American Journal of
Epidemiology & Public Health

Short Communication

Astonishing Undocumented COVID-19 Infections in the United States as of Mid-March 2020 and Potentially Unstoppable Epidemics without Intensified Measures

Zhifu Fu[1#], Cheng Long[2#], Tieyong Zeng[3*] and Xinmiao Fu[4*]

[1]*Anxi AIER Eye Hospital, Anxi County, Fujian Province 362400, China*

[2]*Department of Orthopaedics, Sichuan University West China Hospital, Chengdu City, Sichuan Province 610041, China*

[3]*Department of Mathematics, the Chinese University of Hong Kong, Shatin, NT, Hong Kong 99999, China*

[4]*College of Life Sciences, Fujian Normal University, Fuzhou City, Fujian Province 350117, China*

[#]*These authors contributed equally to this work*

***Address for Correspondence:** Xinmiao Fu, College of Life Sciences, Fujian Normal University, Fuzhou City, Fujian Province 350117, China, ORCID ID: orcid.org/0000-0003-3361-6904; E-mail: xmfu@fjnu.edu.cn
Tieyong Zeng, Department of Mathematics, the Chinese University of Hong Kong, Shatin, NT, Hong Kong 99999, China, E-mail: zeng@math.cuhk.edu.hk

Submitted: 23 July 2020; **Approved:** 29 July 2020; **Published:** 30 July 2020

Cite this article: Fu Z, Long C, Zeng T, Fu X. A Astonishing Undocumented COVID-19 Infections in the United States as of Mid-March 2020 and Potentially Unstoppable Epidemics without Intensified Measures. American J Epidemiol Public Health. 2020;4(3): 070-074. https://dx.doi.org/10.37871/ajeph.id33

ISSN: 2644-0032

ISSN: 2644-0032

ABSTRACT

People raise deep concerns about the COVID-19 outbreaks in the United States. Here we explored Central Limit Theorem-based approach to estimate potential COVID-19 infections in the US. Assuming that the 18000 attendees of the 2020 AIPAC conference are a random sample chosen from the population of the US (i.e., mimicking a sampling survey of nationwide COVID-19 infection) and that 480 NBA players are a unique sample with 10 times higher susceptibility to COVID-19 infections than common American males, total infections in the US as of 10 and 19 March 2020 were estimated to be over 90,000 and 680,000 (99.99% Confidence Interval), respectively. We speculate that daily new confirmed cases in the US are likely less than daily new infections under current testing capacity and astonishing basal infections such that the undocumented infections are still expanding, resulting in unstoppable epidemics. Simulation analyses reveal that increasing the testing capacity and decreasing the transmissions of both undocumented and confirmed COVID-19 cases are critical for containing this severing crisis in the US. To our best knowledge, this is the first study estimating the undocumented COVID-19 infections by applying Central Limit Theorem.

Keywords: COVID-19; Coronavirus; United States; Pandemic; Central limit theorem; NBA; AIPAC; Testing density; Basic reproductive number; Undocumented COVID-19 cases

INTRODUCTION

The COVID-19 outbreak is hitting many countries including the United States, in which there have been over 310,000 confirmed cases as of April 4 [1]. People raise deep concerns about the future trajectory of the outbreaks in the US. However, it is technically difficult to precisely estimate the potentially infected cases in an area [2], largely due to the limitation of testing capacity [3,4] and the presence of a large portion of undocumented COVID-19 infections [5]. Here we achieved this goal by applying Central Limit Theorem-based analysis on two unique social events. First, among 18,000 attendees of the 2020 American Israel Public Affairs Committee (AIPAC) conference (held 1-3 March 2020), there were reported at least five COVID-19 positive cases as of 10 March [6]. Second, among a total of 480 NBA players (excluding 17 players from Toronto Raptors, Canada) there have been ten COVID-19 positive players as of 19 March [7].

