Fundamental limitations of contact tracing for COVID-19

Paul Tuppera, Sarah P. Ottob, and Caroline Colijn*

aDepartment of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; bDepartment of Zoology, University of British Columbia, Vancouver, BC V6T1Z4, Canada

*ccolijn@sfu.ca

Abstract

Contact tracing has played a central role in COVID-19 control in many jurisdictions and is often used in conjunction with other measures such as travel restrictions and social distancing mandates. Contact tracing is made ineffective, however, by delays in testing, calling, and isolating. Even if delays are minimized, contact tracing triggered by testing of symptomatic individuals can only prevent a fraction of onward transmissions from contacts. Without other measures in place, contact tracing alone is insufficient to prevent exponential growth in the number of cases in a population with little immunity. Even when used effectively with other measures, occasional bursts in call loads can overwhelm contact tracing systems and lead to a loss of control. We propose embracing approaches to COVID-19 contact tracing that broadly test individuals without symptoms, in whatever way is economically feasible—either with fast and cheap tests that can be deployed widely, with pooled testing, or with screening of judiciously chosen groups of high-risk individuals. These considerations are important both in regions where widespread vaccination has been deployed and in those where few residents have been immunized.

Key words: contact tracing, COVID-19, serial interval, nonpharmaceutical interventions

Introduction

The effectiveness of contact tracing for any infectious disease is limited by how quickly contacts can be informed compared with the infectious period. If contact tracing teams reach an individual’s contacts only toward the end of their infectious period, very few further infections will be prevented. Several delays in the process make rapid contact tracing challenging: the time to develop symptoms, to seek a test, to get test results, and for contact tracing teams to reach contacts. Contact tracing is particularly challenging for COVID-19 because transmission often occurs before symptoms appear (He et al. 2020), and some individuals who never develop symptoms can transmit the virus (Byambasuren et al. 2020).

Before the widespread deployment of vaccination, symptomatic testing followed by contact tracing, alongside widespread distancing measures, has been central to efforts to control COVID-19 in many jurisdictions in Europe, the United Kingdom, and North America. However, broad and restrictive distancing measures have been considered too costly to be palatable in the long term, both economically and in terms of unintended consequences for public health, mental health, and inequality (Bonaccorsi et al. 2020; Donohue and Miller 2020; Pfefferbaum and North 2020). This left most of North America, Europe, and the United Kingdom, among others, in the difficult position of repeatedly reopening their
economies following periods of declining COVID-19 numbers. This reopening has occurred in a context where COVID-19 immunity was very low, substantial costs had been incurred, but COVID-19 had not been eliminated and (or) was continually reintroduced. In many areas, contact tracing capacity was dramatically increased to allow reopening.

After reopening and before vaccination, almost all jurisdictions in North America and Europe saw substantial resurgence of COVID-19. Some overwhelmed their health care systems, for example exceeding intensive care unit capacity, cancelling elective surgeries, diverting patients, and being depleted of nursing staff, leading in some cases to health care workers being asked to work while testing positive for COVID-19 (Reimann 2020; Kottasova et al. 2020; Romero 2020). Groups of doctors have written open letters calling for wider shutdowns while politicians hesitate, knowing the costs and the unintended damages that these shutdowns will create (Bernhardt 2020; Greenhalgh et al. 2020; Wellington 2020). This has occurred despite symptomatic testing and contact tracing being in place. The level of widespread distancing that is tolerable and sustainable is insufficient for robust COVID-19 control with the testing and contact tracing systems that have been in place.

