

## Research



**Cite this article:** Kim H, Paul A. 2021

Automated contact tracing: a game of big numbers in the time of COVID-19. *J. R. Soc. Interface* **18**: 20200954.

<https://doi.org/10.1098/rsif.2020.0954>

Received: 24 November 2020

Accepted: 1 February 2021

### Subject Category:

Life Sciences—Mathematics interface

### Subject Areas:

biomathematics

### Keywords:

COVID-19, SARS-CoV-2, contact tracing, disease mitigation

### Author for correspondence:

Ayan Paul

e-mail: [apaul2@alumni.nd.edu](mailto:apaul2@alumni.nd.edu)

<sup>†</sup>These authors contributed equally to this work.

# Automated contact tracing: a game of big numbers in the time of COVID-19

Hyunju Kim<sup>1,2,3,†</sup> and Ayan Paul<sup>4,5,†</sup>

<sup>1</sup>Beyond Center for Fundamental Concepts in Science, and <sup>2</sup>School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA

<sup>3</sup>ASU-SFI Center for Biosocial Complex Systems, Arizona State University and Santa Fe Institute, Tempe, AZ, USA

<sup>4</sup>DESY, Notkestraße 85, 22607 Hamburg, Germany

<sup>5</sup>Institut für Physik, Humboldt-Universität zu Berlin, 12489 Berlin, Germany

HK, 0000-0002-0768-4017; AP, 0000-0002-2156-4062

One of the more widely advocated solutions for slowing down the spread of COVID-19 has been automated contact tracing. Since proximity data can be collected by personal mobile devices, the natural proposal has been to use this for automated contact tracing providing a major gain over a manual implementation. In this work, we study the characteristics of voluntary and automated contact tracing and its effectiveness for mapping the spread of a pandemic due to the spread of SARS-CoV-2. We highlight the infrastructure and social structures required for automated contact tracing to work. We display the vulnerabilities of the strategy to inadequate sampling of the population, which results in the inability to sufficiently determine significant contact with infected individuals. Of crucial importance will be the participation of a significant fraction of the population for which we derive a minimum threshold. We conclude that relying largely on automated contact tracing without population-wide participation to contain the spread of the SARS-CoV-2 pandemic can be counterproductive and allow the pandemic to spread unchecked. The simultaneous implementation of various mitigation methods along with automated contact tracing is necessary for reaching an optimal solution to contain the pandemic.

## 1. Introduction

A relentless and damaging battle is being fought against the spread of COVID-19. While several countries have managed to significantly slow down its spread, severe measures have had to be taken to do so and at great cost to the economic and social well-being of the nations. It is still not certain when a significant control over the spread of SARS-CoV-2 can be attained. Recent projections propose surveillance for the next few years [1], with several measures that will need to be put in place to minimize the cost of the pandemic to humankind. Automated contact tracing is one of these measures.

Contact tracing has been observed to be effective in previous pandemics (or epidemics) like the Ebola virus outbreak in 2014–2015 [2]. This pre-emptive method allows for the containment of the pathogen by isolating potentially infected individuals that have been traced. Extensive studies of manual contact tracing were done during the previous outbreak of the Ebola virus [3–5], SARS-CoV and MERS-CoV [6]. More recently, mathematical models have been formulated to study contact tracing assuming the disease spread to be quantifiable by the SIR model [7]. However, the efficacy of automated contact tracing during the SARS-CoV-2 pandemic requires a more detailed examination given the distinct difference in the prevalence of this pandemic from the ones in the recent past and the different modes of transmission of the pathogen.

Manual contact tracing is not very effective against pathogens that spread like the influenza virus but is more effective for containing smallpox and SARS-CoV and partially effective in containing foot-and-mouth disease [8].

The viral shedding patterns of SARS-CoV and MERS-CoV are similar [9,10] and show almost no presymptomatic transmission [11],<sup>1</sup> while Ebola is known to be transmitted through the bodily fluids of infected individuals after the onset of symptoms [13]. On the other hand, influenza shows a significant rate of viral shedding in the presymptomatic stage [14]. The important transmission characteristics of SARS-CoV-2 that set it apart from other HCoV pathogens like SARS-CoV and MERS-CoV and from Ebola are:

- SARS-CoV-2 transmission is driven by presymptomatic spreading like the influenza virus [15–17].
- The pathogen can be transmitted through the air in high contamination regions and through contaminated dry surfaces for several days [15,18,19] leading to its high transmission rates. This brings about additional challenges when the disease cannot be contained within an isolated envelope of a healthcare system. While a similar spreading pattern is seen in SARS-CoV and MERS-CoV, this makes SARS-CoV-2 more easily transmittable than Ebola.
- The ACE2 binding of SARS-CoV-2 is estimated to be relatively stronger than SARS-CoV and might explain its observed spreading characteristics [20–22].

In the early stages of the pandemic the reproduction number  $R_0$ , for SARS-CoV-2 was estimated to be 2.2–2.7 [23–27], similar to SARS-CoV.<sup>2</sup> The dispersion parameter is estimated to also be similar to that of SARS-CoV (close to 0.1), which could be causing superspreading [26,30–32].