METHODS

Central limit theorem-based analysis

The central limit theorem states that when an infinite number of successive random samples are taken from a population, the sampling distribution of the means of those samples will become approximately normally distributed with mean μ and standard deviation σ/√N as the sample size (N) becomes larger, irrespective of the shape of the population distribution [8]. Accordingly, Central Limit Theorem can be applied to estimate the status of the total sample at certain Confidence Interval (CI) based on the status of the surveying sample [8,9]. Total COVID-19 infections in the US were estimated by Central Limit Theorem-based analysis as follows.

Suppose that X_1, X_2, X_3, . . . , X_N are independent identically distributed random variables with mean value μ and variance σ_2. We are interested in the sum of these variables:

$$S_N := X_1 + X_2 + \ldots + X_N . \quad (1)$$

The celebrated Central Limit Theorem [8] states that

$$\frac{S_N - \mu N}{\sigma\sqrt{N}} \approx N(0.1) \quad (2)$$

where N(0, 1) follows the normal distribution when N is big.

Now assume that X_1, X_2, X_3, . . . , X_N following the Bernoulli distribution are all the American people, and each person is infected by COVID-19 with probability p (Note: after the COVID-19 test, X_i is numbered as 1 indicating the infection or as 0 indicating no infection). Accordingly, the quantity S_N is just the total number of the infected persons in the US. Readily, we have,

$$\mu = p, \sigma^2 = pq,$$

with q = 1 - p. We set N = 332 million (i.e., the population size of the US in 2020) and $p = \frac{5}{18000}$ (i.e., 5 COVID-19 positive cases among the 18000 attendees).

Set a fixed, and let $b = \frac{a-pN}{\sqrt{pqN}}$. Then we have the probability calculation as follows,

$$P(S_N \geq a) = P\left(\frac{S_N - pN}{\sqrt{pqN}} \geq \frac{a - pN}{\sqrt{pqN}}\right)$$

$$= P\left(\frac{S_N - pN}{\sqrt{pqN}} \geq b\right)$$

$$= \frac{2}{\sqrt{2\pi}} p \int_b^{+\infty} e^{-\frac{x^2}{2}} dx$$

$$= \frac{2}{\sqrt{2\pi}} p \int_{\infty}^{-b} e^{-\frac{x^2}{2}} dx$$

$$= \phi(-b)$$

where we let $\phi(z) = \frac{2}{\sqrt{2\pi}} p \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$

If we want that $\Phi(b)$ = 99.99%, then we can simply take b =-3.8.

Accordingly, we have,

$$a = pN + b\sqrt{pqN} \quad (3)$$

For this a, we are 99.99% sure that $S_N \geq a$.

Simulation of undocumented infections over time

Daily increase of undocumented COVID-19 infected cases is expressed as follows:

$$\Delta I = \frac{IR_o}{D} + \frac{C_a R_{ao}}{D} - \Delta C - A \quad (4)$$

where I is the number of undocumented COVID-19 infected cases in a defined region/country; R0 is the basic reproductive number of the

ISSN: 2644-0032

undocumented COVID-19 infected cases; D is time (number of days) for transmission; Ca is the number of active confirmed COVID-19 cases; Ra0 is the basic reproductive number of active confirmed COVID-19 cases; C is the cumulative number of confirmed COVID-19 cases; A is the self-healing number of undocumented cases (i.e., without transmission ability any more) and is expressed as 0.8*I/D (assuming 80% of undocumented COVID-19 cases can be self-healing).

RESULTS

Total COVID-19 infections in the US were over 91,068 as of 10 March 2020

It was reported that among 18,000 attendees of the 2020 AIPAC conference (held 1-3 March 2020), there were at least five COVID-19 positive as of 10 March [6]. Total COVID-19 infections in the US as of 10 March 2020 were then estimated by Central Limit Theorem-based analysis (according to equation 3) to be over 91,068 (99.99% CI) (black squares in figure 1A). Here we assumed that the 18,000 attendees of AIPAC are a random sample chosen from the total population (332 million) of US, i.e., mimicking a sampling survey for nationwide COVID-19 infection. As a matter of fact, these attendees include: "More than 3,600 students from more than 630 campuses" and "AIPAC members from across the country" as major ones [10]. As such, the attendees would be diversified in occupation, geographic location and age, and their COVID-19 infection status will not overestimate but can largely reflect the nationwide infection status of the US.

