

Original Paper

# Age-Stratified Infection Probabilities Combined With a Quarantine-Modified Model for COVID-19 Needs Assessments: Model Development Study

Vena Pearl Bongolan<sup>1</sup>, PhD; Jose Marie Antonio Minoza<sup>1</sup>, BSc; Romulo de Castro<sup>2</sup>, PhD; Jesus Emmanuel Sevilleja<sup>3</sup>, MD, MPhil

<sup>1</sup>Department of Computer Science, University of the Philippines Diliman, Quezon City, Philippines

<sup>2</sup>Center for Informatics, University of San Agustin, Iloilo, Philippines

<sup>3</sup>National Center for Mental Health, Mandaluyong, Philippines

**Corresponding Author:**

Vena Pearl Bongolan, PhD

Department of Computer Science

University of the Philippines Diliman

UP AECH Bldg

Velasquez St

Quezon City, 1101

Philippines

Phone: 63 915 877 2298

Email: [bongolan@up.edu.ph](mailto:bongolan@up.edu.ph)

## Abstract

**Background:** Classic compartmental models such as the susceptible-exposed-infectious-removed (SEIR) model all have the weakness of assuming a homogenous population, where everyone has an equal chance of getting infected and dying. Since it was identified in Hubei, China, in December 2019, COVID-19 has rapidly spread around the world and been declared a pandemic. Based on data from Hubei, infection and death distributions vary with age. To control the spread of the disease, various preventive and control measures such as community quarantine and social distancing have been widely used.

**Objective:** Our aim is to develop a model where age is a factor, considering the study area's age stratification. Additionally, we want to account for the effects of quarantine on the SEIR model.

**Methods:** We use the age-stratified COVID-19 infection and death distributions from Hubei, China (more than 44,672 infections as of February 11, 2020) as an estimate or proxy for a study area's infection and mortality probabilities for each age group. We then apply these probabilities to the actual age-stratified population of Quezon City, Philippines, to predict infectious individuals and deaths at peak. Testing with different countries shows the predicted number of infectious individuals skewing with the country's median age and age stratification, as expected. We added a Q parameter to the SEIR model to include the effects of quarantine (Q-SEIR).

**Results:** The projections from the age-stratified probabilities give much lower predicted incidences of infection than the Q-SEIR model. As expected, quarantine tends to delay the peaks for both the exposed and infectious groups, and to "flatten" the curve or lower the predicted values for each compartment. These two estimates were used as a range to inform the local government's planning and response to the COVID-19 threat.

**Conclusions:** Age stratification combined with a quarantine-modified model has good qualitative agreement with observations on infections and death rates. That younger populations will have lower death rates due to COVID-19 is a fair expectation for a disease where most fatalities are among older adults.

(*J Med Internet Res* 2021;23(5):e19544) doi: [10.2196/19544](https://doi.org/10.2196/19544)

**KEYWORDS**

COVID-19; epidemic modeling; age stratification theory; infection probability; SEIR; mathematical modelling

## Introduction

The initial impression that came out of Wuhan, China, in late 2019 and early 2020 was that COVID-19 most affects older adult males with pre-existing conditions. Classic compartmental models like the susceptible-exposed-infectious-removed (SEIR) model all assume a homogenous population, and that everyone has equal chances of getting infected. The SEIR model is initialized by “dividing” the population into four compartments; people “progress” through being susceptible, to getting exposed, to being infectious, to getting removed, either via recovery with permanent immunity or death. Permanent immunity is a common assumption when modelling viral infections, and it is assumed here. In future, the model may be modified for temporary immunity, as soon as we get reliable data on reinfection rates. A scan of the preprints from various modeling efforts during the first quarters of 2020 gave high estimates for the peaks of the exposed and infectious groups (40% of the population, by