For COVID-19, the average number of new cases per infected individual, known as the reproductive number $R$, depends on the actions taken by a society to control the pandemic. In the absence of social restrictions, contact tracing, and immunity, the basic reproductive number, $R_0$, is estimated to be between 2 and 5 (Liu et al. 2020) and even higher for variants of concern (CDC 2021) (see Table 1 for a summary of notation). Social distancing and other nonpharmaceutical interventions (NPIs) without the use of contact tracing reduce $R$ to some lower value $R_{NCT}$ (NCT = no contact tracing). Contact tracing reduces this further to $R = (1 - \rho)R_{NCT}$, where $\rho$ is the average fraction of cases a contact would infect that are prevented by contact tracing. This fraction depends on the capacity to contact trace and on the timing of symptoms and transmission. Two important features are the fraction of contacts that can be reached and who self-isolate (we call this fraction $\alpha$ or “coverage”) and the time from symptom onset until the contacts are reached (we call this time $\tau$). Coverage $\alpha$ generally falls below 100% because some contacts are unknown or unidentifiable, some are not disclosed by the index for whatever reason, and some refuse to isolate even when contacted. The factor $\alpha$ also accounts for the fraction of index cases who are never detected at all, perhaps because they are asymptomatic, and therefore have none of their contacts informed or because they did not seek or access testing. The delay $\tau$ includes the time from symptom onset to seeking a test, from testing until the results are available, and then the time it takes for a contact to be reached and isolate. When cases are rare, contact tracers can rapidly reach a large proportion of a case’s contacts. As the number of cases rises, however, contact tracing capacity can become increasingly taxed, reducing

<table>
<thead>
<tr>
<th>Table 1. Different reproductive numbers used for different (hypothetical) conditions for COVID-19 in a given jurisdiction.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
</tr>
<tr>
<td>$R_0$</td>
</tr>
<tr>
<td>$R_{NCT}$</td>
</tr>
<tr>
<td>$R$</td>
</tr>
<tr>
<td>$\alpha$</td>
</tr>
<tr>
<td>$\tau$</td>
</tr>
<tr>
<td>$\rho$</td>
</tr>
</tbody>
</table>

*With the term nonpharmaceutical intervention (NPI), we include interventions like distancing and mask mandates but do not include contact tracing.
the ability to reach all contacts (lowering \(\alpha\)) and lengthening the delay (raising \(\tau\)). In Fig. 1 we illustrate the fraction of cases that contact tracing can prevent given the tracing capacity, i.e., how \(\rho\) depends on \(\alpha\) and \(\tau\).

If we already have contact tracing in place, we cannot observe \(R^{NCT}\). But if we have estimates of \(R\), the coverage, and the delay, we can first estimate \(\rho\) and then estimate \(R^{NCT}\) from \(R^{NCT} = R/(1 - \rho)\). Then we can determine the fraction of new cases (\(\rho_{crit}\)) we would have to prevent to bring \(R\) below 1, by solving \(1 = (1 - \rho_{crit})R^{NCT}\). We can use \(\rho_{crit}\) in turn, to determine whether contact tracing can bring cases under control and how much the delay would have to be reduced, and (or) the coverage increased, to accomplish this. See the Appendix for the derivation of our formula for \(\rho\) in terms of \(\tau\) and \(\alpha\).

As an example, in British Columbia, on 1 April 2021, \(R\) was estimated to be approximately 1.2 (BCCDC 2021). In ideal cases \(\tau\) (time from symptom onset to contact notification) was approximately 5 days (Daflos 2020) (roughly 1 day for symptom onset to test, 2 days for the test result to be available, and another 2 days for the contacts to be notified, though this varies from one to many days). With coverage of \(\alpha = 0.5\), contact tracing only prevents 8% of onward cases (\(\rho = 0.08\), using the formula in the caption of Fig. 1), suggesting that without contact tracing \(R^{NCT} = 1.3\). To reduce \(R\) further down to 1 we would need \(\rho = 1 - (1/1.3) = 0.23\). This could be achieved by reducing delays in contacting from 5 to 3 days, potentially by combining traditional contact tracing with contact tracing apps and automatic text messaging, as achieved in Taiwan (Jian et al. 2020). Alternatively, we could increase coverage, but \(R\) is only reduced to 1.08 even if we reached 100% of contacts (\(\alpha = 1\)), which would not bring spread under control. In reality, striving to both decrease delays (\(\tau\)) and increase coverage (\(\alpha\)) would be best.

