

This is an Accepted Manuscript for *Infection Control & Hospital Epidemiology* as part of the Cambridge Coronavirus Collection.

DOI: 10.1017/ice.2020.61

**Words: 1400 (w/o Abstract, References, Title, Figure, Table)**

**Commentary/editorial material**

# Identification of COVID-19 Can be Quicker through Artificial Intelligence framework using a Mobile Phone-Based Survey in the Populations when Cities/Towns Are Under Quarantine.

Arni S.R. Srinivasa Rao, Ph.D.\*

Professor,

Division of Health Economics and Modeling,

Department of Population Health Sciences, Medical College of Georgia, Augusta

University, Augusta, USA Director, Laboratory for Theory and Mathematical

Modeling,

Division of Infectious Diseases, Department of Medicine,

Medical College of Georgia, Augusta, GA 30912, USA.

Department of Mathematics

Augusta University,

1120, 15th Street, AE 1015

Augusta, GA, 30912, USA

Tel: +1-706-721-3786 (office)

Email: [arrao@augusta.edu](mailto:arrao@augusta.edu)

\*Corresponding author (email: [arrao@augusta.edu](mailto:arrao@augusta.edu))

Jose A. Vazquez, M.D., FACP, FIDSA

Chief and Professor, Division of Infectious Diseases,

Department of Medicine, Medical College of Georgia,

Augusta University,

1120, 15th Street, AE 1015

Augusta, GA, 30912, USA

Tel: +1-706-723-0105 (office)

**Abstract:** We are proposing to use machine learning algorithms to be able to improve possible case identifications of COVID-19 more quicker when we use a mobile phone-based web survey. This will also reduce the spread in the susceptible populations.

Key words: modeling, machine learning, novel coronavirus, disease spread.

Emerging and novel pathogens are a significant problem for global public health and technology can assist in faster identification of possible cases to bring timely interventions. This is especially true for viral diseases that are easily and readily transmissible and have asymptomatic infectivity periods. The novel Coronavirus (SARSCoV2) described in December 2019 (COVID-19) has resulted in major quarantines throughout the world, including major cities, villages and public areas throughout China [1-3] to prevent further spread. As of February 25<sup>th</sup> 2020, the World Health Organization's situational data indicates that there were about 77780 confirmed cases, including 2666 deaths due to COVID-19, including cases in 25 countries [4]. The majority of the deaths reported so far have been in China only [5]. The Centers for Disease Control and Prevention (CDC) and World Health Organization have issued interim guidelines in order to protect the population, and to attempt to prevent the further spread of COVID-19 from infected individuals[6]. Since cities and villages throughout China are unable to accommodate such large numbers of infected individuals, and be able to maintain the quarantined. China has built several new hospitals in an attempt to manage the infected individuals [7]. It is imperative that we evaluate novel models in an attempt to control the rapidly spreading virus [8].

In order to reduce the time to identification of a person under investigation (PUI) for the COVID-19 infection, and the rapid isolation of this individual, we propose to collect the basic travel history along with the more common manifestations using a phone-based online survey. Such collected data can be used to assist in the preliminary screening and early identification of possible COVID-19 infected individuals. Thousands of data points are able to be collected and processed

through an artificial intelligence (AI) framework which can ultimately evaluate individuals that may be infected and stratify them into no-risk, minimal-risk, moderate-risk, and high-risk of being infected with the virus. The identification of the high-risk cases can then be quarantined earlier, thus decreasing the chance of spread.

Table 1 is inserted here.

See Appendix I for the details on the steps involved in data collection on all the respondents independent of whether or not they think they are infected. The AI algorithm described in Appendix II is to identify possible case identifications and send alerts to the nearest health clinic as well as to the respondent for an immediate health visit, we call this as an “alert for health check recommendation for COVID-2019. In case the respondent is unable to commute to the health center, the health department can then send an alert to a mobile health unit so they can then do door-to-door assessments and even testing for the virus. This generates alert for mobile health check recommendation for 2019-nCoV (MHCRC). If a respondent does not have an immediate risk of having symptoms or signs related to the viral infection, then the AI-based health alert will be sent to the respondent to notify them that there is no current risk of COVID-2019. Figure 1 summarizes the outcomes of data collection and identification of possible cases.