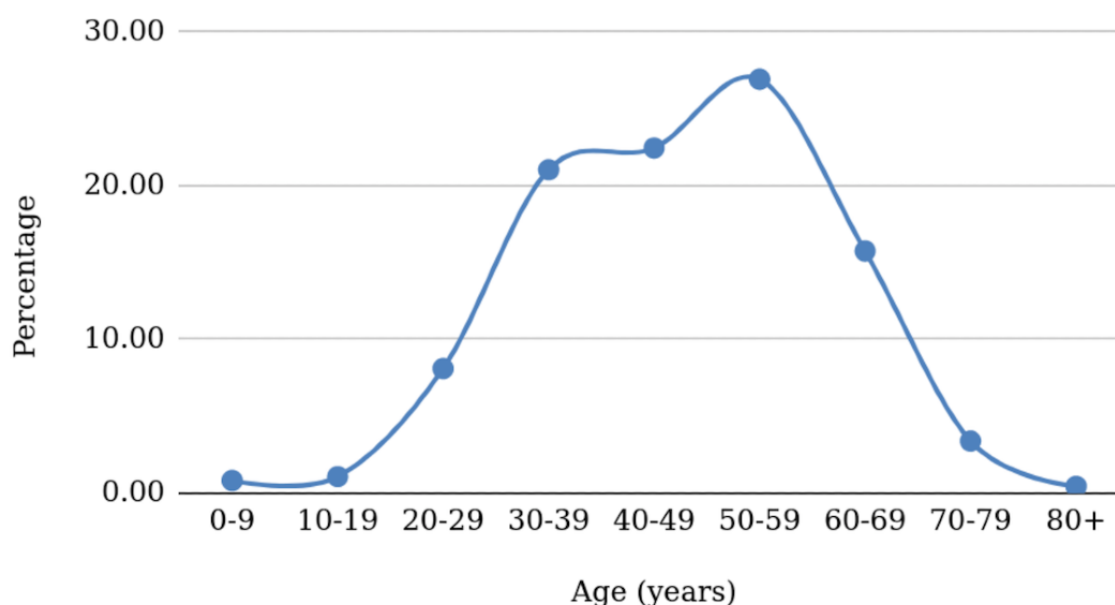
some estimates). Even our quarantine-modified model suffered from this, and this inspired us to use age-stratified infection probabilities, which gave us a lower bound for estimates.

## Methods

### Estimates by Age Stratification

This calculates the “mathematical expectation” of future infections per age group, by multiplying an age group’s infection probability by the population in that age group. Initially, we used the data of patients with COVID-19 in Hubei [1], stratified by ages. As data came in, we repeated the calculations with updated Quezon City data. We treat the percentages of incidence in each age group as a proxy or estimate for the corresponding probabilities of infection for people in the corresponding age group. The true probabilities are and continue to be unknown, but the scatter of the data from Hubei is consistent with the virus affecting older adults with pre-existing conditions more than other groups (Figure 1).

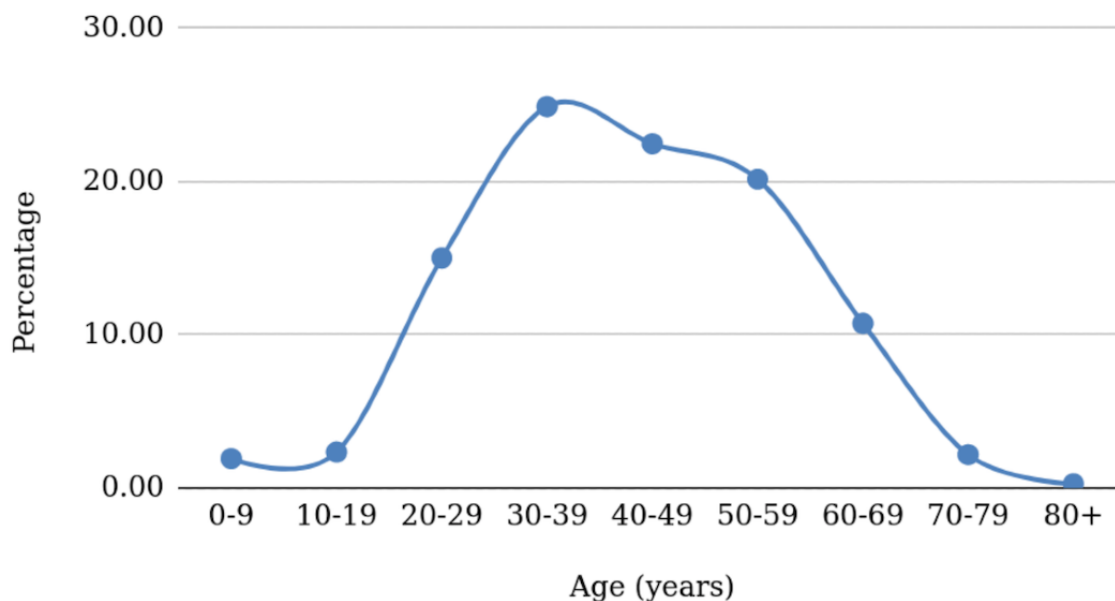
**Figure 1.** Projected age-stratified percentages of cases for China.



