Anatomy of digital contact tracing: role of age, transmission setting, adoption and case detection

Jesús A. Moreno López[1], Beatriz Arregui García[1], Piotr Bentkowski[1], Livio Bioglio[2], Francesco Pinotti[1], Pierre-Yves Boëlle[1], Alain Barrat[3,4], Vittoria Colizza[1], Chiara Poletto[1]*

[1]INSERM, Sorbonne Université, Pierre Louis Institute of Epidemiology and Public Health, Paris, France

[2]Department of Computer Science, University of Turin, Turin, Italy

[3]Aix Marseille Univ, Université de Toulon, CNRS, CPT, Turing Center for Living Systems, Marseille, France

[4]Tokyo Tech World Research Hub Initiative (WRHI), Tokyo Institute of Technology, Tokyo, Japan

*chiara.poletto@inserm.fr

Abstract

The efficacy of digital contact tracing against COVID-19 epidemic is debated: smartphone penetration is limited in many countries, non-uniform across age groups, with low coverage among elderly, the most vulnerable to SARS-CoV-2. We developed an agent-based model to precise the impact of digital contact tracing and household isolation on COVID-19 transmission. The model, calibrated on French population, integrates demographic, contact-survey and epidemiological information to describe the risk factors for exposure and transmission of COVID-19. We explored realistic levels of case detection, app adoption, population immunity and transmissibility. Assuming a reproductive ratio $R = 2.6$ and 50% detection of clinical cases, a ~20% app adoption reduces peak incidence of ~36%. With $R = 1.7$, >30% app adoption lowers the epidemic to manageable levels. Higher coverage among adults, playing a central role in COVID-19 transmission, yields an indirect benefit for elderly. These results may inform the inclusion of digital contact tracing within a COVID-19 response plan.

Introduction

In the absence of vaccination and effective drugs for COVID-19, preventing transmission remains key to mitigating the pandemic. To this aim, lockdown was adopted in many European countries and have successfully slowed down the epidemic to manageable levels[1], [2]. These interventions had however a huge economic and societal cost and were relaxed as incidence went down. Yet, population immunity remains low [3], [4] and re-emerging outbreaks are possible [5], [6]. Sustainable strategies are required to keep the epidemic under control while enabling a return to a close-to-normal functioning of the society. Widespread testing, case finding and isolation, contact-tracing, use of face masks and enhanced hygiene are believed to be crucial components of these strategies.

Contact-tracing aims at avoiding transmission by isolating at an early stage only those individuals who are infectious or potentially infectious, in order to minimize the societal costs associated to isolation. Considerable resources are therefore directed at improving surveillance capacities to allow efficient and rapid investigation and isolation of cases and their contacts. To enhance tracing capacities, the use of digital technologies has been proposed, leveraging the wide-spread use of smartphones. Therefore, proximity-sensing applications have been designed and made available – e.g. in Australia, France, Germany, Iceland, Italy, Switzerland – to automatically trace contacts, notify users about potential exposure to COVID-19 and invite them to isolate. In many countries the application is privacy preserving and on a voluntary basis.

The utility of these digital applications is however debated. Some built-in features make it more efficient than manual contact tracing: it is automated, reducing the burden of manual contact tracing and limiting recall bias; it is faster, as information can be transmitted in real time. However, coverage is uneven. In particular, most children and elderly do not own a smartphone or are less familiar with digital technologies. The overall adoption of the app among smartphone owners will also be a limiting factor, as well as the fraction of cases

actually triggering the alert to the contacts and the adherence to isolation of the app adopters who receive an alert.

These variables must be gauged in light of the risk factors for exposure and transmission driving the COVID-19 epidemic. First, individuals of different age contribute differently to the transmission dynamics of COVID-19. Younger individuals tend to have more contacts than adults or the elderly. On the other hand, a marked feature of COVID-19 is the strong age imbalance among cases, that may be explained by both a reduced susceptibility [7], [8] and an increased rate of subclinical infections in children compared to adults [8]–[10]. As subclinical cases are harder to detect, this implies that identification of cases and of their contacts may be dependent on age. Second, SARS-CoV-2 transmission risk varies substantially by setting. Transmissions were registered predominantly in households, in specific workplaces and in the community (linked to shopping centers, meals, parties, sport classes, etc.) [11], [12]. This is due, at least in part, to the higher risk of contagion of crowded and indoor environments [11], [13]. Notably, contacts occurring in the community are also the ones more affected by recall biases, thus more difficult to trace with manual contact tracing.

Several modelling studies have quantified the impact of contact tracing [14]–[18], with some of them addressing specific aspects of digital contact tracing [15]–[17]. Still the interplay between age and setting heterogeneity in determining the efficacy of this intervention is largely unexplored. We provide here a systematic exploration of the different variables at play. We considered France as a case study and integrated different sources of data to realistically describe the French population, in terms of its demography and social contact behavior. We accounted for the dynamics of contacts according to age and setting, and for the setting-specific risk of transmission. We used COVID-19 epidemiological characteristics for parametrization. We then modelled case detection and quarantining, isolation of their household contacts and digital contact tracing, under different hypotheses of potential reduction in transmissibility due to other effects (e.g. face-masks and increased hygiene). We quantified the impact of digital contact tracing on the whole population and on different

population groups and settings, as a function of several variables such as the rate of app adoption, the probability of detection of clinical cases, population immunity and transmission potential. Our results provide quantitative information regarding the impact of digital contact tracing within a broader response plan.

Results

Dynamic multi-setting contact network

We modelled the French population integrating available demographic and social-contact data. We collected population statistics on age, household size and composition (Figure 1 A, B), workplace and school size, smartphone penetration (Figure 1 E), and commuting fluxes. Then, by following standard approaches in the literature [19], [20] individuals were created in-silico with given gender and age and assigned to a municipality, a household, and a workplace/school according to the statistics. Smartphones were assigned to individuals depending on their age according to available statistics on French users (Figure 1 E) [21]. Overall smartphone penetration is 64%, that represent the upper bound limit of app adoption in the population – reached when 100% of individuals owning a smartphone download the app. This synthetic population reproduced the location statistics of individuals in different settings, yielding the basis of a multi-setting network of daily face-to-face contacts in household, school, workplace, community and transport (Figure 1 E) [22]–[24]. We parametrized the network from a social contact survey providing information on contacts by age and setting [25] (Figure 1 C, D). As contacts may occur repeatedly, we associated an activation rate to each contact and sampled each day contacts based on their activation rate (Figure 1 G). We imposed that 35% of the contacts registered during one day occur with daily frequency, as found in [25]. Figure 1 F and H show that the features of the resulting daily contact network matched the data: the distribution of the number of contacts was right-skewed as the empirical one reported in [25] and the contact matrix showed age assortativity and the characteristic parent-children (off-diagonal) contact pattern. As a case

study we restricted our study to a municipality with a population size of ~100,000 individuals (see Material and Methods and Supplementary Material for additional details).

COVID-19 epidemic dynamics

We modelled coronavirus transmission and outcome as shown in Figure 2 A, B. Individuals could be susceptible, S, exposed, E, pre-symptomatic preceding subclinical infection, $I_{p,sc}$, pre-symptomatic preceding clinical infection, $I_{p,c}$, subclinical infectious, I_{sc}, clinical infectious, I_c, and recovered, R. Subclinical cases had symptoms that ranged from no symptoms to mild and continued their normal activity throughout the infectious period. Clinical cases had moderate to critical symptoms and stayed at home after the onset of symptoms [9], [10] – we did not consider hospitalization. We accounted for the heterogeneous susceptibility and clinical manifestation by age as parametrized from [7], [10] (Table 1). In order to parametrize infection natural history, we combined evidence from epidemiological and viral shedding studies. We used 5.2 days for the incubation period [26], 2.3 days for the average length of the pre-symptomatic phase [27], and 7 days on average for the infectivity period after symptoms' onset [27].