Figure 1 is inserted here

The data recorded in step 5 of the algorithm using signs and symptoms will be collected prior to both the groups who have received alerts HCRC or MHCRC (for possible identification and assessment) and NCRC (for non-identified respondents). These are explained in steps (iii) and (iv) in the Appendix II. The extended analysis proposed will help to understand if there is any association with different socio-demographic variables and the manifestations such as fever and signs and lower respiratory infections, including cough and SOB in individuals defined as either with and without possible infection.

Applications of AI and deep learning argued to be useful tools in assisting diagnosis and treatment decision making [10-11]. There were studies which promoted disease detection through AI models [12-15]. Use of mobile phones [16-19] and web based portals [20-21] have been tested successfully in health related data collection. However, one need to apply such techniques in a timely way for faster results. Apart from cost-effectiveness, the proposed modeling will be of great assistance in identifying and controlling when populations are closed due to virus spread. In addition to these, our proposed algorithm can be easily extended to identify individuals who might have any mild symptoms and signs.

**Acknowledgements:** We thank Professor N.V. Joshi, Indian Institute of Science, Bengaluru and Mr. P. Sashank, CEO Exaactco Compusoft Global Solutions Pvt. Ltd., Hyderabad, India for their editorial comments.

**Authors contributions:** All authors contributed to the writing and approved the final version. ASRSR developed the methods and wrote the first draft of the paper. JAV contributed in clinical verbiage writing, inputs and editing into the draft.

**Funding:** We do not have funding to report to this study.

**Conflicts:** Not applicable.

## References

1. More Chinese cities shut down as novel coronavirus death toll rises, Channelnewsasia.com (accessed on February 10, 2020).
2. Chinese villages build barricades to keep coronavirus at bay, Financial Times NewsPaper (accessed on February 10, 2020).
3. Transit going in and out of Wuhan, China is being shut down to contain coronavirus, Businessinsider.com (accessed on February 10, 2020)
4. The WHO COVID-19 situation report 36 (accessed on February 26, 2020).
5. Coronavirus explained: Death toll crosses 900, surpasses SARS and MERS, www.cnet.com (accessed on February 10, 2020).
6. Interim guidance for persons who may have 2019 Novel Coronavirus (2019-nCoV) to prevent spread in homes and residential communities, [www.CDC.gov](http://www.CDC.gov) (accessed on February 10, 2020).
7. How China Built Two Coronavirus Hospitals in Just Over a Week, The Wall Street Journal (accessed on February 10, 2020).
8. Expert: Better models, algorithms could help predict and prevent virus spread, The Augusta Chronicle Newspaper (accessed on February 11, 2020)
9. Flowchart to Identify and Assess 2019 Novel Coronavirus, [www.CDC.gov](http://www.CDC.gov) (accessed on February 11, 2020).
10. Liang, H., Tsui, B.Y., Ni, H. et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med* 25, 433–438 (2019). <https://doi.org/10.1038/s41591-018-0335-9>
11. Rao, A.S.R.S, Diamond, M.P. Deep Learning of Markov Model-Based Machines for Determination of Better Treatment Option Decisions for Infertile

Women. *Reprod. Sci.* 27, 763–770 (2020). <https://doi.org/10.1007/s43032-019-00082-9>

12. D. B. Neill, "Using Artificial Intelligence to Improve Hospital Inpatient Care," in *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 92-95, March-April 2013.  
doi: 10.1109/MIS.2013.51
13. Rajalakshmi, R., Subashini, R., Anjana, R.M. *et al.* Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye* **32**, 1138–1144 (2018). <https://doi.org/10.1038/s41433-018-0064-9>
14. Zeinab A, Roohallah AI, Mohamad R, Hossein M, Ali AY (2017). Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm, *Computer Methods and Programs in Biomedicine*, Volume 141, 2017, Pages 19-26.
15. V. B. Kumar, S. S. Kumar and V. Saboo, "Dermatological disease detection using image processing and machine learning," *2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR)*, Lodz, 2016, pp. 1-6.
16. Tomlinson, M., Solomon, W., Singh, Y. *et al.* The use of mobile phones as a data collection tool: A report from a household survey in South Africa. *BMC Med Inform Decis Mak* 9, 51 (2009). <https://doi.org/10.1186/1472-6947-9-51>
17. Ballivian, A, Azevedo, J P, and Durbin, W. 2015. Using Mobile Phones for High-Frequency Data Collection. In: Toninelli, D, Pinter, R & de Pedraza, P (eds.) *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, Pp. 21–39. London: Ubiquity Press. DOI: <http://dx.doi.org/10.5334/bar.c>. License: CC-BY 4.0.