The fact that there is a critical value of the contact tracing delay beyond which contact tracing is not able to prevent a sufficient fraction of cases to bring COVID-19 under control is a fundamental limitation of contact tracing that comes about because many transmissions occur before and soon after symptom onset before cases are detected and contacts are notified. As a consequence, only if NPI causes \(R\) to be low enough (\(R\) below the curves in Fig. 1, bottom right panel) will contact tracing effectively control an epidemic. In such cases, there is a tipping point: a value of a parameter where a system has qualitatively different behaviour when the parameter is above or below it (Lamberson and Page 2012). If the delay \(\tau\) is above its critical value, we have exponential growth in the number of cases, otherwise cases decline. There is another tipping point for the coverage; increasing coverage (for example by expanding the definition of a contact to include more people, taking extra measures to ensure compliance with self-isolation) could push \(R\) below 1 if \(\tau\) is sufficiently short and other measures are in place.

Without any mitigation measures or immunity, contact tracing on its own is unable to bring COVID-19 under control, leading to exponential growth in cases. Even with the optimistic parameters of \(\alpha = 0.5\) and \(\tau = 3\) days we get \(\rho = 0.23\), which would only bring \(R\) below 1 if \(R_0 < 1.3\), far below the estimated range of 2–5 for the “wild type” (Liu et al. 2020) and even higher for variants of concern (CDC 2021).

Even if exponential growth is slow at first, the situation will become worse as contact tracing reaches its capacity. As the number of incident cases increases, the contact tracing system will be put under a heavier load, with \(\tau\) increasing and \(\alpha\) decreasing, meaning an even lower fraction of cases are prevented by contact tracing. \(R\) increases even further, which in turn causes even faster growth in the number of incident cases. This positive feedback cycle unabated will overwhelm the contact tracing system, because contact tracing capacity cannot be expanded as quickly as case numbers rise.
Fig 1. Top: Contact tracing of an index case eliminates infections from a secondary case only after contact and self-isolation occur (blue dashed line). Bottom left: $\rho$ as determined by $\alpha$ and $\tau$. Bottom right: The maximum value of $R_{NCT}$ (reproductive number without contact tracing) for which contact tracing with a given $\alpha$ and $\tau$ can bring growth under control. Parameters and calculations: If a proportion $\alpha$ of contacts are reached and self-isolate, the fraction of tertiary infections averted (pink shading) is

$$\rho = \alpha \int_0^\infty f(t)[1 - F(t_{\text{sympt}} + \tau - t)]dt$$

where $f(t)$ is the distribution of the generation interval (red/pink curve; assumed to be Gamma distributed) and $F(t)$ is its cumulative distribution. $\rho$ is calculated by assuming that the secondary case started at a random time $t$ while the index case was infectious (integrating over $f(t)$) and then calculating what fraction of the infectious period of the secondary case is averted (the $1 - F$ term), accounting for the time delay until the index case develops symptoms ($t_{\text{sympt}}$ since infection) and then the additional delay until contacts are traced ($\tau$). See Appendix for more details.
If a region does have COVID-19 cases under control \((R < 1)\), we can also determine how large an increase in call load can be handled before contact tracing will fail to limit spread. As the number of individuals who need to be contacted daily \((n)\) rises above the tracing capacity \((c)\), the best strategy is to reduce coverage without also adding further delays. This is because the benefits of contact tracing decline proportionately with the number of people contacted but decline faster than linearly with delays (as illustrated in Fig. 1). This in turn is because at longer delays, most contacts will not be called until after their peak infectivity. Assuming that further delays are avoided (holding \(\tau\) constant), coverage will decline in proportion to how far over capacity the call demand is (e.g., coverage will drop by half if twice as many people need to be called as can be called, \(n/c = 2\)). If we take our equation for how \(R\) is determined from \(\rho\) and \(R_{NCT}\), and our considerations of how \(\rho\) is affected by excess call demand, we can determine the call volume that will cause \(R\) to rise above one. We find that contact tracing breaks down once \(n/c > \rho R_{NCT}/(R_{NCT} - 1)\) or, equivalently, \(n/c > (R_{NCT} - R)/(R_{NCT} - 1)\).

For example, in a region brought down to \(R = 0.9\) from \(R_{NCT} = 1.33\), contact tracing will fail to prevent spread once the demand requires 33% more calls than can be placed in a day.