Next, we took the proxy probabilities from Hubei, and applied them to the actual age stratification of the Philippines. Due to its young population, this resulted in a “skewed to the right”

distribution (Figure 2) compared to the Hubei distribution, and the true distribution for the study area will be revealed as actual cases are reported.

**Figure 2.** Projected age-stratified percentages of cases for the Philippines.



The Philippines has a median age of 25.7 years [2] compared to China’s 38.4 years [2]. This means half of the population is aged <25.7 years, so more than half of the population will be in the “safer” age groups, with lower probabilities of getting infected, and significantly greater chances of survival if they should contract the disease. Those who do get infected account for only 10.2% of cases (see Table 1, sum of the percentages

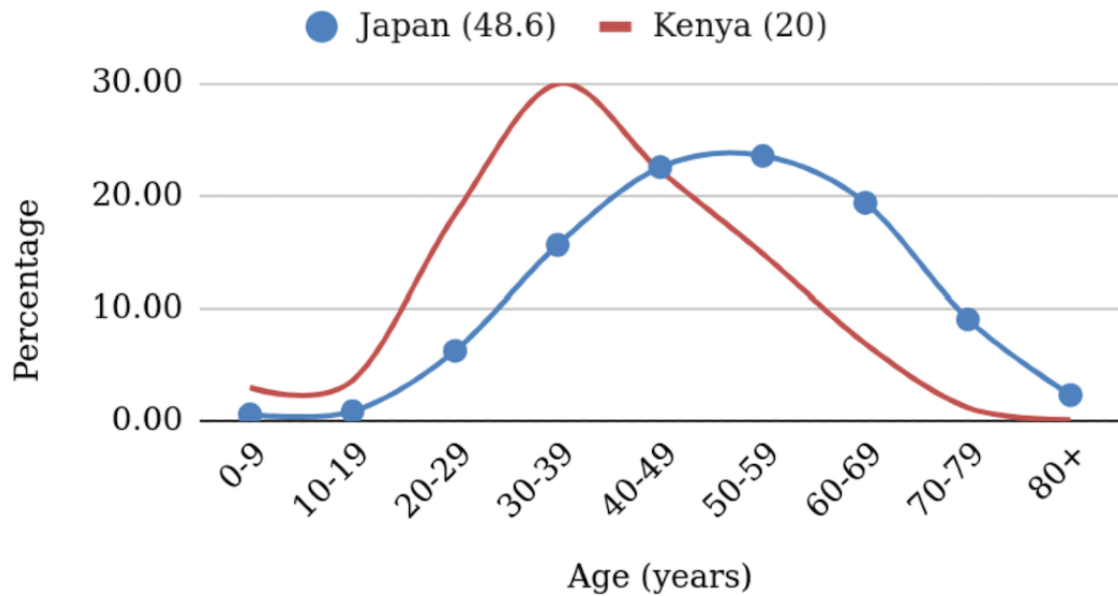
of those aged 0-29 years old). We expect this to be true for other countries with low median ages.

Using the United Nations World Population Prospects 2019 data [3], we did a similar experiment with Japan (median age 48.6 years) and Kenya (median age 20 years) [2]. We see the younger population also skewing right (Figure 3). Martinez [4] did similar calculations, but did not treat the age-stratified incidence as a proxy for infection probability.