Total COVID-19 infections in the US were over 686,762 as of 19 March 2020

It was reported that among a total of 480 NBA players (excluding 17 players from Toronto Raptors, Canada) there have been ten COVID-19 positive players as of 19 March [7]. Total COVID-19 infections in the US as of 19 March were then estimated by Central Limit Theorem-based analysis to be over 686,762 (red circles in figure 1A). Here, we assumed that NBA players are a unique sample chosen from the total population of American males but they are arbitrarily 10 times more susceptible to COVID-19 infections than the latter based on the following considerations: i) NBA teams are located in typical major cities with high populations across the country, thus conceivably reflecting the overall COVID-19 status in the US; ii) NBA players are largely aged 20-39 years and would not be more susceptible to COVID-19 infections than older males in general [11,12]; iii) However, NBA players have more direct contacts with each other, more frequent travels, and more close interactions with the general public than common people, thus being more susceptible to COVID-19 infections. In addition, American females are assumed equally susceptible as males [11].

In contrast to our estimation, only 472 and 7087 confirmed cases were reported in the US as of 10 and 19 March by WHO situation reports [1] (open red circles in figure 1B), respectively, and the corresponding numbers of confirmed cases were 949 and 13548 [13] according to the data from Worldometer (open squares in figure 1B), respectively. Furthermore, total COVID-19 tests in the US as of March 19, 2020 were approximately 100, 000 (solid blue circles in figure 1B), far less than potential total infections we estimated (over 680,000). Such huge number of undocumented infected cases definitely lead to the uncontrolled spread of COVID-19 in the US.

Interruption of COVID-19 transmission and increasing COVID-19 testing capacity are both critical for containing the COVID-19 epidemics in the US

To find out whether intensified prevention and control measures could contain the extremely severe COVID-19 outbreaks in the US, we performed simulation studies to evaluate the effects of transmission level and testing capacity on the progression of undocumented infections according to equation 4 (for detail, refer to the Methods section). Data displayed in figure 2A indicate that, as expected, when the basic reproductive number (R0) is decreased from 2.0 to 1.2, the undocumented infections will dramatically decrease, and if it is further decreased to 0.8 then the undocumented infections will be diminished. Notably, when R0 is 1.2 the number of undocumented infections first decreases and then goes up (blue curve in figure 2A), which is due to the transmission from the active confirmed cases.