In principle, automated contact tracing can be shown as an effective means of containing SARS-CoV-2 [31]. However, factors such as long delays from symptom onset to isolation, fewer cases ascertained by contact tracing, and increasing presymptomatic transmission can significantly impact how effective automated contact tracing will be in practice. Normally, a significant contact is defined as being within 2 m for at least 15 min.<sup>3</sup> Keeling *et al.* demonstrated that this can result in the detection of more than four out of five secondary infections but at the cost of tracing 36 contacts per individual [33]. Changes to the definitions of a significant contact can reduce the numbers traced. For example, if the minimum time required to be considered a significant contact is increased, the number of people needed to be traced will decrease at the cost of not being able to identify potentially infected individuals. Detailed modelling of SARS-CoV-2 transmission shows that the pandemic can be sustained just by presymptomatic transmission and that automated contact tracing can be used to contain the spread of the pathogen if there are no significant delays to identifying and isolating infected individuals and their contacts [34].

Considering all the factors that make contact tracing a different game for SARS-CoV-2, in this paper, we will examine in detail how much data and participation from the population will be needed to make automated contact tracing effective. This will give an estimate of the necessary scale of implementation of automated contact tracing and whether it will be feasible. The model that we build with parameters that are mostly independent of each other or factorized, will also allow for the estimation of the effects of various mitigation methods like the use of personal protective equipment (PPE) in enhancing the efficacy of automated contact tracing which we discuss before the discussion section. In this work, we address voluntary and automated contact tracing using proximity data alone

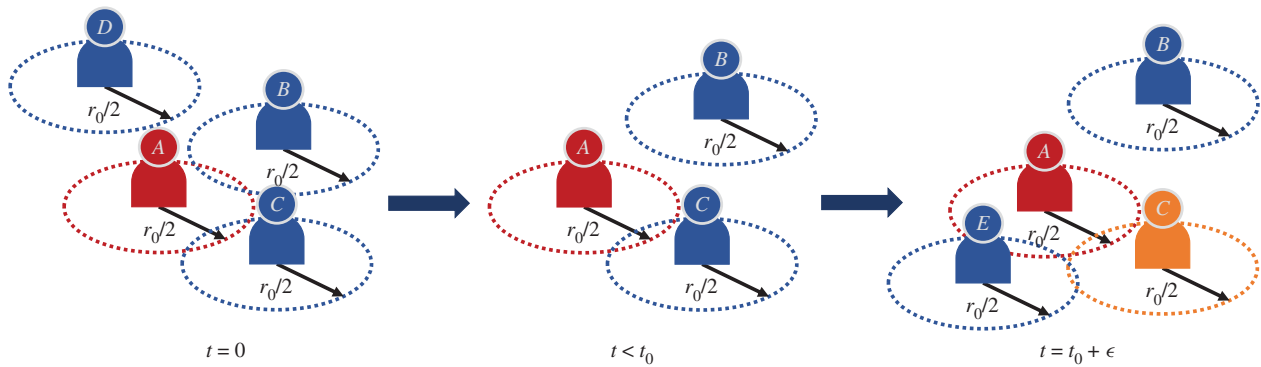
excluding methods such as the use of CCTV, credit card information, logging of identities of individuals during visits to locations and travels, etc., that have been successfully used by many countries like Singapore [35], Taiwan [36], South Korea [37] and China [34] for contact tracing.

## 2. Contact tracing for COVID-19

To judge the efficiency of contact tracing, it is crucial to determine whether an infectious disease can spread in the presymptomatic stage or from asymptomatic individuals. For a disease that can spread only in the symptomatic stage, the infected individuals can spread the disease to their contacts before they are isolated and to medical workers after they are isolated with varying probabilities. Of significance here is that after the initial period of ignorance of the population about a rising pandemic, infected individuals will be isolated with higher efficiency (even with manual contact tracing) resulting in the curtailment of the spread of the pathogen. How is contact tracing more effective in such diseases? Since the mobility of the infected individual usually sees a decline after the onset of symptoms, the number of contacts at risk become limited to only those who are most often in contact with the individuals and hence traceable manually. This allows the implementation of a manual contact tracing algorithm that identifies these neighbours and isolates or tests them as suggested in reference [8]. This was seen to be effective during the Ebola, MERS-CoV and SARS-CoV outbreaks.

However, the spreading of SARS-CoV-2 follows a very different pattern. With the prevalence of spreading of infection through presymptomatic and subclinical hosts, the number of individuals that might need to be traced can be very large. This has led to the belief that automated contact tracing in a wider gamut should be implemented. Most of the proposed solutions [31,33,34] require the use of historical proximity data to trace contacts. In the context of COVID-19, there are some obvious pitfalls in the algorithm:

- It is estimated that about 86% (95% CI: [82–90%]) of the infected cases in China were undocumented prior to the travel ban on 23 January 2020 generating 79% of the documented infections [23]. A large number of these undocumented cases experienced mild, limited or no symptoms and can hence go unrecognized. Similar results were reported by other studies [38,39]. It is not possible to trace all the contacts of these individuals since they will be partially reported leading to incomplete coverage of contact tracing.
- While it is assumed that the SARS-CoV-2 spreads within a proximity radius of  $r_0$  (assumed to be 2 m), not much is known about the probability of transmission,  $p_t$  when two individuals come within this domain of contact for a minimum contact time  $t_0$ . Assuming  $p_t$  to be large will lead to an unreasonably large estimate of the number of potential infections required to be traced in a crowded region like supermarkets, which remain open even during the period of social distancing. On the other hand, assuming  $p_t$  to be small will underestimate the number of infected contacts, especially because there might be other modes of transmission of SARS-CoV-2 that are not being considered. By definition  $p_t$  depends on the dynamics of disease transmission when



**Figure 1.** A depiction of automated contact tracing. The cross-section is denoted by the dashed circle and is of radius  $r_0/2$ . Interactions occur from  $t = 0$  to  $t = t_0 + \epsilon$  where  $\epsilon \ll t_0$ . A will be confirmed as COVID-19 positive in the future and C will be notified having come in contact with A. E might be notified if E stays in contact with A for a time period greater than  $t_0$ .

a healthy individual comes in significant contact with a infected individual. Moreover,  $p_t$  is not constant over  $r_0$  and also varies with the stage of infection the infected individual is at [17]. Several other factors contribute to the value of  $p_t$  in addition to the contagiousness of the disease including, but not limited to, the use of PPE, public awareness of the disease, whether the surroundings are an open or a closed area, air circulation (freely circulating as opposed to air conditioned), etc. [40].