**Table 1.** Quezon City, Philippines, projection of COVID-19 cases and mortality using COVID-19 data from Hubei, China.

Age group (years)	Hubei, China		Quezon City, Philippines		
	COVID-19 cases, %	Case fatality rate, %	Projected COVID-19 cases, n	Projected COVID-19 case distribution, %	Projected mortality, n
0-9	0.9	0	5164.60	1.60	0.00
10-19	1.2	0.2	7345.98	2.28	14.69
20-29	8.1	0.2	53,444.14	16.57	106.89
30-39	17	0.2	87,156.90	27.02	174.31
40-49	19.2	0.4	75,013.55	23.25	300.05
50-59	22.4	1.3	61,753.09	19.14	802.79
60-69	19.2	3.6	27,420.07	8.50	987.12
70-79	8.8	8.0	4588.82	1.42	367.11
≥80	3.2	14.8	699.83	0.22	103.57
Total	100	28.7	322,586.99	100	2856.54

**Figure 3.** Projected age-stratified percentages of cases for Japan and Kenya. The average age in each country is given in parentheses.



**Estimates Using a Quarantine-Modified SEIR Model**

The quarantine-modified SEIR equations are shown below.

$$S' = -\beta Q(t) \frac{SI}{N} \quad (1)$$

$$E' = \beta Q(t) \frac{SI}{N} - \sigma E \quad (2)$$

$$I' = \sigma E - \gamma I \quad (3)$$

$$R' = \gamma I \quad (4)$$

The insertion of the  $Q(t)$  term serves to control the  $S \times I$  or susceptible-infectious interactions, hence lowering exposure.  $Q(t)$  equal to one means no quarantine (ie, no change to the model). A  $Q(t)$  value of 0.4 means a 60% effective quarantine, while  $Q(t)=0.6$  means the quarantine is only 40% effective; therefore, the lower the  $Q(t)$ , the better. We allowed  $Q(t)$  to vary day by day (since cases began before the quarantine), and we also estimated the success of the quarantine. Henceforth, we refer to this model as Q-SEIR. Solution was via the Euler method, and time stepping was one day. We used Excel (Microsoft Corp) worksheets for our calculations.

The model was ground-truthed to the estimated number of exposed individuals at the national level on April 1, 2020 (N=7400) [5]. From the nationally reported number of exposed individuals (patients under investigation + persons under monitoring), Quezon City represents almost 10% of cases (~740); the Q-SEIR model predicted 705 cases.

**Results**

Applying the Hubei infection probabilities on Quezon City with an age distribution as shown in Table 1 (from the 2015 Census, projected to 2020 at a 2% growth rate [6]) gave an estimate of 322,586 infectious individuals (accumulated, which we equate with the Q-SEIR peak), which accounts for less than 10% of the population of Quezon City. Deaths were predicted at 22,390

cases (6.94%), which lies between the World Health Organization (WHO) morbidity estimates of 5.58% [7] and the 12.47% reported in Italy [8]. These estimates were what was available as of April 2020, and will be updated in the discussion.

**Discussion**

**Principal Results**

Initial reports (around April 2020) estimated the Philippines' death or mortality rate to be at 4.70% (4.05%-5.43%) [8]. This high estimate may be explained by sampling bias, wherein severe cases may have been overrepresented because of a lack of testing. Those who are infectious but are asymptomatic or exhibit mild symptoms should also be equally represented in the testing guidelines (at the moment, they are not), as well as those who were infectious with no symptoms and have recovered.

We tried the calculations from [4] using the death rate, which was reported at 2.3% for China [1]. This gave a much lower number of around 2857 deaths, for a Quezon City death rate of 0.89%. This figure is surprisingly low compared to the 6.94% projected using the estimated infection probability. The latest estimate of the Philippine COVID-19 death rate is 1.82% [9].

The delay in test reporting (estimate of 5-7 days [10]) factors in the estimation of the initial E-I-R values. In addition, this delay is compounded by the incubation period and, in our opinion, moves the quarantine effect further down from the actual date of implementation (March 15, 2020). We started the Q-SEIR simulation on March 20, 2020, with no quarantine assumed because the steep jump in cases occurred on this date; 60% effective quarantine was set for April 2, 2020.