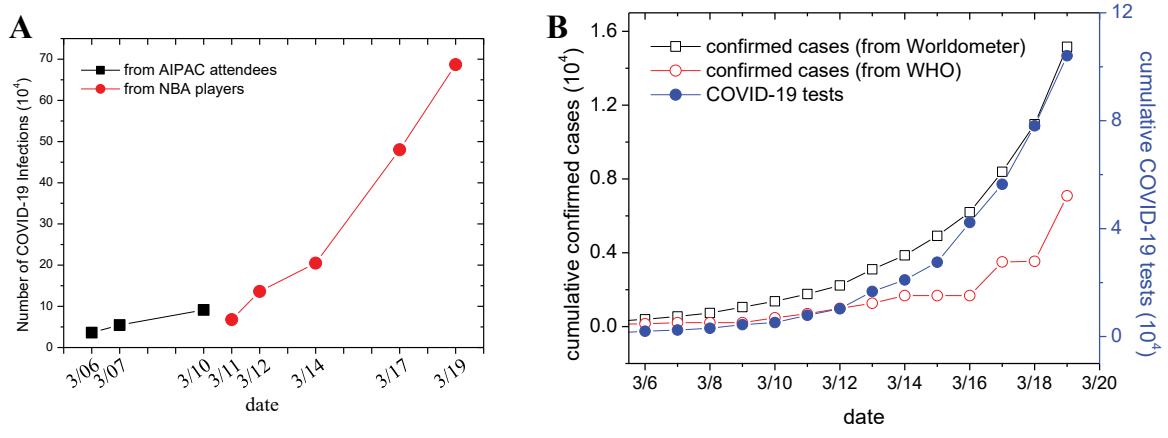

Figure 1: Estimation of COVID-19 infections in the US over time.
(**A**) Number of COVID-19 infections was calculated according to equation (3) based on the number of infected attendees of the AIPAC conference (black square) and NBA players (red circles) as of the indicted dates. There have reported 2, 3 and 5 positive persons among 18000 attendees as of 6, 7 and 10 March, respectively [6]; and 1, 2, 3, 7 and 10 infected players as of 11, 12, 14, 17 and 19 March, respectively [7]. The susceptibility of NBA players to COVID-19 infections is assumed 10 times higher than that of common American males, and American females are assumed equally susceptible to infections as males. (**B**) Cumulative COVID-19 confirmed cases and tests over time. Cumulative confirmed cases are obtained from the websites of both Wordometer [13] and WHO [1].