The first pitfall can be alleviated by increasing the testing rate of individuals for viral RNA in the hope that a larger fraction of the asymptomatic or mildly symptomatic carriers can be traced. Increasing awareness can also help. The second pitfall can be alleviated when more detailed knowledge of the spread of SARS-CoV-2 is available and with the help of simulation of the spread of the disease in a population. For the rest of the work, we will assume  $p_t$  to be a variable and  $r_0$  to be fixed to 2 m [33].

The real-world applicability of automated contact tracing requires the examination of the effects of partial sampling of the population. The assumption that we are working with is that enrolment in automated contact tracing will be voluntary and individuals remain free to do one of the following:

- Choose not to enrol in the programme by either not using the application or the devices needed for tracing, including discontinuity in participation.
- Choose not to report on their health condition which is assumed to be voluntary.

Both types of occurrences have an effect of reducing the efficacy of automated contact tracing but in slightly different manners. In the first case, not subscribing to the service would not only remove an individual from the pool that is being notified but it also removes them from the pool of individuals that are reporting. In the second case, only the latter happens.

### 3. Modelling automated contact tracing

Since, in automated contact tracing, a significant contact has to be less than  $r_0$  distance away for time  $t_0$ , we describe every individual by a circle with a radius of  $r_0/2$  which we shall call the cross-section of the individual. The cross-section is chosen such that any overlap between two cross-sections can be taken as a significant contact between the two

respective individuals. Temporally, the cross-sections have to overlap for a time  $t_0$  which is the threshold interaction time that is assumed critical for an individual to infect another by proximity. For the sake of simplicity and with some loss of generality of our argument, we can assume that the probability of getting infected,  $p_t$ , is independent of the degree of overlap of the cross-sections<sup>4</sup> and for any time  $t > t_0$  as is done normally in automated contact tracing.

Figure 1 gives a depiction of what automated contact tracing would be for a group of individuals. In the left-most panel, B and C are in contact with A at  $t=0$  but not with each other. D is isolated from all of them. After a period of time  $t < t_0$ , B is isolated but C stays in contact with A. Then at time  $t = t_0 + \epsilon$ , where  $\epsilon \ll t_0$ , we see that C is still in contact with A, B remains isolated and E has come in contact with both A and C. Using the methods of automated contact tracing, if A reports as being tested as infected within 14 days of the encounter with C, C will be deemed as having had significant contact with A. E might also be deemed as such depending on how long he maintains proximity with A, but the proximity of E with C need not be counted even if E spends  $t > t_0$  in contact with C (if only primary contacts are traced) unless C reports as being infected too.

This method of automated contact tracing will work as long as A and C (and possibly E) are enrolled in the service even if B and D are not. However, D is completely isolated and by remaining so for a long time is observing social distancing from any other individual. B is representative of an individual who observes partial social distancing. Hence, for D this service is not necessary and for B it is of limited value. If C is not enrolled in the service C will never get notified if A gets tested as infected. C might get confirmed as infected or become an asymptomatic carrier and continue contaminating others. If A does not enrol in this service then C never gets notified leading to the same conclusions but E might get notified if C declares being infected and E is enrolled in the service.

An estimated 45% of person–person virus transmissions occur from individuals who are in the presymptomatic phase [34]. Prevalence of subclinical infections of SARS-CoV-2 further reduces the effectiveness of contact tracing. With automated contact tracing using a definition of  $r_0 = 2$  m and  $t_0 = 15$  min more than 80% of the cases can be traced [33] if every infected case is reported. In what follows, we create a simplified model of automated contact tracing to deduce the minimum fraction of the population that needs to enrol in the programme for it to be effective.

- Let  $N$  be the number of individuals in a population and  $f_i$  the fraction of the population that is infected, regardless of whether they know it or not. Therefore, the true number of infected individuals is  $f_i N$ .
- If testing is conducted only when mild or severe symptoms are seen (i.e. excluding testing of asymptomatic cases), the number of confirmed cases is  $r_c f_i N$  with  $r_c$  being the fraction of the infected that will be confirmed as infected by testing.
- We define  $f_e$  as the fraction of the population that is enrolled for automated contact tracing and  $f_c$  as the fraction of the users that will confirm that they have been diagnosed positive. Hence, the number of individuals that have tested positive, are using automated contact tracing and will confirm that they have been tested as infected is  $f_c f_e r_c f_i N$ .
- We define  $a_c$  as the average number of individuals that a contagious individual has significant contact with over the period in which they are contagious, significant contact being defined as lasting for a period of time greater than  $t_0$  and within a radius of  $r_0$ . The period over which an individual is contagious is about 5 days on an average for those who spread the disease in the presymptomatic phase and can be longer for asymptomatic and sub-clinical cases [41].

