

An application of the geometric distribution for assessing the risk of infection with SARS-CoV-2 by location



Athanasios Polymenis

Assistant Professor, Department of Economics, University of Patras, University Campus at Rio, Rio-Patras, Greece

Submission: 02-08-2021

Revision: 29-08-2021

Publication: 01-10-2021

ABSTRACT

Background: COVID-19 disease has quickly become a pandemic because of high infectiousness among populations. In order to assess for infectiousness reproduction numbers have been widely used in existing literature. A recent research has estimated reproduction numbers by location in order to account for differences among infectiousness. **Aims and Objective:** The aim of the present work was to use reproduction numbers in order to obtain an estimate for the mean number of contacts of some patient up to the first transmission of the disease to some other person in a specific location. The objective is to help authorities collect even more information on disease transmission. **Materials and Methods:** Our study was based on results obtained in a recent research work where replicated values of reproduction numbers have been generated in order to estimate medians and 95% confidence intervals for these numbers and thus assess infectiousness. We proposed a method that relies on a geometric distribution such that the parameter for this distribution is related to these reproduction numbers. **Results:** Our method showed that large reproduction numbers were associated with small mean numbers of contacts up to first transmission of the disease whereas small reproduction numbers were associated with a large mean number of contacts up to first transmission of the disease for all the locations under investigation. **Conclusion:** Our method has shown a nice performance and thus it can be considered as an alternative tool, by comparison with reproduction numbers, for assessing the risk of disease transmission, which is particularly useful when locations are investigated separately.

Key words: COVID-19; Reproduction number; Geometric distribution; Infectiousness

INTRODUCTION

Since the outbreak of a disease called COVID-19, caused by coronavirus SARS-CoV-2, there has been a huge scientific effort for preventing its quick spread among populations around the world. For this purpose, reproduction numbers appeared to be useful for evaluating the risk of transmitting this disease, and so, they have been widely used in existing literature.¹⁻⁵ In a somehow different approach, the present paper aims to estimate the mean number of contacts made by some patient up to the first person that this patient will infect, using simulated reproduction numbers reported in a very recent research paper⁶. Our method is implemented using a geometric

distribution that models probabilities of number of contacts up to first infection. A comparison between our results and those from reproduction numbers⁶ is provided. Since locations are investigated separately, it has been argued that their particular characteristics are taken into account, thus yielding more accurate reproduction numbers than for the whole population.⁶ Our method is expected to also have this advantage, and should help authorities adopt even more efficient countermeasures like, for example, better managing the number of people in some location, taking into consideration that the first infected person is in turn expected to transmit the disease to other people gathered in the same location.

Access this article online

Website:

<http://nepjol.info/index.php/AJMS>

DOI: 10.3126/ajms.v12i10.38783

E-ISSN: 2091-0576

P-ISSN: 2467-9100

Copyright (c) 2021 Asian Journal of Medical Sciences



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Address for Correspondence:

Dr. Athanasios Polymenis, Assistant Professor, Department of Economics, University of Patras, University Campus at Rio, Rio-Patras, Greece, Pin: 26504. **Mobile No:** +30- 6977742999. **E-mail:** athanase@upatras.gr

MATERIALS AND METHODS

As mentioned in the previous section, our method is based on a model which has been recently proposed by Hata et al.⁶ for estimating reproduction numbers in the context of infectiousness from COVID-19. Since we also use this model for our purpose, we now provide a short presentation of the model, as follows. Let us denote by p_i the probability that a patient infects some other person at place i ; then according to this model, the probability of secondary infections from this patient is a geometric sequence of the form p_i, p_i^2, p_i^3, \dots and the reproduction number considered at place i , denoted as R_i , can be written as a power series of the form $R_i = p_i + 2p_i^2 + 3p_i^3 + \dots = \sum_{k=1}^{\infty} k p_i^k = p_i / (1 - p_i)^2$. A crucial point that insures that this model is valid is the assumption that information concerning patients who have been reported as not having infected anyone has been excluded from the analysis.⁶ Concerning the numerical experiment for this model, a non-parametric bootstrap procedure provided estimates for the median reproduction numbers as well as 95% confidence intervals from 1000000 replicated values of the reproduction number concerning 10 different locations⁶. We will use these results for our purpose, as we now explain.