ISSN: 2644-0032

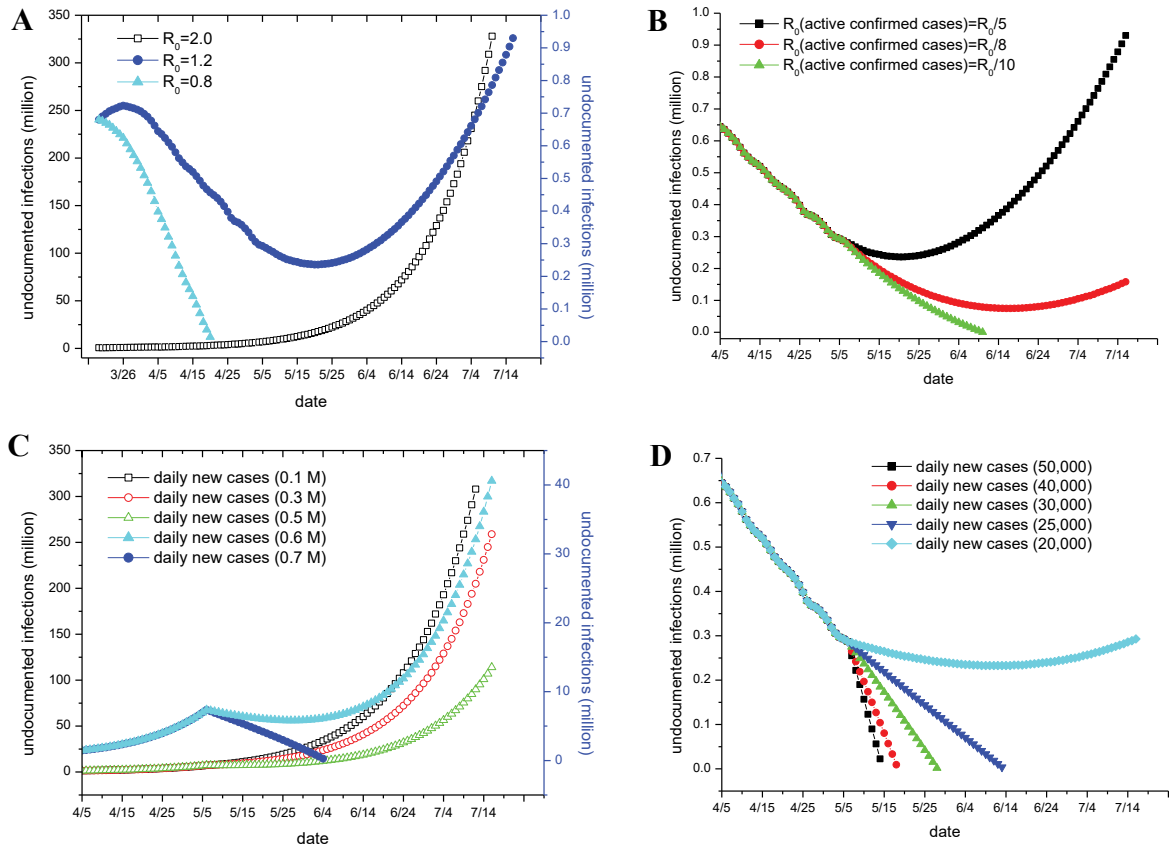

Figure 2: Simulation of undocumented COVID-19 infections in the US over time under different conditions.
(**A**) Undocumented COVID-19 infections in the US since March 19, assuming that R_0 is 2.0, 1.2 or 0.8. Initial infections on March 19 are assumed as 680,000 by referring to the results in figure 1A. (**B**) Undocumented COVID-19 infections in the US, assuming that R_0 for undocumented infections is 1.2 and that R_0 for active confirmed cases since May 6 is 1.2/5, 1.2/8 or 1.2/10. (**C**,**D**) Undocumented COVID-19 infections in the US under the conditions of different testing capacity, assuming that R_0 for undocumented infections is 2.0 (panel C) or 1.2 (panel D). Daily new confirmed cases since May 6 are assumed as 0.1, 0.3, 0.5, 0.6 and 0.7 million in panel C, and 50,000, 40,000, 30,000, 25,000 and 20,000 in panel D.

Further analyses reveal that if R0 of the active confirmed cases could be reduced, for instance, by strict hospitalization rather than by home-based isolation [14], the undocumented infections can also be diminished (red curve in figure 2B).

We also found that testing capacity can significantly change the trends of the epidemics. Under the condition of rapid spread of the virus (i.e., R0 = 2.0), the epidemic cannot be stopped unless daily new cases are up to 0.7 million (blue curve in figure 2C), which is far beyond the current testing capacity of the US. However, if R0 can be decreased to 1.2 (i.e., more quarantine measures are applied to common people and confirmed cases), then the epidemic can be stopped when daily new cases are over 25,000 (blue curve in figure 2D).

Together, these simulations results suggest that reducing the transmissions of the virus from undocumented infected cases and active confirmed cases, as well as increasing the testing capacity are both critical for containing the potentially unstoppable epidemics in the US. In support of our conclusion, recent modelling study suggests institutional, not home-based, isolation could contain the COVID-19 outbreak [14], which is apparently achieved by reducing the transmissions of the active confirmed cases (particularly the transmission in households). Consistently, Fangcang shelter hospitals are considered to be critical for successfully containing the COVID-19 epidemic in Wuhan City, China [15,16].

DISCUSSION

In many countries, the epidemics are mitigating and daily new cases are declining [13]. However, the US is still in growing phase. Most likely, in the context of a large pool of undocumented or unidentified infected cases, testing capacity is relatively insufficient such that a largely number of daily new infections has been made. More severely, such daily new infections might be larger than daily new confirmed cases such that undocumented infections in the US are still expanding. In other words, the epidemic is unstoppable till the virus has spread to all the populations.

Although our simulation analyses reveal that increasing the testing capacity (Figure 2C & 2D) and reducing the transmissions of the virus from undocumented and active confirmed cases (Figure 2A & 2B) enable the unstoppable epidemic to become stoppable, the measures to achieve these goals appear to be infeasible in the US at present. First, hospitalization level of COVID-19 patients in the US cannot be increased much, i.e., the spread of the virus from active confirmed cases cannot be limited. Second, societal and economic restorations as well as anti-racism rally take place cross the country, resulting in the uncontrolled spread of the virus from undocumented

cases. The only feasibly measure, in our opinion, is the dramatic expansion of COVID-19 testing capacity. As an example, COVID-19 testing capacity in China on late June has been expanded to 3,780,000 per day [17]. As such, the resurgence of COVID-19 in Beijing on early June has been successfully shutdown by mass testing (over 300,000 per day) [18].