18. Braun R, Catalani C, Wimbush J, Israelski D. Community health workers and mobile technology: a systematic review of the literature. *PLoS One*. 2013;8(6):e65772. Published 2013 Jun 12. doi:10.1371/journal.pone.0065772
19. Bastawrous, A., & Armstrong, M. J. (2013). Mobile health use in low- and high-income countries: an overview of the peer-reviewed literature. *Journal of the Royal Society of Medicine*, 106(4), 130–142.  
<https://doi.org/10.1177/0141076812472620>
20. D. Paolotti A. Carnahan V. Colizza K. Eames J. Edmunds G. Gomes C. Koppeschaar M. Rehn R. Smalenburg C. Turbelin S. Van Noort A. Vespignani (2013). Web- based participatory surveillance of infectious diseases: the InfluenzaNet participatory surveillance experience, *Clin Microbiol Infect* 2014; 20: 17– 21
21. Fabic MS, Choi YJ, Bird S (2012) A systematic review of Demographic and Health Surveys: data availability and utilization for research. *Bull World Health Organ* 90: 604–612.

## Appendix I: Steps involved in data collection through mobile phones

We have developed our data collection criteria based on CDC's Flowchart to Identify and Assess 2019 Novel Coronavirus [9] and added additional variables for the extended utility of our efforts in identifying infected and controlling the spread (see Table 1).

## Appendix II: Algorithm

Let  $O_1, O_2, O_3, O_4, O_5$  be the outputs recorded during the data collection steps 1 through 5 described in the Appendix I. There are three outputs within  $O_2$  which are given as

$$O_2 = \{O_{2G}, O_{2A}, O_{2R}\},$$

and nine pairs of outputs within  $O_5$  which are given as

$$O_5 = \left\{ \begin{array}{l} (O_{5A}, D_{5A}), (O_{5B}, D_{5B}), (O_{5C}, D_{5C}), (O_{5D}, D_{5D}), \\ (O_{5E}, D_{5E}), (O_{5F}, D_{5F}), (O_{5G}, D_{5G}), (O_{5H}, D_{5H}), \\ (O_{5I}, D_{5I}) \end{array} \right\},$$

where the pair  $(O_{5i}, D_{5i})$  for  $i = A, B, \dots, I$  represents the respondents response on presence or absence of  $i^{th}$  sign and symptom ( $O_{5i}$ ) and duration of corresponding sign and symptom ( $D_{5i}$ ).

(i) If the set of identifiers,  $I_1$ , for

$$I_1 = \{O_3, O_{5A}, O_{5B}, O_{5C}\}$$

is equal to one of the elements of the set  $C_1$ , for

$$C_1 = \left\{ \begin{array}{l} (1,1,1,1) \\ (1,1,1,0) \\ (1,1,0,0) \\ (1,0,1,1) \\ (1,0,0,1) \\ (1,0,1,0) \\ (1,1,0,1) \end{array} \right\},$$

for a respondent, then, send HCRC or MHCRC. If  $I_1$  is not equal to any of the elements of the set  $C_1$ , then proceed to test criteria (iii).

(ii) If the set of identifiers,  $I_2$ , for

$$I_2 = \{O_4, O_{5A}, O_{5B}, O_{5C}\}$$

is equal to one of the elements of the set  $C_1$  then send HCRC or MHCRC to that respondent, else proceed to the test criteria (iv).

(iii)

If  $I_1$  is equal to one of the elements of the set  $C_2$ , for

$$C_2 = \left\{ \begin{array}{l} (0,1,1,1) \\ (0,1,1,0) \\ (0,1,0,0) \\ (0,0,1,1) \\ (0,0,0,1) \\ (0,0,1,0) \\ (0,1,0,1) \end{array} \right\}$$

then the respondent will be sent an NCRC alert.