Since only  $f_e$  fraction of contacts are using the service, we can estimate the number of individuals that can be traced as  $f_c f_e r_c f_i N a_c f_e$ . Note that we assumed  $f_e$  and  $f_c$  are uniform even though  $r_c f_i N$  is not a random sample of the overall population with the purpose of estimating the number in the most conservative scenario. In the real world,  $f_e$  will be likely lower among the set of individuals that actually get infected (and their immediate contacts) and higher in the conjugate set due to different levels of caution exercised by the two groups, which, in turn, results in the decrease of the number of traceable contacts.<sup>5</sup>

To compute the number of individuals that need to be quarantined or isolated since they are now at risk of being infected from coming in contact with a contagious person, we define the following.

- Since  $p_t$  is defined as the probability of transmission of infection within the proximity radius  $r_0$  being exposed for a time greater than  $t_0$ , the number of individuals who are potentially newly infected is, on average,  $p_t f_i N a_c$ , i.e.  $p_t$  multiplied by the number of contacts of the group of infected individuals.<sup>6</sup>
- Finally, we define  $f_T$  as the fraction of the individuals at risk of being infected that needs to be successfully quarantined to quell the spread of the pathogen. In addition to other factors,  $f_T$  also depends on the delay in isolating potentially infected individuals [34].

Therefore, the number of individuals that should be quarantined is  $f_T p_t f_i N a_c$ . For automated contact tracing to work effectively, we have,

$$f_e^2 f_c r_c f_i N a_c \geq f_T p_t f_i N a_c. \quad (3.1)$$

## 4. The game of big numbers

Equation (3.1) simply states that the number of individuals that can be notified by automated contact tracing (on the left-hand side) has to be greater than or equal to the number of individuals who need to be notified (on the right-hand side). Note

that  $a_c$ , the average number of contacts, drops out of the inequality and hence, the inequality is independent of the population density of the region since equation (3.1) is in terms of fraction of the population and not the absolute number of individuals. This simply implies that in a region of denser population a larger number of people need to be contacted and quarantined but leaves  $f_e$  independent of the population density. Since the right-hand side is the minimum fraction of the population that needs to be traced we arrive at

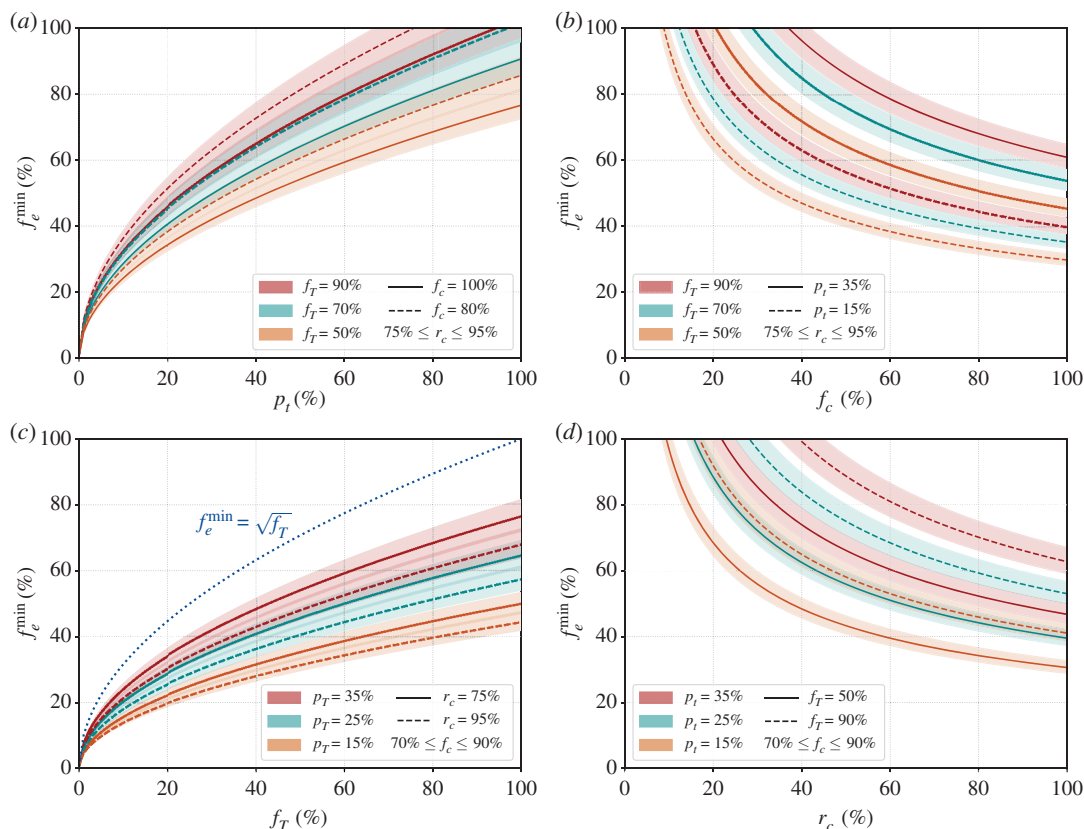
$$f_e^{\min} = \sqrt{\frac{f_T p_t}{f_c r_c}}. \quad (4.1)$$