ETHICAL APPROVAL

Ethical approval is not required for our study, as all the data utilized in our study were collected from newspapers and public-accessible websites.

ACKNOWLEDGMENTS

This work is support by the National Natural Science Foundation of China (No. 31972918 and 31770830 to XF).

AUTHOR CONTRIBUTIONS

X.F. conceptually designed the study; Z.F. and C.L. collected data and performed analyses; T.Z. established the Central Limit Theorem-based approach; X.F. wrote the manuscript.

REFERENCES

1. WHO: Coronavirus disease (COVID-2019) situation reports. https://bit.ly/3hOwl9w
2. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. Lancet. 2020; 395: 689-697. DOI: 10.1016/S0140-6736(20)30260-9.
3. The lost month: How a failure to test blinded the U.S. to COVID-19. https://bit.ly/3jWl4Vp
4. Maxmen A. Thousands of coronavirus tests are going unused in US labs. Nature. 2020; 580: 312-313. DOI: 10.1038/d41586-020-01068-3.
5. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science. 2020; 368: 489-493. DOI: 10.1126/science.abb3221 (2020).
6. 2 more AIPAC attendees test positive for coronavirus, bringing total to 5. https://bit.ly/3096umE
7. Coronavirus: Kevin Durant, Donovan Mitchell, Rudy Gobert among NBA players who tested positive for COVID-19. https://bit.ly/2X5eYJv
8. Durrett R. Probability Theory and Examples Part of Cambridge Series in Statistical and Probabilistic Mathematics. 5th ed. Duke University, North Carolina: Cambridge University Press; 2019. ISBN: 9781108473682
9. Perez-Rodriguez F, Zwietering MH. Application of the Central Limit Theorem in microbial risk assessment: High number of servings reduces the coefficient of variation of food-borne burden-of-illness. Int J Food Microbiol. 2012; 153: 413-419. DOI: 10.1016/j.ijfoodmicro.2011.12.005
10. AIPAC: AIPAC Policy Conference 2020. https://bit.ly/3jRr4zt.
11. Epidemiology Working Group for NCIP epidemic response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an oubreak of 2019 novel coronavirus diseases (COVID-19) in China. Chin J Epidemiol. 2020; 41: 145-151. DOI: 10.3760/cma.j.issn.0254-6450.2020.02.003
12. COVID-19 cases by age in some states of USA. https://bit.ly/33690fk, https://bit.ly/2BDDMke, https://on.nyc.gov/39yVORi, https://bit.ly/3faGw6e
13. Worldometer: COVID-19 coronavirus pandemic. https://bit.ly/30ZARuH
14. Dickens BL, Koo JR, Wilder-Smith A, Cook AR. Institutional, not home-based, isolation could contain the COVID-19 outbreak. Lancet. 2020; 395: 1541-1542. DOI: 10.1016/S0140-6736(20)31016-3.
15. Chen S, Zhang Z, Yang J, Wang J, Zhai X, Barnighausen T. Fangcang shelter hospitals: A novel concept for responding to public health emergencies. Lancet. 2020; 395: 1305-1314. https://bit.ly/3f35jcA
16. Zhang Z, Yao Y, Wang Y, Long C, Fu X. Wuhan and Hubei COVID-19 mortality analysis reveals the critical role of timely supply of medical resources. Journal of Infection. 2020; 81: 147-178. DOI: 10.1016/j.jinf.2020.03.018
17. COVID-19 testing capacity of China is up to 3,780,000 per day (in Chinese). https://bit.ly/39zpBcD
18. COVID-19 testing capacity in Beijing has been expanded up to 300,000 per day (in Chinese). https://bit.ly/2DddN3x