(iv)

If  $I_2$  is equal to one of the elements of the set  $C_2$ , then the respondent will be sent an NCRC alert.

Comparison of test criteria results of (iii) and (iv) with their corresponding geographic and socio-demographic details will yield further investigations of signs and symptoms based on whether or not an individual in the survey has traveled to coronavirus affected areas or had contact with any person who is known to have COVID-19. As we are focusing, in this note, only identification of cases, further analysis techniques are out of the scope. However, our approach is flexible to capture various other associations within the populations.

### **Appendix III: Further computations on the data collected**

Suppose  $n$  and  $m$  are individuals in a region who have responded and not responded, respectively for a mobile-based online survey. Responses are randomly associated and not depended on the sickness due to the virus. The pair

$$\left( \frac{n}{n+m}, \frac{m}{n+m} \right)$$

gives us the proportions who have responded and not responded in that region. Note that we can compute  $\frac{m}{n+m}$  because the value  $m$  is known to us in that region. Let  $n_1$  out of  $n$  are possible cases identified through our algorithm and  $m_1$  out of  $m$  are possible cases of the virus which were never identified by the algorithm (because  $m$

number of individuals never responded to the survey). Since  $n$  and  $m$  are known to us, one of the following relations will hold:

$$\{n > m, n = m, n < m\} \text{ --- (A2.1)}$$

We will know which one of the relations listed in (A2.1) is true. When  $n > m$ , one of the following relations will hold:

$$\left\{ \frac{n_1}{n} > \frac{m_1}{m}, \frac{n_1}{n} = \frac{m_1}{m}, \frac{n_1}{n} < \frac{m_1}{m} \right\} \text{ --- (A2.2)}$$

But we will never know which one of the relations in (A2.1) is true because  $m_1$  were never identified by the algorithm. For example, suppose 2000 individuals respond and 500 individuals do not respond to the survey, and 400 were identified as a possible number of cases by the algorithm. If there are 100 possible cases of virus (which we do not have a mechanism to count) out of 500 who never responded, then the relation

$$\frac{n_1}{n} = \frac{m_1}{m}$$

is true. Similarly, we will see other relations of (A2.2) could arise when  $n > m$ . Using a similar argument, we can verify that when other relations of (A2.1) are true, we still be unsure which of the relations in (A2.1) is true. The  $2 \times 2$  contingency options are provided in Figure 2 to visualize the data to be generated through the proposed method.

Figure 2 is inserted here.

**Theorem:** Let there be  $N$  individuals in a region. The probability that  $n_1$  cases identified through the AI framework given that there are  $n$  individuals responded to the survey is  $\frac{n_1 N}{n^2}$ .

**Proof:**

Let  $N = n + m$ , and let

$$U = \{u_1, u_2, \dots, u_n\}$$

be the collection of  $n$  individuals responded,

$$V = \{v_1, v_2, \dots, v_m\}$$

be the collection of  $m$  individuals not responded. Suppose

$$U_1 = \{u_{a_1}, u_{a_2}, \dots, u_{a_{n_1}}\} \subset U$$

be the collection of respondents who are identified as possible cases. Here  $U \cup V$  can be thought of the region as shown in (a),  $U$  as shown in (b) and  $U_1$  as (c) shown in Figure 1.