The fraction  $f_e^{\min}$  is the minimum fraction of the population that needs to be enrolled in automated contact tracing for it to be effective as a means of slowing down the spread of the pandemic. In equation (4.1),  $p_t$  depends on the spreading dynamics of the pathogen determined by individual-to-individual interactions and, therefore, also depends on the mitigating measures taken at both the population level and the individual level. Naively, in automated contact tracing,  $p_t$  is taken as one if the contact has lasted for over time  $t_0$  with the subjects being less than  $r_0$  apart. This can be reduced by use of PPE or other mitigation methods as we discuss later. The parameter  $f_T$  depends on the disease spreading dynamics and can be estimated from modelling the disease spreading among a population [34]. From both Hellewell *et al.* [31] and Ferretti *et al.* [34], it is seen that 60–80% of the contacts need to be successfully traced and quarantined instantly to contain the outbreak over a period of time which makes  $f_T \sim 0.6$ – $0.8$ . The slower the response to the identification of contact at risk higher is  $f_T$  for the same reduction rate of the reproduction number. We assume that identification of contact at risk takes less than a day in automated contact tracing. The parameter  $r_c$  is governed by the ability to identify infected individuals through testing and depends on the protocols of the testing programme and its coverage. On the other hand,  $f_c$  is determined solely by the degree to which individuals are willing and able to confirm that they have been tested positive.

Let us examine the limit  $p_t = f_c = r_c = 1$ . This is the limit where every significant contact is assumed to be at risk, everyone who is enrolled in the automated contact tracing programme reports as infected when tested positive and every infected individual can be successfully identified by testing. Then we arrive at the relation  $f_e^{\min} = \sqrt{f_T}$  (blue dotted line in the third from left panel of figure 2). Since  $f_T$  is the fraction of contacts that need to be successfully isolated, it can be extracted from the abscissa of fig. 3 of ref. [34]. For example, if 100% of the infected cases can be isolated, then for a change in the epidemic growth rate by  $-0.1$ , one needs  $f_T \sim 60\%$ . Hence  $f_e^{\min} \sim 77\%$ . It is intuitive that  $f_e^{\min}$  scales as the square root of  $f_T$  since both the infected and the contact at risk need to be enrolled and the probability that each are enrolled is  $f_e$  leading to  $f_T \propto f_e^2$ . It gives the threshold which  $f_e^{\min}$  cannot exceed for any given  $f_T$ .

Lastly, we define the effectiveness of the automated contact tracing,  $\eta$ , as the ratio of the actual number of individuals that will be notified ( $f_e^2 f_c r_c f_i N a_c$ ) to the minimum number of individuals that should be notified to quell the spread of the disease ( $f_e^{\min 2} f_c r_c f_i N a_c$ ) and get

$$\eta \equiv \frac{f_e^2}{f_e^{\min 2}}. \quad (4.2)$$



**Figure 2.** Percentage of the population that needs to be enrolled ( $f_e^{\min}$ ) for automated contact tracing to be successful. Starting from the left, the solid and dashed lines represent  $f_c = 100\%$ ,  $80\%$ , respectively, for the first panel,  $p_t = 35\%$ ,  $15\%$  for the second panel,  $r_c = 75\%$ ,  $95\%$  for the third panel and  $f_T = 50\%$ ,  $90\%$  for the fourth panel. For the left two panels, the fraction of truly infected individuals that will be confirmed as infected by testing,  $r_c$  is varied between  $75\%$  and  $95\%$ . For the right two panels, the fraction of people who will confirm they have been tested as infected if they are enrolled,  $f_c$  is varied between  $70\%$  and  $90\%$ . Three cases for the minimum fraction of the individuals at risk that need to be traced are considered with  $f_T = 50\%$ ,  $70\%$ ,  $90\%$  in orange, green and red, respectively, in the left two panels and similarly, three cases are considered for  $p_t = 15\%$ ,  $25\%$ ,  $35\%$  in the right two panels. The blue dotted line in the third panel from the left gives the threshold variation of  $f_e^{\min}$  with  $f_T$  when all other parameters are set to 1. The y-axes are identical for all panels. See text for more details.

Figure 2 depicts how the fraction of the population that needs to be enrolled for the automated contact tracing programme to be successful ( $f_e^{\min}$ ) varies with the four factorized parameters. In the left-most panel of figure 2, we show the minimum percentage of the population that needs to be enrolled in automated contact tracing  $f_e^{\min}$  (in %) versus the transmission probability  $p_t$ . We consider two values for  $f_c = 80\%$ ,  $100\%$ , the fraction of individuals who test positive and will confirm their symptoms to trigger automated contact tracing, by the solid and dashed lines, respectively. The solid and dashed lines represent  $p_t = 15\%$ ,  $25\%$ , respectively. The bands are generated by varying the fraction of infected individuals that can be confirmed as infected by testing,  $r_c$ , between  $75\%$  and  $90\%$ . The other panels show the variation of  $f_e^{\min}$  with  $f_c$ ,  $f_T$  and  $r_c$ .

If we take a closer look at equation (3.1) and the left-most panel of figure 2 we see that even with a modest probability of transmission  $p_t$  (e.g. about  $30\%$ ) quite a large fraction of the population (about  $40\text{--}60\%$ ) needs to be enrolled in automated contact tracing even when we assume almost all of them will be actively participating in confirming when they get infected. Assuming all the traced contacts within radius  $r_0$  lasting for more than  $t_0$  period of time are going to be infected is equivalent to stating  $p_t = 100\%$ . From the panel on the right, we can see how a fall in the fraction of individuals that confirm that they are infected,  $f_c$ , can increase  $f_e^{\min}$ . Even with quite low values of  $p_t$  nearly half the population needs to be enrolled in automated contact tracing.