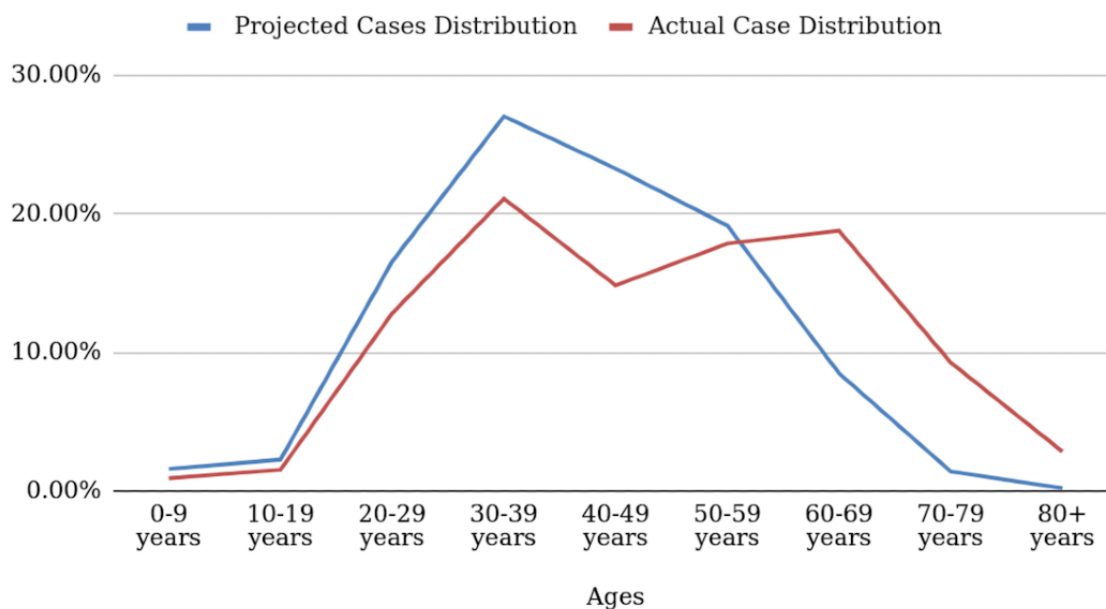
The Q-SEIR model predicted 14.00% of the population will be infectious (I) at the peak. The two methods now give us a low and high estimate for Quezon City: infectious individuals will peak between 9.95% (from age stratification) and 14.00% (from Q-SEIR) of the population, around the third week of May 2020.

These figures seemed high compared to the reported incidence during the same period, but were not in any way unique (compared to modeling done in other countries). At that time, the suspicion was that actual cases were undertested/underreported by as much as 90% (ie, only 10% of cases were being detected). Nevertheless, this range of values

serves as a guide for planners in anticipating the need for personal protective equipment, mass testing, hospital beds, and other basic needs.

Figure 4 shows the scatter of cases for Quezon City projected by the Hubei data, and the actual scatter as of May 2020.

**Figure 4.** Projected versus actual age-stratified case distribution for Quezon City.



### Limitations

Like many research groups, we were and continue to be hampered by a lack of reliable data in a usable format. Even now, we refer to multiple data sets including the Philippine Department of Health's DataDrop, WHO, and other data sources.

### Comparison With Prior Work

We were one of the first groups to forward the theory of age stratification when it comes to modeling COVID-19 infections. Recently, we came across works by Balabdaoui et al [11], Undurraga et al [12], and the WHO [13]. However, we are uniquely using incidence percentages as proxies for infection probabilities with good results.

### Conclusions

In conclusion, age stratification predicted the scatter of cases for Quezon City fairly well. It also predicted later observations of lower cumulative confirmed cases for the same time period (eg, 6587 cases per million compared to Italy's 58,417 cases per million as of March 28, 2021); there is a color-coded world map in [14]. Some African countries have lower median ages than the Philippines, and they have generally lighter colors than Europe and North America. The best prediction is the noticeably lower death or mortality rate of 1.82% compared to Italy's 3.05% and Indonesia's 2.7% (all as of March 29, 2021) [9]. Since COVID-19 disproportionately affects older adults more than younger populations, we expect the Philippines to have a lower mortality rate than countries with older populations (eg, Italy has a median age of 47.3 years and Indonesia has a median age of 29.7 years, compared to 25.7 years in the Philippines). Later work can use actual infection probabilities to include the effect of age stratification in the model.

### Acknowledgments

VPB thanks the Alumni Engineers of the University of the Philippines Diliman for their support of the study (UPCoE Covid-19 Response). Additionally, she thanks her coauthors, RdC and JES, who worked pro bono; true heroes! Finally, the authors are grateful to the Engineering Research and Development for Technology (ERDT) Consortium, Republic of the Philippines, for shouldering publication costs.