This is the best-case scenario—only coverage was impacted. In reality, as the incidence or call volume rises, delays are likely to grow. Testing backs up so that fewer onward cases can be prevented. Thus, any event—a burst in transmission caused by a superspreading event, a cluster of importations, or changes in behaviour around a holiday—that causes the call volume to rise above this limit cannot be reversed by contact tracing alone. Importantly, the number of calls needed can rise either because of true increases in incidence or due to increases in contacts per case—either can cause this kind of collapse. When cases rise, the effectiveness of contact tracing declines just when it is most needed. Another important factor is how much social distancing is in place when contact tracing is performed. If much of the population is isolating at home and large indoor gatherings are prohibited, the number of contacts an infected individual has will be few, and contact tracers will have a relatively light load. But as restrictions are lifted and reopening occurs, each case will have many more contacts.

Where does this leave us? The answer depends on the strategy employed by a region to contain the pandemic. In areas with many active cases, contact tracing is used to reduce case numbers (“mitigation phase”), requiring an ongoing contact tracing system that can handle large case loads. In regions aiming to keep COVID-19 out, contact tracing is needed to prevent reintroductions from leading to new outbreaks (“containment phase”), which requires rapid deployment of contact tracing. In either case, the success of contact tracing will rely on strategies that reduce delays (e.g., automated contact tracing, focusing on cases with the most recent symptom onset dates) and that increase the proportion of contacts reached (e.g., asymptomatic testing, use of apps to monitor contacts).

One option for considerably strengthening the power of contact tracing is to go beyond merely instructing contacts to isolate by testing all contacts of a known case as rapidly as possible, whether or not they are symptomatic. This approach, which is used in New Zealand (NZMOH 2020), has the advantage that if a secondary contact tests positive, tracing for their contacts can be initiated much earlier than if testing only occurs after symptom onset, which might never happen if the secondary contact remains asymptomatic. In contrast, if we simply ask individuals to self-isolate, this isolation is imperfect, and we do not obtain information about their contacts early enough to prevent onward transmission from them.

Mass testing is another approach used, for example, in Slovakia, where two-thirds of the country were tested with rapid antigen tests over two days, and 57 500 COVID-19 cases were identified, which is almost three-quarters the number of cases discovered by polymerase chain reaction (PCR) tests in that country since the beginning of the pandemic (Shotter 2020). These tests are less accurate than standard PCR tests, but lower cost means that they can be deployed much more widely. Pooled
sample testing is another approach: samples are collected from a group or from environmental samples such as wastewater and tested at once, reducing the costs of testing (Cleary et al. 2021). A positive result can lead to instructions to isolate for the whole group and (or) to subsequent individual tests. In addition, when call load exceeds contact tracing capacity in a region, mass testing and isolation of positive cases could be used to bring case numbers down to where contact tracing would become effective again. What these strategies have in common is that they aim to find cases, and even contacts of those cases, before symptom onset, at the start of or before infectiousness (effectively reducing \( r \)). Measures initiated by testing symptomatic individuals cannot “get ahead of transmission” in the same way. Many jurisdictions that have successfully taken a containment strategy have relied on getting ahead of transmission in multiple ways (e.g., use of a tracer app combined with testing of asymptomatic contacts in New Zealand (NZMOH 2020), Taiwan (Jian et al. 2020), and Prince Edward Island (PEI 2020)).

In regions without high levels of immunity through infection or vaccination, we can either find ways to get ahead of transmission or accept that contact tracing focused on symptomatic cases must be complemented with long, sustained, and widespread distancing measures to control COVID-19. If we want to avoid continual shutdowns and resurgences, as well as the extremely high economic, social, and health costs of shutdown measures, we need to build robust testing strategies and capacity that can stop transmission much earlier before symptoms appear.