Our approach stems from remarking that if $p_i (\neq 0,1)$ is the probability of success (that is the probability of infecting a person) and so, $q_i = 1 - p_i$ is the probability of failure (that is the probability of not infecting a person), then the reproduction number R_i can be used in order to compute q_i and subsequently, p_i as this will be described in the sequel. Then, let X be the number of contacts from some patient at place i up to the first success, with $X=1,2,3,\dots$, then X follows a geometric distribution with parameter p_i (for the definition, properties and applications of a geometric distribution see, for example, Holmes et al.⁷). Our goal is then to find an estimate and 95% confidence intervals

(denoted as CIs in this article) for $E(X)$, using the fact that the geometric distribution has the advantage of being quite easy to apply. So in the sequel we will estimate values of $E(X)$ that correspond to the median reproduction number and to the lower and upper bounds of the 95% CIs for the reproduction number presented in Table 1. Since $E(X)$ accounts for the mean number of contacts by a patient up to his first success, it is expected to take smaller values in case $R_i > 1$ (in which case there may be an epidemic) and larger values in case $R_i \leq 1$ (in which case the disease will decline or stay stable and so no epidemic is expected). A comparison of values of $E(X)$ with corresponding values for R_i is provided in the next sections. In order to estimate $E(X)$ we need to estimate the probabilities p_i that correspond to R_i by solving equation $R_i = p_i / (1 - p_i)^2$, using for R_i on one hand, values of median(R_i) from Table 1, and, on the other hand, values of the lower and upper bounds of the 95% CIs for the replicated reproduction numbers of Table 1. Calculation of $E(X)$ follows then very easily since, by theory pertaining to the geometric distribution⁷, $E(X)$ is just $1/p_i$. This procedure is then repeated for all infection locations under consideration, which are reported in Table 1. This is simply done as follows. Let $1 - p_i = q_i$, then we have $R_i = p_i/q_i^2$ and so $p_i = R_i q_i^2$ i.e. $1 - q_i = R_i q_i^2$, thus leading to solving second order equation $R_i q_i^2 + q_i - 1 = 0$. The positive root of this equation will then provide an estimate for q_i , and thus p_i can be estimated using equation $p_i = 1 - q_i$.

RESULTS

We now implement our method, which was previously described, using the results provided by Hata et al.⁶ and our results are presented in Table 1. We digress to say that the reproduction numbers of Hata et al.⁶ have been also included in Table 1 (in brackets) since we use them for our calculations and, furthermore, they can be immediately compared to our results for inferential reasons. In order to show how our method is practically performed, let us consider, for example, infection place “Home”. In this case

Table 1: Comparison of mean numbers of contacts up to first transmission with reproduction numbers from Hata et al.⁶ (reproduction numbers are in brackets)

Location i	Median $E(X)$ (Median R_i)	95% CI	
		Lower bound $E^l(X)$ (Upper bound R_i^u)	Upper bound $E^u(X)$ (Lower bound R_i^l)
Home	1.869 (2.4752)	1.464 (6.7998)	2.3971 (1.2943)
Hospital	1.4391 (7.0849)	1.204 (28.9211)	1.7295 (3.2501)
Facility for Elderly People	1.7328 (3.2267)	1.202 (29.4718)	2.6266 (0.9927)
School	2.3931 (1.2331)	1.1715 (39.8456)	4.9628 (0.3160)
Nursery School	2.4565 (1.1580)	1.1327 (67.5018)	24.0327 (0.0453)
University	1.732 (3.2321)	1.1522 (48.2487)	6.8252 (0.20114)
Restaurant	1.7705 (2.9824)	1.2712 (17.2831)	8.7253 (0.1462)
Night Entertainment	3.2337 (0.6481)	1.1293 (67.5018)	28.7026 (0.0374)
Karaoke	1.9953 (2.0143)	1.1656 (42.4974)	27.5391 (0.0391)
Workplace	2.3308 (1.3161)	1.4647 (6.7833)	3.7351 (0.4993)