### Conflicts of Interest

None declared.

### References

1. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020[J]. *China CDC Weekly* 2020 Mar 03;2(8):113-122. [doi: [10.46234/ccdcw2020.032](https://doi.org/10.46234/ccdcw2020.032)]
2. Central Intelligence Agency. Field Listing - Median age. The World Factbook. URL: <https://www.cia.gov/the-world-factbook/field/median-age/> [accessed 2020-04-11]
3. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2019. URL: <https://population.un.org/wpp/DataQuery/> [accessed 2020-04-11]
4. Martinez R. COVID-19 mortality calculator. Public Tableau. URL: <https://public.tableau.com/profile/ramon.martinez#!/vizhome/COVID-19mortalitycalculator/COVID-19mortalitycalc> [accessed 2020-04-06]
5. Department of Health. COVID-19 Case Tracker. URL: <https://www.doh.gov.ph/covid-19/case-tracker> [accessed 2020-04-01]
6. Philippines: Metro Manila. City Population. URL: <https://www.citypopulation.de/php/philippines-metromanila-admin.php?adm2id=137404> [accessed 2020-04-06]
7. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> [accessed 2020-04-06]
8. Oke J, Heneghan C. Global Covid-19 Case Fatality Rates. The Centre for Evidence-Based Medicine. URL: <https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/> [accessed 2020-04-06]
9. Coronavirus (COVID-19) death rate in countries with confirmed deaths and over 1,000 reported cases, by country. Statista. URL: <https://www.statista.com/statistics/1105914/coronavirus-death-rates-worldwide/> [accessed 2020-04-01]
10. COVID-19 test results from RITM out in 5 to 7 days, but not for long, DOH says. CNN Philippines. URL: <https://www.cnnphilippines.com/news/2020/3/27/COVID-19-test-results-from-RITM-out-in-5-to-7-days,-but-not-for-long,-DOH-says.html> [accessed 2020-04-01]
11. Balabdaoui F, Mohr D. Age-stratified discrete compartment model of the COVID-19 epidemic with application to Switzerland. *Sci Rep* 2020 Dec 04;10(1):21306 [FREE Full text] [doi: [10.1038/s41598-020-77420-4](https://doi.org/10.1038/s41598-020-77420-4)] [Medline: [33277545](https://pubmed.ncbi.nlm.nih.gov/33277545/)]
12. Undurraga E, Chowell G, Mizumoto K. COVID-19 case fatality risk by age and gender in a high testing setting in Latin America: Chile, March-August 2020. *Infect Dis Poverty* 2021 Feb 03;10(1):11 [FREE Full text] [doi: [10.1186/s40249-020-00785-1](https://doi.org/10.1186/s40249-020-00785-1)] [Medline: [33531085](https://pubmed.ncbi.nlm.nih.gov/33531085/)]
13. World Health Organization. Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, 17 March 2020. 2020 Mar 17. URL: <https://apps.who.int/iris/handle/10665/331656> [accessed 2021-05-12]
14. Philippines: Coronavirus Pandemic Country Profile. Our World in Data. URL: <https://ourworldindata.org/coronavirus/country/philippines> [accessed 2020-04-01]

## Abbreviations

**SEIR:** susceptible-exposed-infectious-removed

**WHO:** World Health Organization

*Edited by C Basch; submitted 22.04.20; peer-reviewed by N Mohammad Gholi Mezerji, D Cebo; comments to author 26.10.20; revised version received 31.03.21; accepted 05.04.21; published 31.05.21*

*Please cite as:*

Bongolan VP, Minoza JMA, de Castro R, Sevilleja JE

*Age-Stratified Infection Probabilities Combined With a Quarantine-Modified Model for COVID-19 Needs Assessments: Model Development Study*

*J Med Internet Res* 2021;23(5):e19544

URL: <https://www.jmir.org/2021/5/e19544>

doi: [10.2196/19544](https://doi.org/10.2196/19544)

PMID: [33900929](https://pubmed.ncbi.nlm.nih.gov/33900929/)

©Vena Pearl Bongolan, Jose Marie Antonio Minoza, Romulo de Castro, Jesus Emmanuel Sevilleja. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 31.05.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.