We first simulated an uncontrolled epidemic assuming transmission levels corresponding to $R_0 = 3.1$, within the range of values estimated for COVID-19 in France at the early stage of the pandemic [1], [4]. The generation time resulting from our model and parameters had mean value of 6.0 days (95% CI [2,17]), in agreement with epidemiological estimates [9], [27], [28]. Figure 2 B and C show the repartition of cases among age groups and settings at the early stage and during the whole course of the epidemic. Age-specific infection probability was higher among young adults, while clinical infections were shifted towards older population with respect to the overall (clinical and subclinical) cases, as noted in previous observational and modelling works [8]. The age profile changed in time with children infected later as the epidemic unfolded [8], [29]. Transmissions occurred predominantly in household and workplaces followed by the community setting [11].

Contact tracing

We quantified the impact of combined household isolation and digital contact tracing considering the possible scenario of a second epidemic wave in the country. We thus assumed some level of immunity to the virus – exploring a range from 0 to 15% of the population. We considered interventions based on the use of digital contact tracing, coupled with testing and isolation of clinical cases and households. 50% of individuals with clinical symptoms were assumed to get tested after consulting a doctor and to isolate if positive. Higher and lower percentages were also considered.

Case tracing was assumed to start when a case with clinical symptoms tested positive and was invited to isolate. Household members were also invited to isolate. If the index case had the app installed, the contacts he/she registered in the previous $D = 7$ days were notified and could decide to isolate. We explored different probabilities of case detection and several levels of app adoption in the population. In addition to the detection of clinical cases, we assumed that a small proportion (5%) of subclinical cases was also identified. These may be cases with very mild, aspecific symptoms who decided to get tested as part of vulnerable groups (i.e. co-morbidity) or because highly exposed to the infection (health care professionals). Isolated individuals resumed normal daily life if infection was not confirmed. We took 7 days as the time needed for being confirmed negative because multiple tests and some delay since the exposure are needed for a negative result to be reliable. Infected individuals got out of quarantine after 14 days unless they still have clinical symptoms after the time is passed. They may, however, decide to drop out from isolation each day with if they don't have symptoms [17].

Figure 3 summarizes the effect of the interventions. We compared the uncontrolled scenario ($R = R_0 = 3.1$) with scenarios where the transmissibility is reduced due to the adoption of barrier measures (R down to 1.5). We also assumed 10% of the population to be immune to the infection [4]. Panels A-C shows the results for $R = 2.6$ and $R = 1.7$. With $R = 2.6$ (Figure

3 A, C), the relative reduction of peak incidence due to household isolation only would be 30%. The inclusion of digital contact tracing would increase the relative reduction to 36% with ~20% app adoption, and to 67% with ~60% app adoption – i.e., 90% of individuals owning a smartphone use the app. This corresponds to an additional mitigation effect ranging from 20% to 120% provided by contact tracing compared to household isolation only. With $R = 1.7$ (Figure 3 B, C), we find that ~20% app adoption would reduce the peak incidence by 45% (additional mitigation effect of 25%), while the reduction would reach 89% in a scenario of ~60% app adoption (additional mitigation effect of 147%). According to the projections in Ref.[1], intensive care units occupation would remain below the saturation level with incidence below 0.4 /1000 hab. In the scenario with $R = 1.7$, this would be reached with app adoption greater than ~30% (grey dashed line in Figure 3 B). Stronger reductions could be obtained with more efficient detection of clinical cases (Figure 3 E, H). Results show similar trends across different levels of population immunity, with higher relative impacts predicted for low immunity (Figure 3 F, I).

We analyzed the simulation outputs to characterize index cases and their contacts and relate this to the reduction in number of cases by age and setting. We found that adults represented the majority of index cases (Figure 4 D), while their household contacts were mostly children. The app registered mostly contacts with adults, and the tracked contacts were occurring predominantly in workplaces and in the community (Figure 4 A). This results in a heterogenous reduction in transmission (TRR) by setting and age group. Household isolation reduced transmission in all settings, with the smallest effect in workplaces (Figure 4 B). Digital contact tracing has instead a high TRR at work, in the community and in transports (Figure 4 C). Household isolation reached mostly children (< 15 years old) and the elderly (especially the 75+ group) with the smallest effect in the 15-59 years old (Figure 4 E). Adopting digital contact tracing led to an increased TRR with age, even among the oldest age range (Figure 4 F). This result shows the indirect effect of digital tracing: due to the central role of adults in the transmission of SARS-CoV-2 towards all age-groups, avoiding

adult infections led to less transmission to the elderly. We also tested the case in which elderly people (70+) owning a smartphone did not install the app at all, because less familiar with digital technologies, and we found no appreciable effect. These results and additional details are provided in the Supplementary Material.

Traced and Isolated individuals

Feasibility of contact tracing depends on the number of traced contacts who require assistance and virological tests. In a scenario with high detection rate (80%), we found that for each detected case 1.5 contacts were identified on average through household isolation but up to 7.5 with app adoption at 57% for $R = 2.6$ and 10 for $R = 1.7$ (Figure 5 D). This number was however subject to fluctuations (Figure 5 A). Overall, the maximal fraction of the population quarantined at any given time was ~50 per 1000 habitants in a scenario with $R = 2.6$, and was between ~1 and ~4.5 per 1000 habitants when $R = 1.7$ (Figure 5 B and E). The latter case corresponded to the situation in which high levels of app adoption were able to strongly reduce spreading, thus the proportion of isolated individuals declined in time, signaling the success of quarantining in preventing the propagation of the infection. A total of 30 per 1000 habitants were isolated in a scenario with $R = 1.7$, assuming high app adoption. At $R = 2.6$, 1030 per 1000 habitants were isolated at the end of the epidemic meaning that certain individuals were isolated more than once. In all scenarios, the increase of app adoption inevitably determined an increase in the proportion of people that were unnecessarily isolated, i.e. of individuals that were not infected but still isolated (Figure 5C, F): this proportion increased from 61% to 84% with the increase of app adoption from 0% to 57% (note that the case of 0% app adoption implies that 60% of individuals who were isolated through household isolation were not infected). These numbers were similar for the two tested values of $R = 1.7$ and 2.6 (Figure 5 C and F).

Discussion

Quantifying the impact of digital contact tracing is essential to envision this strategy within a wider response plan against the COVID-19 epidemic. We modelled this intervention together with household isolation assuming a 50% detection of clinical cases. In a scenario of high transmissibility ($R = 2.6$), we found that household isolation by itself would produce a reduction in peak incidence of 30%, while the inclusion of digital contact tracing could increase this effect by 20% for a reasonably achievable app adoption (~20% of the population), and by 120% for a large-scale app adoption (~60%). At a moderate transmissibility level ($R = 1.7$), the app would substantially damp transmission (36% to 89% peak incidence reduction for increasing app adoption), bringing the epidemic to manageable levels if adopted by 32% of the population or more. Importantly, the app-based tracing and household isolation have different effects across settings, the first intervention efficiently preventing transmissions at work that are not well targeted by the second. Moreover, app-based contact tracing also yields a protection for the elderly despite the lower penetration of smartphones in this age category.