Let us try to understand why the effectiveness of automated contact tracing seems to drop so drastically with the enrolment fraction  $f_e$ . From the left-hand side of equation (3.1), we see that the effectiveness of automated contact tracing drops as  $f_e^2$ . We see that  $\eta$  drops to  $64\%$  when  $f_e = 0.8f_e^{\min}$  and  $25\%$  when  $f_e = 0.5f_e^{\min}$ . This nonlinearity exists because  $f_e$  not only reduces the number of infected individuals who can report their status but also the number of individuals who can receive a notification that they have come in contact with an infected person. The primary reason behind this is the fact that the automated contact tracing depends on voluntary participation whereas manual contact tracing or the use of CCTV, credit card information or identity logging at visited location to trace contact are not voluntary in their current form of implementation.<sup>7</sup>

Furthermore, as seen in figure 2, when the percentage of infected individuals who report that they have been tested as infected,  $f_c$  is lower than  $100\%$ , automated contact tracing becomes even less effective. In addition, the percentage of cases that can actually be detected,  $r_c$ , will realistically be less than  $100\%$  for SARS-CoV-2 because of the prevalence of subclinical cases that will escape detection and other clinical factors.

## 5. Assisted contact tracing

The necessary scale of implementation of automated contact tracing appears to be too large for it to be considered an

effective measure to slow down the ongoing pandemic. For automated contact tracing to be a viable option,  $f_e^{\min}$  has to be as low as possible. To achieve this either the product  $f_T p_t$  needs to be decreased or the product  $f_c r_c$  needs to be increased as seen from equation (4.1).

- Both  $f_T$  and  $p_t$  depend on the dynamics of the disease spreading among humans. The fraction of traced cases that need to be quarantined to stop the spread of the disease  $f_T$  can be reduced by extensive monitoring of the disease to make sure infected cases are isolated as soon as possible and their contacts are traced. Even a day or two of delays can increase  $f_T$  making automated contact tracing ineffective [34].
- Variations in  $p_t$  can be caused by several factors some of which are controllable. Since  $p_t$  depends on the contagiousness of the disease and any protective measures taken against the spread of the infection,  $p_t$  can be reduced by measures of limited social distancing, the use of PPE and raising public awareness about the contagiousness of COVID-19. This can pose a significant challenge in densely populated regions and regions with poor living conditions and might lead to the breakdown of the applicability of automated contact tracing.
- $f_c$  is somewhat more difficult to control assuming the reporting of those who are confirmed as infected is voluntary. This can only be increased by increasing the population's willingness to contribute to automated contact tracing.
- $r_c$  is the parameter that is least under control since without very large-scale testing, asymptomatic and mildly symptomatic cases will be difficult to find. This is especially true if the infection can spread by means other than proximity alone as might be the case for SARS-CoV-2 [15,18,19].

Thus we see that a combination of several measures along with a large participation of the population in contact tracing would be the optimal solution for avoiding extensive population-wide social distancing measures and reducing the cost to the economy and well-being of a nation and also allow for greater freedom of movement during a pandemic.

## 6. Discussion

In our analysis, we have inclined towards an optimistic picture of the spread of SAR-CoV-2. We have considered only spreading due to proximity and not considered other means of spreading like contaminated surfaces and aerosol that are common for SARS-CoV-2 [15,18,19] and can increase  $p_t$ . In figure 2, we have taken a minimum  $r_c$  of 75% when this can be even lower if widespread testing is not conducted to identify subclinical cases that can go undetected. We have also neglected the requirement for tracing secondary or tertiary contacts. In addition, we have also ignored events where a large number of individuals are infected in very a crowded location like public events for which thresholds like  $r_0$  and  $t_0$  need to be modified. Despite this optimistic picture, our analysis shows that a majority of the population has to enrol and actively participate in automated contact tracing for the measure to work in the absence of active social distancing measures.

We have not addressed the sociological aspect of selection bias in the enrolment process. Diversity in socio-economic conditions, awareness of technology and willingness to participate in a community effort will create variation in representations

among the population. This can lead to the most vulnerable in society getting the least benefit from the implementation of automated contact tracing. Addressing the challenges of implementing automated contact tracing in developing nations where the necessary technologies might not be accessible to a large fraction of the population lies beyond the scope of this work.

We have shown that in real-world scenarios, automated contact tracing alone cannot contain a pandemic driven by a pathogen like the SARS-CoV-2. Advocating it as such can lead to exasperating the spread of the pathogen. The primary reasons why such a strategy will not work as effectively as projected for SARS-CoV-2 is because of a large degree of spreading from pre-symptomatic and subclinical hosts, and the rapidity with which the virus spreads through proximity alone if no additional measures are taken to mitigate the spread. All of these conjugated with the vulnerability of automated contact tracing to insufficient sampling due to limited participation among the population and possibly incomplete reporting of infected cases will lead to reduction in the efficacy of automated contact tracing. A small fraction of the population being infected with SARS-CoV-2 can quickly lead to a majority of the population being needed to participate in the programme.

We put together all the factors of concern and show that they follow a simple relationship. We further discussed how factors like the transmission probability  $p_t$  should be reduced and the fraction of infected individuals that test positive,  $r_c$ , should be increased to assist in reducing the burden on automated contact tracing while keeping the entire process voluntary. The strength of our model lies in the fact that we separate the various parameters that individually contribute to the efficacy of automated contact tracing. This allows for each parameter to be addressed individually through improved clinical intervention, logistics, mitigation strategies and public awareness of automated contact tracing to increase adoption of the method. While our focus in this paper is to address the feasibility of automated contact tracing for containing the spread of SARS-CoV-2, equation (4.1) can be applied for using automated contact tracing to contain other pathogens too. Our analysis is also independent of the methods of implementation of automated contact tracing and the definitions of  $r_0$  and  $t_0$ . Therefore, our approach is quite general.