median (R_i) = 2.4752 (where index i stands for “Home”), lower bound for the 95% CI of R_i is 1.2943 and upper bound for the 95% CI of R_i is 6.7998. In order to estimate $E(X)$ concerning median (R_i) we need to solve equation $(2.4752) q_i^2 + q_i - 1 = 0$. The positive root of this equation is $q_i = 0.4649$ and thus $p_i = 0.5351$. It results that $E(X) = 1/0.5351 = 1.869$. The same procedure is used in order to compute CIs for $E(X)$ using the lower and upper bounds of R_i (denoted R_i^L as and R_i^U). Indeed, for $R_i^L = 1.2943$ we obtain an upper bound for $E(X)$ denoted as $E^U(X) = 2.3971$, and for $R_i^U = 6.7998$ we obtain a lower bound for $E(X)$ denoted as $E^L(X) = 1.464$, and thus a 95% CI for $E(X)$ is [1.464, 2.3971]. We repeat this procedure for the remaining 9 infection places reported in Table 1, namely: “Hospital”, “Facility for Elderly People”, “School”, “Nursery School”, “University”, “Restaurant”, “Night Entertainment”, “Karaoke”, and “Workplace”. In this way we obtain mean numbers of contacts up to first transmission of the disease for all the locations under consideration.

DISCUSSION

In order to compare our results to those from Hata et al.⁶ we first mention that the reproduction number is significantly larger than 1 for locations “Home”, “Hospital” (since the lower bounds of the 95% CIs are larger than 1), and for “Facility for Elderly People” (where the lower bound of the 95% CI is approximately equal to 1).⁶ Furthermore, all three medians corresponding to these reproduction numbers turn out to be well above 1, which result is compatible with inference resulting from the CIs. Concerning our results, we remark that both the medians and the 95% CIs show that there is a very small mean number of contacts (between 1.2 and 2.6) up to the first success for locations “Home”, “Hospital” and “Facility for Elderly People” (see Table 1). These results are in line with our idea reported in the Materials and Methods section, namely that reproduction numbers clearly larger than 1 would correspond to small mean number of contacts up to the first success (see Materials and Methods section).

Concerning locations “School”, “Nursery School”, “Workplace”, and “Night Entertainment”, the reproduction numbers are not significantly larger than 1 at a 95% confidence level since the 95% CIs pertaining to these numbers include 1 (see Table 1). These results are compatible with corresponding ones for medians of reproduction numbers, since the latter are only marginally larger than 1 for the first three locations and smaller than 1 for the last location (see Table 1). Concerning our 95% CI results, we found the mean number of contacts up to the first success to be between 1 and 5 for “School”, 1 and 24 for “Nursery School”, 1.4 and 3.7 for “Workplace”,

and 1 and 29 for “Night Entertainment” (see Table 1) which are clearly wider intervals than those obtained for locations “Home”, “Hospital” and “Facility for Elderly People”. Thus, the mean number of contacts up to the first success can get much larger values for “School”, “Nursery School”, “Workplace” and “Night Entertainment” by comparison with the locations “Home”, “Hospital”, and “Facility for Elderly People”. Furthermore, values of Median $E(X)$ concerning locations “School”, “Nursery School”, “Workplace”, and “Night Entertainment” are higher than corresponding values obtained for “Home”, “Hospital” and “Facility for Elderly People” (see Table 1), which result complies with the above discussion on the CIs. Finally, note that “Night Entertainment”, which has the smallest median reproduction number among all locations reported in Table 1 (0.6481), has also the largest mean number of contacts up to the first success among all these locations (3.2337), which can get as large as 28 (see Table 1). The above results are thus compatible with our idea reported in the Materials and Methods section, namely that reproduction numbers smaller or approximately equal to 1 lead to large mean numbers of contacts up to the first success.