In several countries, the lockdown has reduced transmission to very low levels. However, low population immunity leaves countries at risk for possible renewed increases of the number of cases. This may potentially occur at any time – as evidenced by the current high incidence levels in the United States [29] and the resurgence of cases reported in the region of Pays de la Loire in France [5] –, and could become increasingly likely after summer. Indeed, during summer, high temperature and humidity, increased ventilation and outdoor activities reduce the risk of contagion [13]. These environmental effects will disappear with the return of the cold season. This urges the planning of sustainable non-pharmaceutical interventions, able to suppress COVID-19 spread while having limited impact on the economy and on individuals' daily life.

Many countries are now increasing their capacity to detect cases and track their contacts. In France, hundreds of transmission clusters have been identified and controlled since the end of the lockdown period [30]. Still, traditional surveillance relies on human intervention and could saturate at increased epidemic activity. The automated tracking of contacts could then provide an important complementary tool. Here we found that digital contact tracing could reduce attack rate and peak incidence, in agreement with previous works [15], [16]. The impact of the measure would depend on population immunity, thus geographical heterogeneities should be expected, since immunity levels are likely higher in those areas more severely hit during the first wave. On the other hand, app adoption as well may be higher in these areas because of risk aversion behavior [31]. Also, higher participation rates may be expected in dense urban areas to protect from exposure from random encounters (e.g. in public transports).

Under realistic hypotheses, the intervention would not be able alone to bring the epidemic under control in a scenario where transmission is high [15], [16], mainly due to the strong role of asymptomatic transmission in fueling the epidemic [1], [4]. We considered also lower values of R as recent studies suggest that the combination of facemask, physical distance and hand hygiene may substantially hinder SARS-COV-2 transmission [32]. The use of facemasks in public spaces is recommended by the majority of public health institutions, and it is mandatory in certain cases. In France the adoption of face masks by the general population has increased over time reaching a peak at ~50% [30]. However, the quantification of the effect of these measures and their effect on COVID-19 reproductive ratio is uncertain. We found that a reduction of the epidemic to a manageable level would be possible with a moderate R (e.g. R =1.7 explored here).

Improved case finding is the first step towards a successful contact-tracing intervention. Many countries decided to screen anyone with symptoms compatible with COVID-19 [33]. According to the weekly report of Public Health France of July 9, tests had been prescribed to 97% of all patients declaring COVID-19 symptoms upon consultation [30]. This policy

could become less feasible during winter, due to the co-circulation of other respiratory infections, that would increase the rate of false positive and testing needs. Given that the majority of cases do not require hospitalization, case detection effectiveness is also influenced by the consultation rate. This has been estimated to be around ~30% with peaks at ~45% by the participatory surveillance platform covidnet.fr [30]. Higher detection levels would substantially improve the impact of contact tracing policies (Figure 3 E, H). Increased population awareness is thus essential for the efficient monitoring of the epidemic and its containment through contact tracing.

Besides detection efficiency, the other important factor determining the efficacy of digital contact tracing is clearly the app adoption. Adoption levels have been low (<5%) in many countries (e.g. Italy 4% [34]), whereas higher levels were observed, e.g., in Australia (6 millions download, 25% of the population) [35] and Iceland (~150 thousand, 38%) [36]. In France, it is reported to be around 3% [37]. These values may increase in case of a rebound of the epidemic, due to increased concern of the population. Individuals may be more incline to use the app if they perceive a direct and immediate benefit from its use. This may be implemented through, e.g., easy access to testing in case they are notified as contacts and assistance by public health professionals. Moreover, increased transparency and ethical debate are essential to reassure the population concerning privacy [38], [39].

The results presented here are based on an agent-based model that describes age-specific risk factors for exposure and transmission: contact rates, contacts by location, susceptibility to the virus, probability of being detected and rate of app adoption. The interplay between these features has a profound impact on COVID-19 spread and affects the efficacy of household isolation and digital contact tracing. To account for contact heterogeneities we used statistics on population demography, combined with social contact surveys to build a multi-setting contact network, similarly to previous works [14], [17], [22]–[24]. The network is also dynamic in time as it captures the repetition of a certain number of contacts (e.g. relationships) and the occurrence of random encounters. Social contact data provide an

invaluable information source to study the current COVID-19 outbreak [1], [4]. Previous projections on the impact of contact tracing rely on a similar approach in some cases [14]. Other works make use of high resolution data [15], [16], that are more reliable than contact surveys, but are restricted to specific settings or population groups. Despite the difference in the data source and approach, the results of these studies are consistent and in agreement with our work on the overall impact of the intervention.

We modelled age-specific epidemiological characteristics based on available knowledge in the literature. Children are less impacted by the COVID-19 epidemic. This may be explained by reduced susceptibility and severity, with accumulating evidence that both effects are acting simultaneously [8]. The strength of these effects is still debated and the infection risk for children should not be minimized. However, these differences have implications for digital contact tracing. Indeed, it is precisely in the group that plays a central role in transmission and where cases are more likely symptomatic (i.e., adults) that the app coverage is already the highest. Our model shows that taken together, these characteristics reinforce the impact of digital tracing and provides indirect protection in the elderly population. This occurs even if no adoption is registered in the elderly population.

Our study is affected by limitations. First, we analyzed the effect of digital contact tracing on COVID-19 incidence on the general population. Crucial information for public health authorities would be to quantify the effect in time of these measures on hospitalizations. This would require to couple our model for COVID-19 transmission in the general population with a model describing disease severity and within-hospital patient trajectories [14], [17]. Second, the model does not account for transmission in nursing homes. This setting is where the majority of transmissions among elderly occurred. At the same time, however, the response to the COVID-19 epidemic in this setting relies mostly on routine screening of symptoms and frequent testing of residents, together with face masks and strict hygiene for visitors. Third, clustering effects are partially captured by the model thanks to the repetition of contacts, but effects may be larger in real contact patterns. Results obtained on real contact

data, however, are similar to ours obtained on synthetically reconstructed contacts [15]. Eventually, parameterized with data from a social contact survey [25], the model cannot account for crowding events. These events were suggested to play an important role in the epidemic dynamics [12], and may also impact the effectiveness of contact tracing. Contact tracing may be more effective in networks showing large fluctuations in the number of contacts per individual [40]. Therefore, results presented here may be conservative.

Material and Methods

Synthetic population

The model simulates the population of Metropolitan France representing individual inhabitants. This approach is similar to studies done previously e.g. for Italy[19] and for USA[20]. The French synthetic population is based on the National Institute of Statistics and Economic Studies (INSEE) censuses. The individuals were grouped by municipalities according to the administrative borders. The number of households and the age structure of their inhabitants, sizes of schools and workplaces, fluxes of commuters between municipalities also followed the distribution of these statistics found in the INSEE data. Population size was kept constant as it aims to simulate one season of the epidemic.

To generate the population, we defined several statistics derived from INSEE publicly available data:

- The list of municipalities ("les communes de France") of Metropolitan France (2015) with each municipality described by its INSEE code, population size, number of schools of six different levels (from kindergarten to university), number of workplaces in given size categories (0-9, 10-49, 50-99, 100-499, 500-999 and over 1000 employees) (Populations légales 2017, INSEE, https://www.insee.fr/fr/statistiques/4265429?sommaire=4265511).
- Statistics regarding the percentage of people in given age groups enrolled in each of six school levels, employed and unemployed (Bilan démographique 2010, INSEE, https://www.insee.fr/fr/statistiques/1280950).