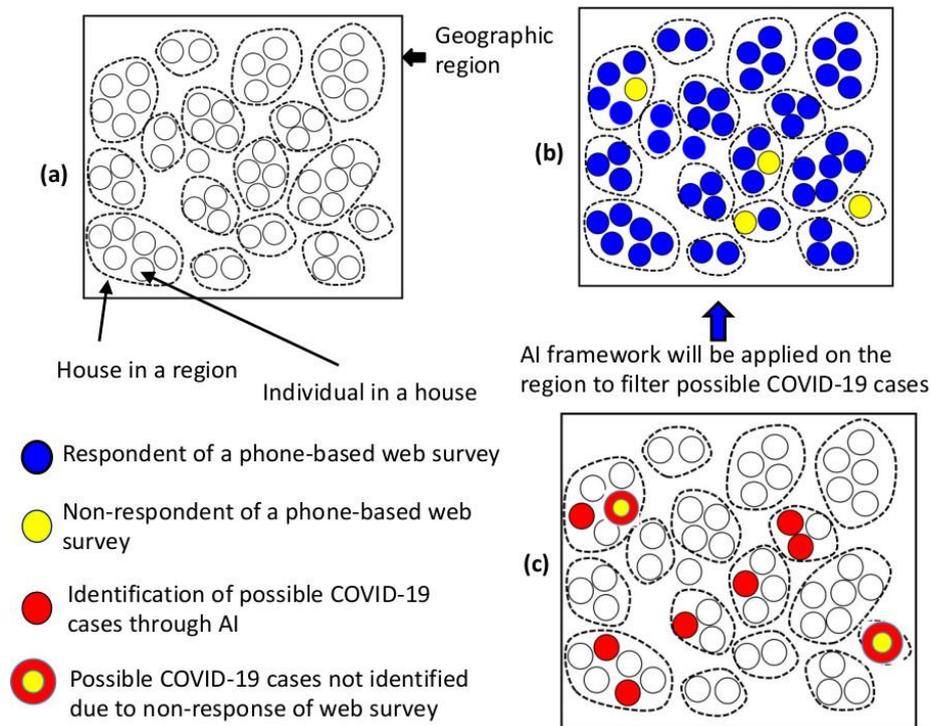
Suppose we define two events  $E_1$  and  $E$  using the sets  $U, V$  and  $U_1$  as below:

$E_1$ :  $n_1$  out of  $n$  responded cases are identified through the algorithm

$E$ :  $n$  out of  $N$  have responded to the survey.

The conditional probability of the event  $E_1$  given the event  $E$ , say,  $P(E_1/E)$ , is computed as

$$P(E_1/E) = \frac{P(E_1 \cap E)}{P(E)} = \frac{\frac{n_1}{n}}{\frac{n}{N}} = \frac{n_1 N}{n^2}. \blacksquare$$



**Figure 1.** Conceptual framework of data collection and possible COVID-19 identification. (a) A geographical region (could be a city/county/town/village) with households in it, (b) Respondents and non-respondents of a phone-based web survey, (c) Possible identified cases of COVID-19 among the survey respondents and possible cases of COVID-19 among non-respondents of the survey.

		Possible cases Identified through AI	
		Yes	No
Individuals responded for a phone-based web survey	Yes	$n_1$	$n - n_1$
	No	0	$m$

**Figure 2.** Number of possible cases identified through AI framework versus number of individuals responded to a phone-based web survey.

Table 1. Steps involved in the collection of data through a mobile phone-based survey.

- Step 1: Record the location details of the house/apartment from where the respondent uses a phone-based web survey / or the respondent's usual place of stay.
- Step 2: Record demographic information like gender (G) (1-male, 2-female, 3-others), age (A), race (R) (1-white, 2-black, 3-Hispanics, 4-Others)
- Step 3: Have you traveled to (or living in) any of the COVID-19 affected areas/countries in the last 14 days?  
(Yes=1/No=0)
- Step 4: Have you had any close contact with a person who is known to have COVID-19 during the last 14 days?  
(Yes=1/No=0)
- Step 5: Record the presence or absence of signs and symptoms listed below and the duration of each of the signs and symptoms if yes to any of the signs and symptoms.  
A) fever (Yes=1/No=0), if yes, then the duration in days ----  
B) cough (Yes=1/No=0), if yes, then the duration in days ----  
C) shortness of breath (Yes=1/No=0), if yes, then the duration in days ----  
D) myalgia or fatigue (Yes=1/No=0), if yes, then the duration in days ----  
E) sputum production (Yes=1/No=0), if yes, then the duration in days ----  
F) headache (Yes=1/No=0), if yes, then the duration in days ----  
G) diarrhea (Yes=1/No=0), if yes, then the duration in days ----  
H) pneumonia in both lungs (Yes=1/No=0), if yes, then the duration in days ----
- Step 6: Enter the details of steps 1-5 above for any dependents or other individuals who live in the same location and do not have access to web-based survey.