During the final stages of this work, a similar result was reached by the authors of [42] using a branching process model and arguments from statistical mechanics. They reached a similar conclusion as we do in our paper showing that nearly 75% to 95% of the population need to participate in automated contact tracing for it to be effective. The results in their work corresponds to ours when  $p_t = f_c = r_c = 1$  or  $f_c = \sqrt{f_T}$ . A more informed approach based to contact tracing has also been suggested which leads to a lower fraction of the population needing to be enrolled based on a probabilistic model disease spread [43].

The trust in contact tracing stems from the effectiveness with which it was used to contain pathogens like Ebola, SARS-CoV and MERS-CoV. However, the dynamics of the spread of SARS-CoV-2 is very different from these pathogens. Hence, the effectiveness of contact tracing in stopping the spread of these pathogens should not be seen as a validation of the effectiveness of automated contact tracing for SARS-CoV-2. To make automated contact tracing work, a majority of the population has to enrol for this service and actively participate in it. If this cannot be established then other measures of mitigating the spread of SARS-CoV-2 should be implemented in addition.

As can be seen by the success of several nations in containing the spread of COVID-19, only a judicious combination of contact tracing with measures such as partial social distancing, wide use of PPE and dissemination of information about the disease can prove to be effective in slowing down the spread of the ongoing pandemic.

**Data accessibility.** All data necessary for this work are provided in the paper. The opinions/conclusions presented in this publication are those of the author(s) and do not necessarily reflect the views of John Templeton Foundation.

**Authors' contributions.** H.K. and A.P. contributed equally to this work.

**Competing interests.** We declare we have no competing interests.

**Funding.** This work was partially supported by a grant from John Templeton Foundation.

**Acknowledgements.** We thank Paul Davies, Christophe Grojean, Luca Silvestrini and Melinda Varga for comments and suggestion.

## Endnotes

<sup>1</sup>One study suggested that MERS-CoV can be transmitted before the onset of symptoms [12].

<sup>2</sup>Much higher reproductive rates have also been estimated with data from Wuhan, China [28]. In general, there are variabilities in the estimation of the reproduction number with time and containment strategies [29].

<sup>3</sup>cf. <https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/contact-tracing.html>

<sup>4</sup>This assumption is to mimic how automated contact tracing is implemented through mobile devices where the probability of infection is not considered as variation depending on distance between two users as long as it is closer than a physical proximity threshold set by the contact tracing. In addition,  $p_t$  as a function of distance is not very well known as yet.

<sup>5</sup>The variations in  $f_e$  within different demographic groups are not accounted for in our work and this can potentially be correlated with the way the disease spreads. Here,  $f_e$  is the fraction of the whole population that continuously use the service. Accounting for these variations within our model is possible but requires more data.

<sup>6</sup>Here, we make a simplifying assumption that the disease has spread to only a small fraction of the population and the probability of a single healthy individual to randomly have significant contact with two contagious individuals in a period of 14 days is negligibly small in general. There will be outliers depending on the habits of individuals but we can neglect them for this analysis.

<sup>7</sup>These effectively makes  $f_e$  close to 100% for both those who have been diagnosed as infected and their contacts.