- The age pyramid for France as the population fractions of individuals of a given age (Bilan démographique 2010, INSEE, https://www.insee.fr/fr/statistiques/1280950).
- The number of people commuting to work between each pair of municipalities (Mobilités professionnelles en 2016: déplacements domicile - lieu de travail, INSEE, https://www.insee.fr/fr/statistiques/4171554).
- The number of people commuting to school between each pair of municipalities (Mobilités scolaires en 2015: déplacements domicile - lieu d'études, INSEE, https://www.insee.fr/fr/statistiques/3566470).
- The probability distributions of sizes of households in France (Couples - Familles - Ménages en 2010. INSEE, https://www.insee.fr/fr/statistiques/2044286/?geo=COM-34150.)
- The probability of age class of individuals depending on its role in the household: child of a couple, child of a single adult, adult in a couple without children, adult in a couple with children (Couples - Familles - Ménages en 2010. INSEE, https://www.insee.fr/fr/statistiques/2044286/?geo=COM-34150).

With the above statistics, the synthetic population was generated in the following steps:

1. Initialization of all the municipalities with an appropriate number of schools of each type and workplaces of given sizes.
2. Creation of schools in each municipality according to given statistics.
3. Creation of workplaces in each municipality according to given statistics.
4. Definition of the commuters fluxes between municipalities.

Each municipality has a defined number of inhabitants and individuals are created (one by one) until this number is reached. Each individual was assigned an age, a school or a workplace (or is assigned to stay at home) according to probability distributions derived from the data mentioned above.

The number of households in municipalities were not defined explicitly, but their number depends on the number of individuals. The municipal population size and statistics regarding family demographics constrain the number of households. Additional details on the algorithm for the population reconstruction are provided in the Supplementary Material.

Face-to-face contact network

The synthetic population encodes information on the school, workplace, household and community each individual belongs to. We used this information to extract a dynamic network representing daily face-to-face contacts. We parametrized this network based on contacts' statistics for the French population[25].

First, we generated a time aggregated network representing all contacts that can potentially occur – we will call this *acquaintance network,* with some abuse of language since it includes also sporadic contacts. Second, to each contact we assigned a daily rate of activation. Then, in the course of the simulation we sampled contacts each day based on their rate.

The acquaintance network has five distinct layers representing contacts in household (layer H), workplace (layer W), school (layer S), community (layer C) and transports (layer T). The household layer is formed by a collection of complete networks linking individuals in the same household. The W, S, C, and T layers are formed by collections of Erdős–Rényi networks generated in each location i, with average degree χ_i. A location can be a workplace (W layer), a school (S layer) and a municipality (C and T layers). χ_i is extracted at random for each place and depends on the type and size of the location. In particular, when the size of a location is small we assume that each individual enters in contact with all the others frequenting the same place. As the size increases the number of contacts saturates.

Once the acquaintance network was built a daily activation rate x was assigned to each link according to a cumulative distribution that depends on the layer s. For simplicity we assumed this distribution to be the same for $s = W, S, C$, while we allowed it to be different in household

(where contacts are more frequent) and in transports (where contacts are sporadic). Parameters were tuned based on average daily number of contacts, proportion of contacts by setting, and contact frequency as provided in [25] (Figure 1 C D). Additional details on the network reconstruction and parametrization are provided in the Supplementary Material.

Transmission model

We defined a minimal model of COVID-19 spread in the general population that accounts for two levels of symptoms: none to mild (subclinical cases, I_{sc}), and moderate to severe (clinical cases, I_c). We assumed that clinical cases stay at home after developing symptoms. Susceptible individuals, if in contact with infectious ones, may get infected and enter the exposed compartment (E). After an average latency period ϵ^{-1} they become infectious, developing a subclinical infection with probability p_{sc}^{A} and a clinical infection otherwise. From E, before entering in either I_{sc} or I_c, individuals enter first a prodromal phase (either $I_{p,sc}$ or $I_{p,c}$), that lasts on average μ_p^{-1} days and where individuals do not show any sign of illness, despite being already infectious. Contact-tracing, population-screening and modelling studies provide evidence that infectivity is related to the level of symptoms, with less severely hit individuals being also less infectious [9]. Therefore, we assumed that subclinical cases, $I_{p,sc}$ and I_{sc} have a reduced transmissibility compared to $I_{p,c}$ and I_c. This is modulated by the scaling factor β_I. We neglected hospitalization and death and assumed that with rate μ infected individuals become recovered.

COVID-19 has heterogeneous impact across age groups [7]–[9]. This may be driven by differences in susceptibility [7], differences in clinical manifestation [9], [10] or both [8]. We considered here both effects in agreement with recent modelling estimates [8]. Susceptibility by age, σ_A, was parametrized from [7], while clinical manifestation, p_{sc}^{A}, was parametrized from a large-scale descriptive study of the COVID19 outbreak in Italy [10].

Transition rates are summarized in Figure 2 B, and parameters and their values are listed in Table 1. The incubation period was estimated to be around 5.2 days from an early analysis of 425 patients in Wuhan [26]. COVID-19 transmission potential varies across settings, populations and social contexts [11], [12]. In particular, indoor places were found to increase the odds of contagion 18.7 times compared to an open-air environment [13]. In our model we assumed all contacts at work, school and transport occur indoor and have the same transmission risk (ω_S). In the contact survey of Béraud et al. [25], 46% of contacts in the community were occurring outdoors. Combining this information with the 18.7 indoor vs. outdoor risk ratio leads to a 60% relative risk of community contacts with respect to workplace/school/transport contacts. Contacts in household are generally associated to a higher risk with respect to other settings, because they last longer and there is a higher risk of environmental transmission. We assumed, in line with [17], that the transmission risk associated with them is twice the one in workplace/school/transport.

Modelling contact tracing

Self-isolation and isolation of household contacts

Self-isolation and isolation of household contacts was modelled according to following rules:

- As an individual shows clinical symptoms, s/he is detected with probability $p_{d,c}$. If detected, case confirmation, isolation and contacts' isolation occur with rate $r_{d,c} = 0.9$ upon symptoms onset
- Subclinical individuals are also detected with probability $p_{d,sc} = 0.05$ and rate $r_{d,sc} = 0.5$.
- The individual's family members are isolated with probability $p_{c,h} = 0.9$
- We assume contacts are tested and the follow up guarantees that all individuals who got infected prior to isolation are detected. Thus, contacts that are negative (either susceptible or recovered at the time of isolation) terminate their isolation after 7 days. The index-case and the positive contacts are isolated for 14 days. Contacts with no clinical symptoms have a probability $p_{drop} = 0.02$ of drop-out each day.

- For both the case and the contacts, isolation is implemented by assuming no contacts outside household and contacts within a household having a weight ω_H reduced by a factor $\iota = 0.5$.

Digital contact tracing

We assumed that contact tracing is adopted in combination with self-isolation and isolation of household members. Therefore, to the rules above, we added the following ones:

- At the beginning of the simulation, a smartphone is assigned to individuals with probability p_{sm}^{A}, based on the statistics of smartphone penetration (0% for [0,11], 86% for [12,17], 98% for [18,24], 95% for [25,39], 80% for [40,59], 62% for [60,69], 44% for 70+) [21].
- Each individual with a smartphone has a probability p_a to download the app (we explored values between 0 and 0.9)
- Only contacts occurring between individuals with a smartphone and the app are traced.
- If the individual owns a smartphone and downloaded the app the contacts that s/he has traced in the period since $D = 7$ days before his/her detection are isolated with probability $p_{c,a} = 0.9$
- We assume contacts are tested and the follow up guarantees that all individuals who got infected prior to isolation are detected. Thus, contacts that are negative (either susceptible or recovered at the time of isolation) terminate their isolation after 7 days. The index-case and the positive contacts are isolated for 14 days. Contacts with no clinical symptoms have a probability $p_{drop} = 0.02$ of drop-out each day.
- For both the case and the contacts, isolation is implemented by assuming no contacts outside household and contacts within a household having a weight ω_H reduced by a factor $\iota = 0.5$.