However, when considering the three remaining locations “University”, “Restaurant” and “Karaoke” some care is needed in interpreting the results. To make things clear, we note from the results reported in Table 1 that, although the 95% CIs for the reproduction number indicate that this number does not significantly differ from 1 at a 95% confidence level, medians of the reproduction numbers were found to clearly exceed 1 (in contrast with the case of locations “School”, “Nursery School”, “Workplace” and “Night Entertainment”). Furthermore, concerning the “Karaoke” location, some findings from previous research showed that the risk of transmission was high in such a location⁸ (and thus advised for countermeasures), which result complies with the median reproduction number but not with the corresponding 95% CI for this number, as shown in Table 1. Our results concerning these three locations have also the same kind of problem since although the mean number of contacts up to first disease transmission associated with median (R_i) varies between 1.7 and 2 (thus indicating large reproduction numbers), the 95% CIs indicate that the mean number of contacts up to first disease transmission vary between 1 and 7 (for “University”), 1.3 and 8.7 (for “Restaurant”), and 1 and 27.5 (for “Karaoke”), thus implying low infectiousness. We believe that data in hand are probably not enough for investigating about infectiousness in these locations, noting that this problem has also been reported by Hata et al.⁶

CONCLUSION

In the present study, a geometric distribution is proposed for modeling numbers of contacts from a patient up to the first transmission of the COVID-19 disease to some other person, that makes use of recent results concerning reproduction numbers. Our analysis shows that results pertaining to the mean number of contacts from a patient up to the first transmission of the disease comply with results from the reproduction numbers for all the locations under investigation, thus indicating a nice performance for our method. Finally, like for reproduction numbers, our method should also be more useful when the analysis is performed for a specific location, rather than for the whole population, since infectiousness varies among locations, and thus countermeasures aiming at reducing transmission of the disease will be taken according to the particulars of each location.

REFERENCES

- Lai CC, Shih, TP, Ko WC, Tang HJ and Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob Agents.* 2020; 55(3).
<https://doi.org/10.1016/j.ijantimicag.2020.105924>
- Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis.* 2020; 92:214-217.
<https://doi.org/10.1016/j.ijid.2020.01.050>
- Sugishita Y, Kurita J, Sugawara T and Ohkusa Y. Effects of voluntary event cancelation and school closure as countermeasures against COVID-19 outbreak in Japan. *Plos One.* 2020; 15(12).
<https://doi.org/10.1371/journal.pone.0239455>
- Locatelli I, Trächsel B and Rousson V. Estimating the basic reproduction number for COVID-19 in Western Europe. *Plos One.* 2021; 16(3):e0248731.
<https://doi.org/10.1371/journal.pone.0248731>
- Billah A, Mamun A and Nuruzzaman K. Reproductive number of coronavirus: a systematic review and meta –analysis based on global level evidence. *Plos One.* 2020; 15(11):e0242128.
<https://doi.org/10.1371/journal.pone.0242128>
- Hata A, Kurita J, Hata T, Sugawara T and Ohkusa Y. Estimating SARS-CoV-2 reproduction number by infection location in Japan. *J Health Sci Dev.* 2021; 4(1):24-29.
- Holmes A, Illowsky B, Dean S and Hadley K. Introductory business statistics. In: Openstax, eds. Rice University, USA, 2017, pp 213-217.
- Gu Y, Lu J, Su W, Liu Y, Xie C and Yuan J. Transmission of SARS-CoV-2 in the karaoke room: an outbreak of COVID-19 in Guangzhou, China, 2020. *Journal of Epidemiology and Global Health.* 2021; 11(1):6-9.
<https://doi.org/10.2991/jeqh.k.201007.001>

Authors Contribution:

AP- Literature review, research methodology, statistical analysis, manuscript writing and revision.

Work attributed to:

Department of Economics, University of Patras, University Campus at Rio, Rio-Patras, Greece, Pin: 26504.

Orcid ID:

Dr. Athanase Polymenis -  <https://orcid.org/0000-0002-6102-7747>

Source of Support: Nil, **Conflict of Interest:** None declared.