## References

- Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. 2020 Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868. (doi:10.1126/science.abb5793)
- Swanson KC. 2018 Contact tracing performance during the Ebola epidemic in Liberia, 2014–2015. *PLoS Negl. Trop. Dis.* **12**, 1–14. (doi:10.1371/journal.pntd.0006762)
- Berge T, Ouemba Tassé AJ, Tenkam HM, Lubuma J. 2018 Mathematical modeling of contact tracing as a control strategy of Ebola virus disease. *Int. J. Biomath.* **11**, 1850093. (doi:10.1142/S1793524518500936)
- Browne C, Gulbudak H, Webb G. 2015 Modeling contact tracing in outbreaks with application to Ebola. *J. Theor. Biol.* **384**, 33–49. (doi:10.1016/j.jtbi.2015.08.004)
- Shahmoradian NM, Ferdousi T, Scoglio C, Sahneh FD. 2018 Quantifying the impact of early-stage contact tracing on controlling Ebola diffusion. *Math. Biosci. Eng.* **15**, 1165–1180. (doi:10.3934/mbe.2018053)
- Kwok KO, Tang A, Wei VW, Park WH, Yeoh EK, Riley S. 2019 Epidemic models of contact tracing: systematic review of transmission studies of severe acute respiratory syndrome and middle east respiratory syndrome. *Comput. Struct. Biotechnol. J.* **17**, 186–194. (doi:10.1016/j.csbj.2019.01.003)
- Okolie A, Müller J. 2020 Exact and approximate formulas for contact tracing on random trees. *Math. Biosci.* **321**, 108320. (doi:10.1016/j.mbs.2020.108320)
- Klinkenberg D, Fraser C, Heesterbeek H. 2006 The effectiveness of contact tracing in emerging epidemics. *PLoS ONE* **1**, 1–7. (doi:10.1371/journal.pone.0000012)
- Corman VM *et al.* 2015 Viral shedding and antibody response in 37 patients with middle east respiratory syndrome coronavirus infection. *Clin. Infect. Dis.* **62**, 477–483. (doi:10.1093/cid/civ951)
- Chowell G, Abdirizak F, Lee S, Lee J, Jung E, Nishiura H, Viboud C. 2015 Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study. *BMC Med.* **13**, 210. (doi:10.1186/s12916-015-0450-0)
- Fraser C, Riley S, Anderson RM, Ferguson NM. 2004 Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151. (doi:10.1073/pnas.0307506101)
- Omrani AS, Matin MA, Haddad Q, Al-Nakhli D, Memish ZA, Albarak AM. 2013 A family cluster of Middle East respiratory syndrome coronavirus infections related to a likely unrecognized asymptomatic or mild case. *Int. J. Infect. Dis.* **17**, e668–e672. (doi:10.1016/j.ijid.2013.07.001)
- Rewar S, Mirdha D. 2014 Transmission of Ebola virus disease: an overview. *Ann. Global Health* **80**, 444–451. (doi:10.1016/j.aogh.2015.02.005)
- Lau LLH *et al.* 2010 Viral shedding and clinical illness in naturally acquired influenza virus infections. *J. Infect. Dis.* **201**, 1509–1516. (doi:10.1086/652241)
- Santaripa JL *et al.* 2020 Aerosol and surface contamination of SARS-CoV-2 observed in quarantine and isolation care. *Sci. Rep.* **10**, 12732. (doi:10.1038/s41598-020-69286-3)
- Wang C, Horby PW, Hayden FG, Gao GF. 2020 A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473. (doi:10.1016/S0140-6736(20)30185-9)
- He X *et al.* 2020 Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675. (doi:10.1038/s41591-020-0869-5)
- van Doremalen N *et al.* 2020 Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **382**, 1564–1567. (doi:10.1056/NEJM2004973)
- Guo Z-D *et al.* 2020 Aerosol and surface distribution of severe acute respiratory syndrome coronavirus 2 in hospital wards, Wuhan, China, 2020. *Emerg. Infect. Dis.* **26**, 1586–1591. (doi:10.3201/eid2607.200885)
- Zhou P *et al.* 2020 A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273. (doi:10.1038/s41586-020-2012-7)
- Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. 2020 Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263. (doi:10.1126/science.abb2507)
- Chen Y, Guo Y, Pan Y, Zhao ZJ. 2020 Structure analysis of the receptor binding of 2019-nCoV. *Biochem. Biophys. Res. Commun.* **525**, 135–140. (doi:10.1016/j.bbrc.2020.02.071)
- Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. 2020 Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493. (doi:10.1126/science.abb3221)
- Li Q *et al.* 2020 Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207. (doi:10.1056/NEJMoa2001316)
- Wu JT, Leung K, Leung GM. 2020 Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697. (doi:10.1016/S0140-6736(20)30260-9)

26. Riou J, Althaus CL. 2020 Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* **25**, 2000058. (doi:10.2807/1560-7917.ES.2020.25.4.2000058)
27. Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ, Meyers LA. 2020 Risk for transportation of 2019 novel coronavirus disease from Wuhan to other cities in China. *Emerg. Infect. Dis. J.* **26**, 1049–1052. (doi:10.3201/eid2605.200146)
28. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. 2020 High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis. J.* **26**, 1470–1477. (doi:10.3201/eid2607.200282)
29. Al-Raei M. 2020 The basic reproduction number of the new coronavirus pandemic with mortality for India, the Syrian Arab Republic, the United States, Yemen, China, France, Nigeria and Russia with different rate of cases. *Clin. Epidemiol. Global Health* **9**, 147–149. (doi:10.1016/j.cegh.2020.08.005)
30. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
31. Hellewell J *et al.* 2020 Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* **8**, e488–e496. (doi:10.1016/S2214-109X(20)30074-7)
32. Endo A, Abbott S, Kucharski A, Funk S. Submitted. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside china. *Wellcome Open Res.* **5**. (doi:10.12688/wellcomeopenres.15842.1)
33. Keeling MJ, Hollingsworth TD, Read JM. 2020 Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *J. Epidemiol. Community Health* **74**, 861–866. (doi:10.1136/jech-2020-214051)
34. Ferretti L *et al.* 2020 Quantifying SARS-COV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936. (doi:10.1126/science.abb6936)
35. Pung R *et al.* 2020 Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *Lancet* **395**, 1039–1046. (doi:10.1016/S0140-6736(20)30528-6)
36. Chen C-M *et al.* 2020 Containing COVID-19 among 627,386 persons in contact with the Diamond Princess cruise ship passengers who disembarked in Taiwan: big data analytics. *J. Med. Internet Res.* **22**, e19540. (doi:10.2196/19540)
37. Park S, Choi GJ, Ko H. 2020 Information technology-based tracing strategy in response to COVID-19 in South Korea—privacy controversies. *JAMA* **323**, 2129–2130. (doi:10.1001/jama.2020.6602)
38. Day M. 2020 Covid-19: four fifths of cases are asymptomatic, China figures indicate. *BMJ* **369**, m1375. (doi:10.1136/bmj.m1375)
39. Verity R *et al.* 2020 Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* **20**, 669–677. (doi:10.1016/S1473-3099(20)30243-7)
40. Morawska L *et al.* 2020 How can airborne transmission of COVID-19 indoors be minimised? *Environ. Int.* **142**, 105832. (doi:10.1016/j.envint.2020.105832)
41. Ferretti L *et al.* 2020 The timing of COVID-19 transmission. *medRxiv*. (doi:10.1101/2020.09.04.20188516)
42. Bulchandani VB, Shivam S, Moudgalya S, Sondhi SL. 2020 Digital herd immunity and COVID-19. (<http://arxiv.org/abs/2004.07237>).
43. Guttal V, Krishna S, Siddharthan R. 2020 Risk assessment via layered mobile contact tracing for epidemiological intervention. *medRxiv*. (doi:10.1101/2020.04.26.20080648)