Reference

[1]L. D. Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza, “Expected impact of lockdown in Île-de-France and possible exit strategies,” *medRxiv*, p. 2020.04.13.20063933, Apr. 2020, doi: 10.1101/2020.04.13.20063933.

[2]S. Flaxman *et al.*, “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe,” *Nature*, pp. 1–8, Jun. 2020, doi: 10.1038/s41586-020-2405-7.

[3]M. Pollán *et al.*, “Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study,” *The Lancet*, vol. 0, no. 0, Jul. 2020, doi: 10.1016/S0140-6736(20)31483-5.

[4]H. Salje *et al.*, “Estimating the burden of SARS-CoV-2 in France,” *Science*, May 2020, doi: 10.1126/science.abc3517.

[5]SPF, “COVID-19 : point épidémiologique du 9 juillet 2020.” /maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-9-juillet-2020 (accessed Jul. 16, 2020).

[6]mbevand, *mbevand/florida-covid19-line-list-data*. 2020.

[7]J. Zhang *et al.*, “Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China,” *Science*, Apr. 2020, doi: 10.1126/science.abb8001.

[8]N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, and R. M. Eggo, “Age-dependent effects in the transmission and control of COVID-19 epidemics,” *Nature Medicine*, pp. 1–7, Jun. 2020, doi: 10.1038/s41591-020-0962-9.

[9]Q. Bi *et al.*, “Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study,” *The Lancet Infectious Diseases*, vol. 0, no. 0, Apr. 2020, doi: 10.1016/S1473-3099(20)30287-5.

[10]F. Riccardo *et al.*, “Epidemiological characteristics of COVID-19 cases in Italy and estimates of the reproductive numbers one month into the epidemic,” *medRxiv*, p. 2020.04.08.20056861, Apr. 2020, doi: 10.1101/2020.04.08.20056861.

[11]Q. J. Leclerc, N. M. Fuller, L. E. Knight, CMMID COVID-19 Working Group, S. Funk, and G. M. Knight, “What settings have been linked to SARS-CoV-2 transmission clusters?,” *Wellcome Open Research*, vol. 5, p. 83, May 2020, doi: 10.12688/wellcomeopenres.15889.1.

[12]B. M. Althouse *et al.*, “Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2,” *arXiv:2005.13689 [physics, q-bio]*, May 2020, Accessed: Jul. 17, 2020. [Online]. Available: http://arxiv.org/abs/2005.13689.

[13]H. Nishiura *et al.*, “Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19),” *medRxiv*, p. 2020.02.28.20029272, Apr. 2020, doi: 10.1101/2020.02.28.20029272.

[14]A. Aleta *et al.*, “Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic,” *medRxiv*, p. 2020.05.06.20092841, May 2020, doi: 10.1101/2020.05.06.20092841.

[15]G. Cencetti *et al.*, “Using real-world contact networks to quantify the effectiveness of digital contact tracing and isolation strategies for Covid-19 pandemic,” *medRxiv*, p. 2020.05.29.20115915, May 2020, doi: 10.1101/2020.05.29.20115915.

[16]A. J. Kucharski *et al.*, “Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a

mathematical modelling study," *The Lancet Infectious Diseases*, vol. 0, no. 0, Jun. 2020, doi: 10.1016/S1473-3099(20)30457-6.

[17]"BDI-pathogens/covid-19_instant_tracing," *GitHub*. https://github.com/BDI-pathogens/covid-19_instant_tracing (accessed May 04, 2020).

[18]L. Ferretti *et al.*, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Science*, Mar. 2020, doi: 10.1126/science.abb6936.

[19]M. Ajelli *et al.*, "Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models," *BMC Infectious Diseases*, vol. 10, p. 190, 2010, doi: 10.1186/1471-2334-10-190.

[20]D. L. Chao, M. E. Halloran, V. J. Obenchain, and I. M. L. Jr, "FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation Model," *PLOS Comput Biol*, vol. 6, no. 1, p. e1000656, Jan. 2010, doi: 10.1371/journal.pcbi.1000656.

[21]"Baromètre du numérique 2019," Nov. 01, 2019. https://www.credoc.fr/publications/barometre-du-numerique-2019 (accessed Apr. 07, 2020).

[22]L. Fumanelli, M. Ajelli, P. Manfredi, A. Vespignani, and S. Merler, "Inferring the Structure of Social Contacts from Demographic Data in the Analysis of Infectious Diseases Spread," *PLOS Computational Biology*, vol. 8, no. 9, p. e1002673, Sep. 2012, doi: 10.1371/journal.pcbi.1002673.

[23]D. Mistry *et al.*, "Inferring high-resolution human mixing patterns for disease modeling," *arXiv:2003.01214 [physics, q-bio]*, Feb. 2020, Accessed: May 31, 2020. [Online]. Available: http://arxiv.org/abs/2003.01214.

[24]Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, and A. Vespignani, "Measurability of the epidemic reproduction number in data-driven contact networks," *PNAS*, vol. 115, no. 50, pp. 12680–12685, Dec. 2018, doi: 10.1073/pnas.1811115115.

[25]G. Béraud *et al.*, "The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases," *PLOS ONE*, vol. 10, no. 7, p. e0133203, Jul. 2015, doi: 10.1371/journal.pone.0133203.

[26]Q. Li *et al.*, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia," *New England Journal of Medicine*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020, doi: 10.1056/NEJMoa2001316.

[27]X. He *et al.*, "Temporal dynamics in viral shedding and transmissibility of COVID-19," *Nature Medicine*, pp. 1–4, Apr. 2020, doi: 10.1038/s41591-020-0869-5.

[28]D. Cereda *et al.*, "The early phase of the COVID-19 outbreak in Lombardy, Italy," *arXiv:2003.09320 [q-bio]*, Mar. 2020, Accessed: Jul. 17, 2020. [Online]. Available: http://arxiv.org/abs/2003.09320.

[29]mbevand, *mbevand/florida-covid19-line-list-data*. 2020.

[30]SPF, "COVID-19 : point épidémiologique du 2 juillet 2020." /maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-2-juillet-2020 (accessed Jul. 14, 2020).

[31]G. Pullano, E. Valdano, N. Scarpa, S. Rubrichi, and V. Colizza, "Population mobility reductions during COVID-19 epidemic in France under lockdown," *medRxiv*, p. 2020.05.29.20097097, Jun. 2020, doi: 10.1101/2020.05.29.20097097.

[32]D. K. Chu *et al.*, "Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis," *The Lancet*, vol. 0, no. 0, Jun. 2020, doi: 10.1016/S0140-6736(20)31142-9.

[33]"Coronavirus (COVID-19) Testing - Statistics and Research," *Our World in Data*. https://ourworldindata.org/coronavirus-testing (accessed Jun. 26, 2020).

[34]"Immuni ferma a 4 milioni di download. Un'occasione persa per gli italiani contro il COVID-19?," *Hardware Upgrade*. https://www.hwupgrade.it/news/telefonia/immuni-ferma-a-4-milioni-di-download-un-occasione-persa-per-gli-italiani-contro-il-covid-19_90539.html (accessed Jul. 14, 2020).

[35]"The curve of people downloading COVIDSafe has flattened. Has the app done enough?," Jun. 01, 2020. https://www.abc.net.au/news/2020-06-02/coronavirus-covid19-covidsafe-app-how-many-downloads-greg-hunt/12295130 (accessed Jun. 26, 2020).

[36]"Nearly 40% of Icelanders are using a covid app—and it hasn't helped much," *MIT Technology Review*. https://www.technologyreview.com/2020/05/11/1001541/iceland-rakning-c19-covid-contact-tracing/ (accessed Jun. 26, 2020).