Now that vaccines are approved and are being widely deployed, and vaccinated populations are moving to reopen economic and social activities, testing and contact tracing will face new challenges. First, there will be more contacts when people are working in person, socializing indoors, and travelling more: were it not for immunity we would effectively be increasing the reproductive number \( R \) back up towards prepandemic levels. Vaccination will counteract this rise, preventing a large fraction of infections and providing good protection against symptomatic disease, but severe cases will remain a problem among the unvaccinated and even among some vaccinated individuals (Teran et al. 2021). Second, detecting infected contacts who are vaccinated will become more challenging, because vaccinated individuals appear to be asymptomatic more often (Teran et al. 2021) or may ignore the symptoms that they do have, assuming that they are immune. Identifying an individual’s infectors and infected contacts will thus become increasingly difficult. Third, although vaccine “breakthrough” cases may have lower viral loads (Levine-Tiefenbrun et al. 2021), which may reduce transmission, vaccinated individuals with no or mild symptoms will likely circulate for longer than they would if they knew they were possibly infectious. Longer periods of activity risk more onward transmissions but also provide a greater opportunity for contact tracing if cases are detected early. To succeed in largely vaccinated populations, contact tracing will likely require more focus on testing asymptomatic contacts.

As societies open up in the context of vaccination, continued contact tracing programs could lower the level of immunity required to keep COVID-19 cases in check using mitigation measures that can be maintained indefinitely. Our work shows, however, that the buffer provided by long-term tracing programs is small unless infected contacts are quickly reached and isolated. As jurisdictions become immunized and begin to open up, contact tracing must be adjusted to handle many more contacts per case and to adjust to shifting disease characteristics, including the real risk that more infections will fly below the radar unless accompanied by asymptomatic testing.

**Author contributions**

PT, SPO, and CC conceived and designed the study. PT, SPO, and CC analyzed and interpreted the data. PT, SPO, and CC contributed resources. PT, SPO, and CC drafted or revised the manuscript.
Competing interests
The authors have declared that no competing interests exist.

Data availability statement
All relevant data are within the paper.

References


Wellington M. 2020. Shut down, start over, do it right. US. PIRG [online]: Available from uspirg.org/resources/usp/shut-down-start-over-do-it-right.

Appendix

In our model the index case is infected at time 0. They infect other individuals (the secondary cases) at random subsequent times. These secondary cases infect further tertiary cases, also at random later times. The index case ideally changes their behaviour when they develop symptoms by isolating, and even more so after testing positive, thereby reducing the number of secondary cases, but contact tracing does not reduce the number of secondary cases itself. Contact tracing works by informing some of these secondary cases that they may be infectious so that they can isolate sooner than they would otherwise.

The time from when an individual is infected to when they infect another is known as the generation interval. We use estimates of the distribution of the generation interval from Ganyani et al. (2020); it is a Gamma
distributed random variable with a mean of 4 days and a standard deviation of 1.5 days. We denote the probability density function of the generation interval by $f$ and its cumulative distribution function by $F$. We assume that this distribution for the generation interval factors in the effects of people changing their behaviour (self-isolating, reducing contact, and preventing onward infections) due to having symptoms.

Let $t$ denote the time after the index case is infected that a secondary case is infected; $t_{\text{symp}}$ is the time from infection to symptom onset (we’ve assumed 5 days) and $\tau$ the time from symptom onset to the index case’s contacts being notified (and therefore when they can begin isolation). The secondary case can then begin isolation time $t_{\text{symp}} + \tau - t$ after they were infected. They will have already infected a portion $F(t_{\text{symp}} + \tau - t)$ of the people they were going to infect if they had not been notified. So the fraction of their infection that could be prevented if they completely isolated from that point onward would be $1 - F(t_{\text{symp}} + \tau - t)$. Of course, not all contacts will be reached, and not all contacts will isolate—which those that do will do so imperfectly. Furthermore, this may not occur altogether if the original case is asymptomatic, is symptomatic and doesn’t seek a test, or fails to provide contact information when tested. We combine these factors together into a single parameter $\alpha$, which represents how much the efficacy of contact tracing is reduced by these considerations. Together we obtain that, for a contact infected at time $t$, the total fraction of their onward transmissions that are prevented is $\alpha[1 - F(t_{\text{symp}} + \tau - t)]$. Averaging this with respect to $t$ the time of infection of the contact, which has density $f(t)$, we obtain the value for $\rho$ used in the main text.