviation. As a first stratification level we consider the usual division, for statistical proposes, of Mainland Portugal in to five regions (NUTS-I): Norte, Centro, Lisboa and Vale do Tejo, Alentejo and Algarve. Next, for carrying out the second stratification level we consider for each of the regions a random sample of 15 Family Doctors (FD) residing in that region. For each of the selected FD a random sub-sample with dimension 20 will be chosen from the set of people he/she assists. The percentages that we are interested are p_1: healthy; p_2: asymptomatic; p_3: symptomatic; p_4: symptomatic + (in hospital); p_5: immune and p_6: deceased. Since we are interested in having a global view, we assume the 5 regions to be homogeneous. Thus, we can aggregate the sub-samples corresponding to FD in each region obtaining 5 independent samples with dimensions approximately 300.

Given the number of cases $n_{i1},\dots,n_{i6}$, in the *i*-th sample, $i = 1,\dots,5$ we have the unbiased estimators, $\tilde{p}_{ij} = \frac{n_{i,j}}{n_i}$, $i = 1,\dots,5$; $j = 1,\dots,6$, where $n_i = \sum_{j=1}^{6} n_{ij}$, $i = 1,\dots,5$.

Now with,

$$\begin{cases} \tilde{\boldsymbol{p}}_{\boldsymbol{i}} = (\tilde{p}_{i1},\dots,\tilde{p}_{i6}), & i = 1,\dots,5 \\ \boldsymbol{p}_{\boldsymbol{i}} = (p_{i1},\dots,p_{i6}), & i = 1,\dots,5 \end{cases}$$

it is know that, see for instance [3],

$$\sqrt{n_i}\,(\tilde{\boldsymbol{p}}_{\boldsymbol{i}} - \boldsymbol{p}_{\boldsymbol{i}}) \xrightarrow[n_i \to \infty]{\mathcal{D}} \mathrm{N}(\boldsymbol{0}, \mathbb{U}(\boldsymbol{p}_{\boldsymbol{i}})), i = 1,\dots,5$$

this is we have a limit normal distribution with null mean vector and covariance matrix

$$\mathbb{U}(\boldsymbol{p}_{\boldsymbol{i}}) = D(\boldsymbol{p}_{\boldsymbol{i}}) - \boldsymbol{p}_{\boldsymbol{i}}\,\boldsymbol{p}_{\boldsymbol{i}}^{t}, i = 1,\dots,5$$

where D($\boldsymbol{p}_{\boldsymbol{i}}$) is the diagonal matrix with principal elements $p_{i1},\dots,p_{i6}$, $i = 1,\dots,5$. This can be used to obtain joint confidence intervals for the probabilities. Moreover, having $n_i \approx 300$, $i = 1,\dots,5$ we will be, at least approximately, near the balanced case, and we can use ANOVA like techniques to compare the regions. From the successive estimators we can infer about the evolution of the pandemic in each region. An important point of interest is that we are interested in obtaining global estimates for the 5 regions. Since, given evens A and B

$$pr(A) - pr(A|B) \leq 1 - pr(B)$$

so, when $pr(B) \approx 1$, estimating properly $pr(A|B)$, gives us an acceptable estimator of $pr(A)$. This is the reason why we centered our scheme on the FD and the people they assist which is the bulk of the population, discarding, for instance, retirement houses and prisons. We want to stress that our goal is to obtain information on the evolution of the pandemic in the 5 regions to enable proper planning. We point out that the numbers of symptomatic, symptomatic + (in hospital) and deceased are available. So, we must use tests to find the numbers of the asymptomatic and immune.

If we look at the initial graph (Figure 1) we see:

- that "death" is an "absorbing state" only entries into it can be considered
- that "immune" is who also an "absorbing state"
- that "asymptomatic" only gives access to "immune"

thus $p_2 + p_5$ can only increase (at least not decrease) with time. Putting $q = p_2 + p_5$, if q estimated periodically, we can detect for a given level q^* the first time for which we have

$$pr(q \geq q^*) \geq \alpha$$

where α will be a safety level and, possibly $q^* = 0.6$. This knowledge will be useful since then we might lighten safety measures.

The relationship between this probability and α (where α will be a safety level define by health authorities, possibly $q^* = 0.6$) will help support the design of an effective deconfinement policy. Deconfinement can be conducted in this way with levels $q_1^*,\dots$, fixed by health authorities with the technical support of their staff.

CONCLUSIONS AND FINAL REMARKS

With the proposed protocol in this study, we show how to estimate infected individuals in a population by COVID-19, but who are asymptomatic. We point out that antivirus test may be used to estimate the percentage of asymptomatic persons and that the COVID-19 Test may be used to estimate the percentage of immune persons.

ACKNOWLEDGMENTS

MO research is partially supported by National Funds through FCT (Fundação para a Ciência e a Tecnologia), project 343 UIDB/04674/2020 (CIMA) and by the project ref. H2020-MSCA-RISE-2020/101007950, with the title "DecisionES - Decision Support for the Supply of Ecosystem Services under Global Change", funded by the Marie Curie International Staff Exchange Scheme.

REFERENCES

1. J. L. P. Fernandes, *Impacto da dimensão da lista de utentes dos médicos de família na qualidade dos cuidados prestados*. Dissertation (in Portuguese), 2017. https://hdl.handle.net/10216/108589
2. R. Arnab, *Survey Sampling Theory and Applications*. Academic Press, 1st Editions, 2017, ISBN-13: 9780128118481.
3. S. S. Wilks, *Mathematical Statistics*, John Wiley & Sons, 1962.
4. T. J. Rao and C. R. Rao, *Handbook of Statistics*. Elsevier, 2016.