[37]"StopCovid : à peine 14 alertes envoyées aux utilisateurs depuis le lancement." https://www.journaldunet.com/media/guide-de-l-iphone/1490935-stopcovid-a-peine-14-alertes-envoyees-aux-utilisateurs-depuis-le-lancement/ (accessed Jul. 02, 2020).

[38]*DP-3T/documents*. DP^3T, 2020.

[39]M. Nanni *et al.*, "Give more data, awareness and control to individual citizens, and they will help COVID-19 containment," *arXiv:2004.05222 [cs]*, Apr. 2020, Accessed: Jun. 23, 2020. [Online]. Available: http://arxiv.org/abs/2004.05222.

[40]I. Z. Kiss, D. M. Green, and R. R. Kao, "Infectious disease control using contact tracing in random and scale-free networks," *Journal of The Royal Society Interface*, vol. 3, no. 6, pp. 55–62, Feb. 2006, doi: 10.1098/rsif.2005.0079.

Acknowledgements

This study was partially supported by the ANR project DATAREDUX (ANR-19-CE46-0008-01) to AB, VC and PYB; EU H2020 grants MOOD (H2020-874850) to VC, CP, and PYB, and RECOVER (H2020-101003589) to VC; the Municipality of Paris (https://www.paris.fr/) through the programme Emergence(s) to JML, BAC, PB, FP and CP; the ANR and Fondation de France through the project NoCOV (00105995) to PYB, CP.

Figures and Tables

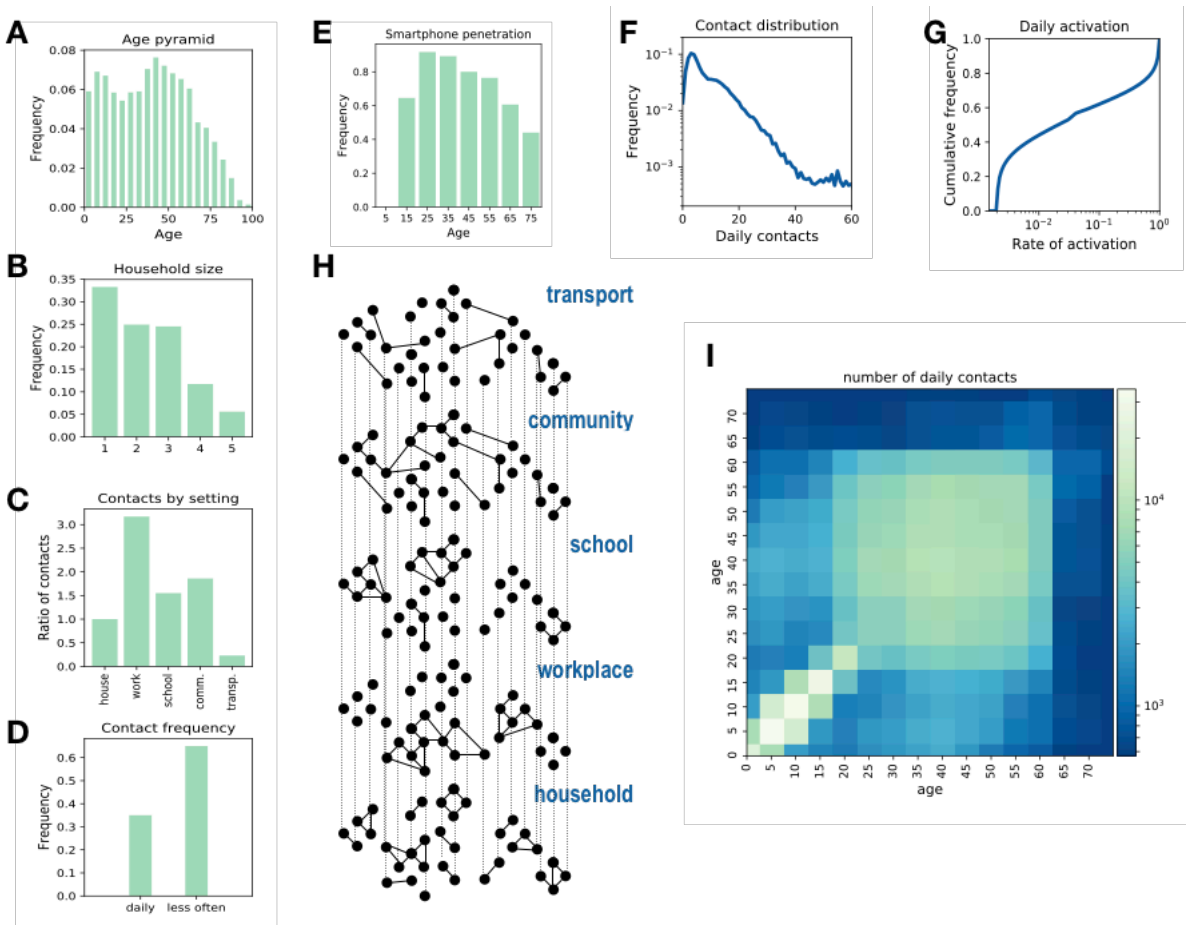

Figure 1 Synthetic population. A-E Key statistics used as input for the synthetic population reconstruction. **A** Age pyramid for France (source INSEE). **B** Household size (source INSEE). **C** Ratio of contacts by setting with respect to household contacts [25]. **D** Fraction of contacts occurring each day or less frequently [25]. **E** Smartphone penetration by age. The overall average adoption in the population is 64% [21]. **F** Distribution of the number of daily contacts. **G** Cumulative distribution of the activation rate associated to the contacts, calibrated in order to be consistent with the information of panel D. **H** Sketch of the construction of the contact network: contacts among individuals were represented as a multi-layer dynamical network, where each layer includes contacts occurring in a specific setting. **I** Age contact matrix computed from the contact network model.

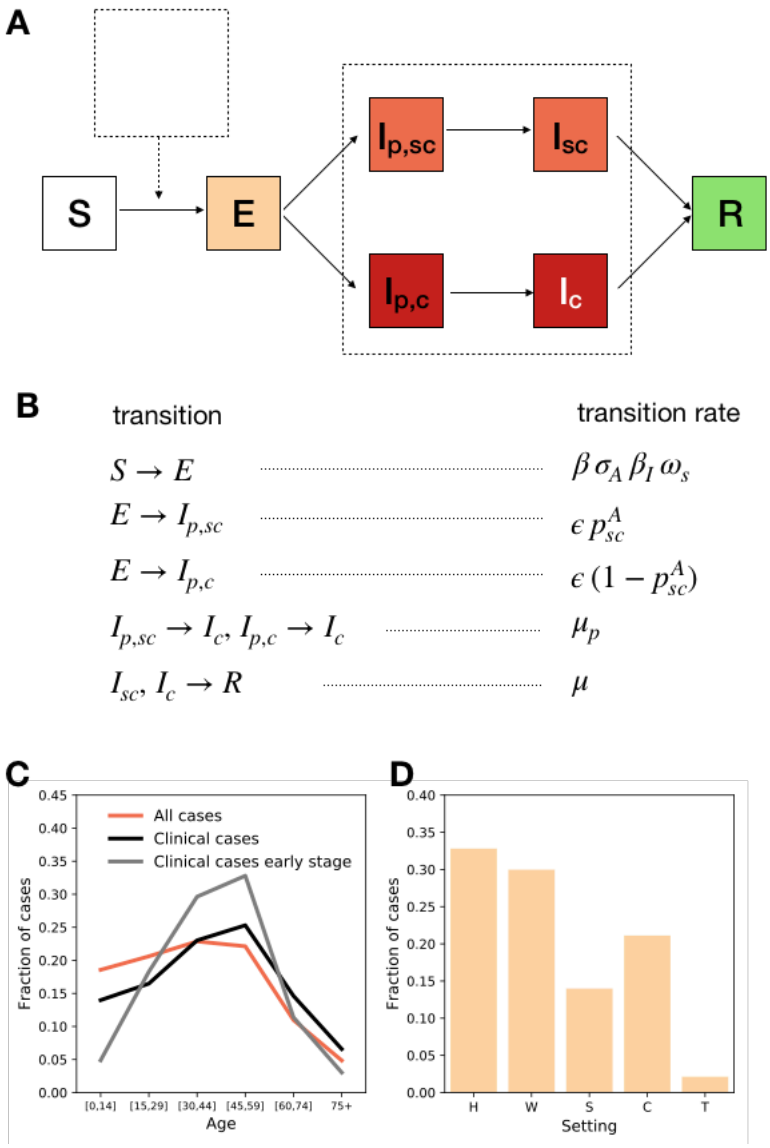

Figure 2. Modelling COVID-19 epidemic. **A, B** Compartmental model summarizing the epidemic states and transitions between states. **C** Cases by age for an uncontrolled epidemic. We show all cases (clinical and subclinical) in red and clinical cases in black. The grey line shows the clinical cases in the early stage of the epidemic (here defined as the first 30 days), with less cases among children than in later stages. **D** Transmission by setting (H, W, S, C, T stand respectively for household, workplace, school, community, transport). The simulations were done with $\beta = 0.25$ corresponding to $R_0 = 3.1$. Additional aspects of the outbreak are reported in the Supplementary Material.

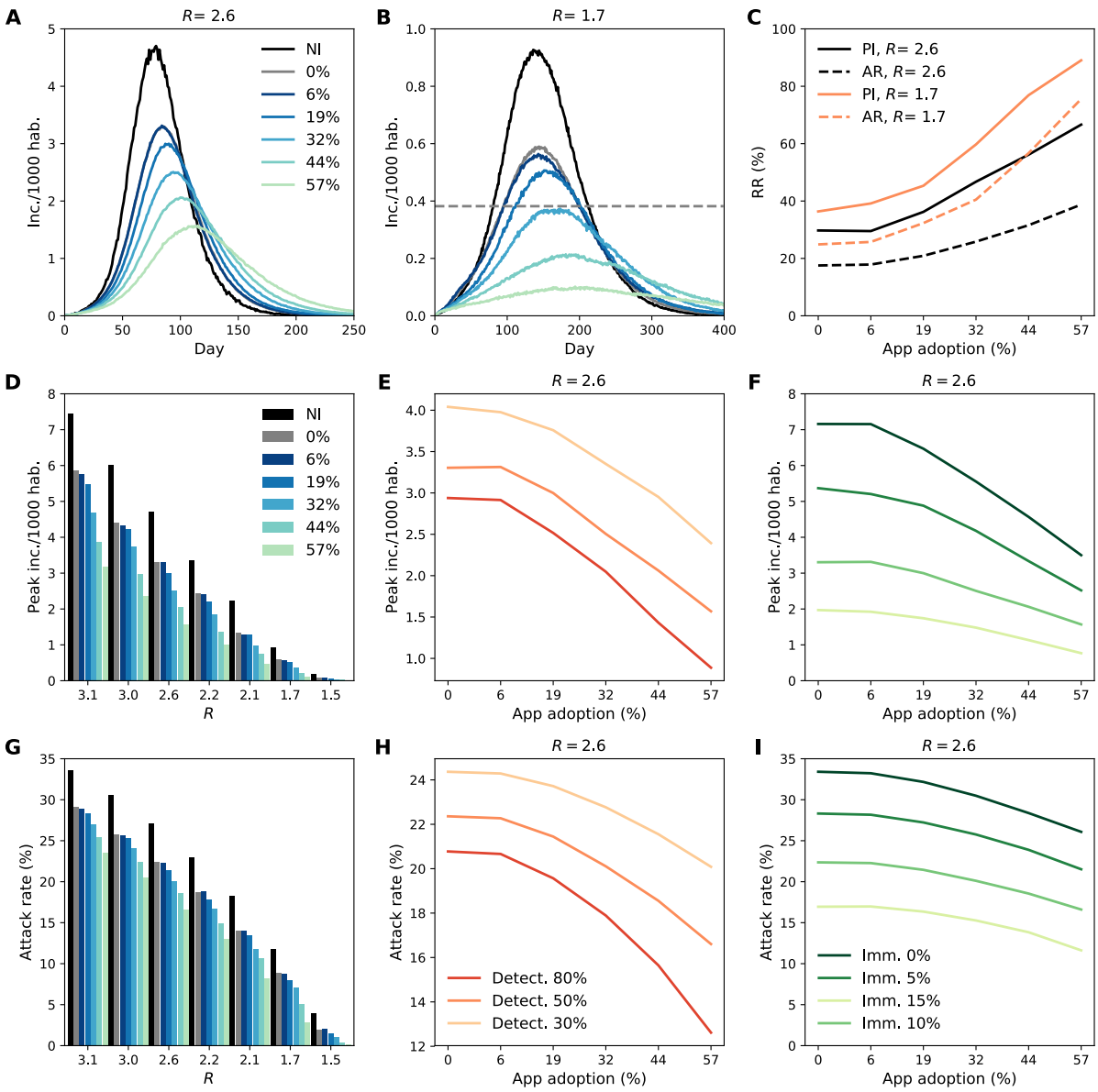

Figure 3 Impact of digital contact tracing and household isolation on the epidemic. A, B Temporal evolution of the incidence (clinical cases) with different app adoptions with $R = 2.6$ and $R = 1.7$, respectively. We explored app adoption levels from 0%, corresponding to household isolation only, to 57%, corresponding to the 90% of smartphone users. The black curve shows the scenario with no intervention (NI). Immunity is 10% and clinical case detection 50%. Incidence threshold level associated to ICU saturation is showed with a dashed grey line in panel B. **C** Relative reduction (RR) in attack rate (AR) and peak incidence (PI) as a function of the app adoption for the same scenario as in A, B. RR is computed as $\frac{x_{ref}-x}{x_{ref}}$, where x is either peak incidence or attack rate and x_{ref} is the value of the quantity in the absence of any measure with the same value of R and immunity. Here the attack rate is the cumulative number of clinical cases since the beginning of the simulation, i.e. the 10% of immune individuals are not included in the computation. **D, G** Peak incidence and attack rate for different values

of R and different app adoptions. Immunity is 10% and case detection 50%. **E, H** Peak incidence and attack rate as a function of app adoption for different values of case detection. Immunity is 10%, $R = 2.6$. **F, I** Peak incidence and attack rate as a function of app adoption for different values of immunity. Case detection is 50%, $R = 2.6$.

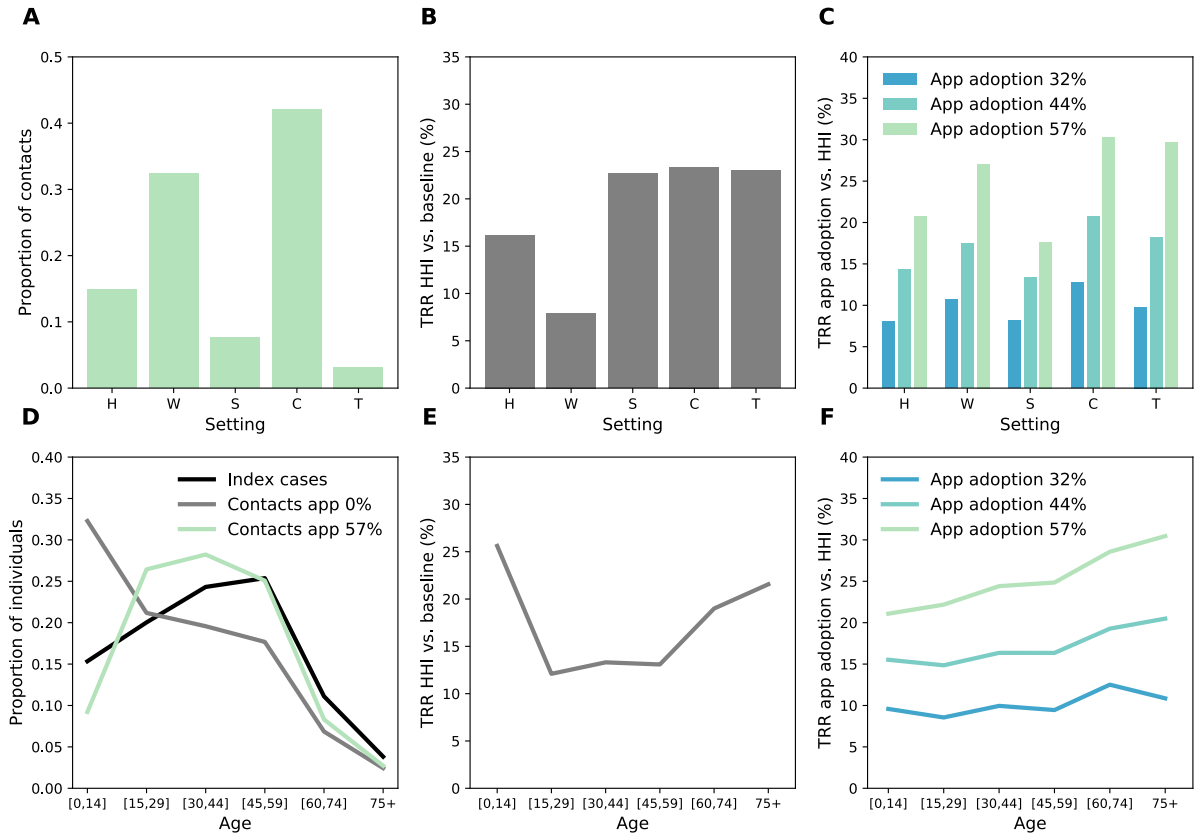

Figure 4 Effect of digital contact tracing and household isolation by age and setting. A Repartition among the different settings of the contacts detected by contact tracing (57% app adoption). **B** Relative reduction in transmission (TRR) by setting obtained with household isolation. **C** TRR obtained with digital contact tracing with respect to household isolation only, for three values of app adoption. **D** Repartition among the different age groups of the index cases and of the detected contacts, in a scenario with household isolation only, and with the inclusion of digital contact tracing (57% app adoption). The repartition of index cases is very similar in the two scenarios, thus only the one with household isolation is shown for the sake of clarity. **E** TRR by age group of the infected as obtained with household isolation only. **F** TRR of digital contact tracing with respect to household isolation only. We assume $R = 2.6$, immunity 10% and probability of detection 50%.

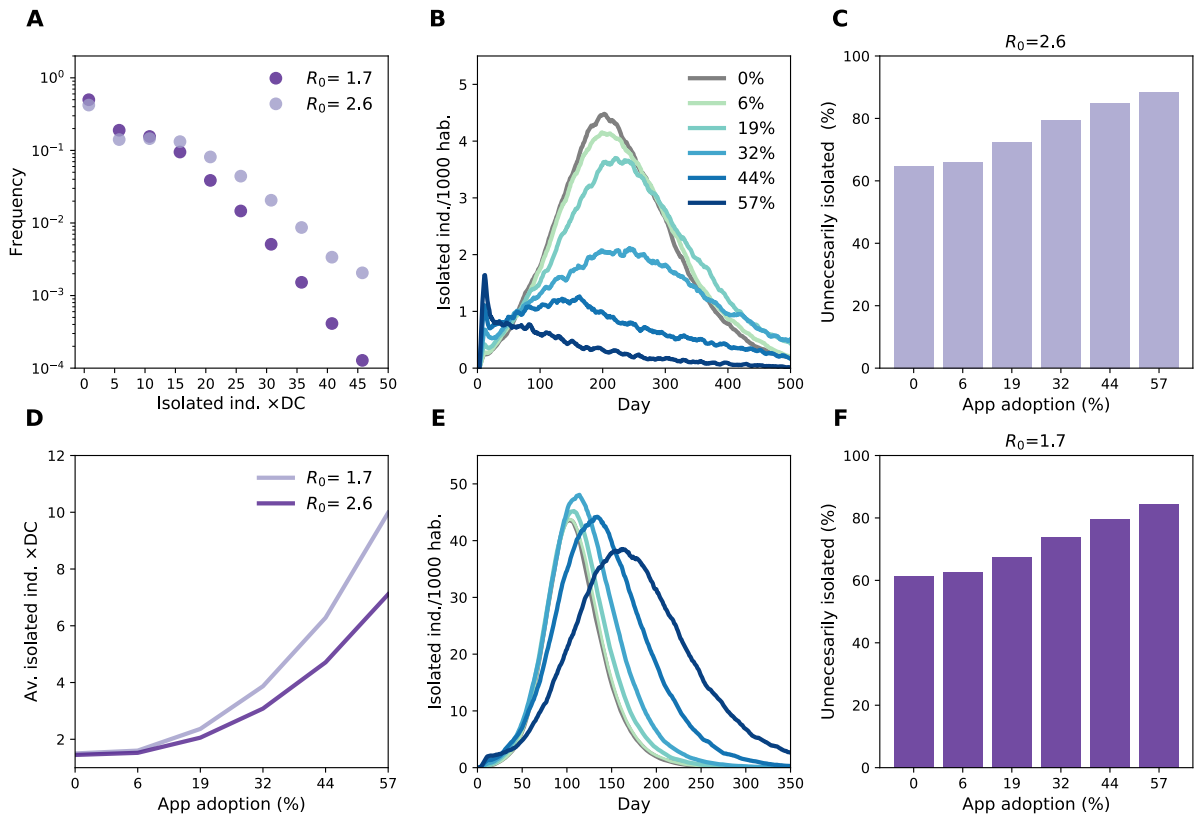

Figure 5 impact of combined digital contact tracing and household isolation on the isolation of individuals. **A** Distribution of the number of isolated individuals per detected case (DC) for 57% of app adoption. **D** Average number of isolated individuals per detected case. **B, E** Percentage of the population isolated as a function of time for $R = 1.7$ (B) $R = 2.6$ (E). **C, F**: Fraction of unnecessary isolated, i.e. fraction of contacts isolated without being positive.

Table 1. Compartmental model parameters and their values

Parameter	Description	Values	Source
IP	Incubation period	5.2 days	[26]
ϵ	Rate of becoming infectious for exposed individuals	(2.9 days)$^{-1}$	$IP - \mu_p^{-1}$
μ_p	Rate of developing symptoms for pre-symptomatic individuals	(2.3 days)$^{-1}$	[27]
μ	Recovery rate	(7 days)$^{-1}$	[27]
β_I	transmissibility rescaling according to the infectious stage	0.51 for $I_{p,sc}$, I_{sc} 1 for $I_{p,c}$, I_c	[1]

ω_s	Transmission risk by layer	1 for H layer 0.3 for C layer 0.5 otherwise	[17]
β	Transmission rate	Explored between 0.1 to 0.25	
σ_A	susceptibility	0.23 for A in $[0,14]$ 0.68 for A in $[15,64]$ 1 for A in $65+$	[7]
p_{sc}^A	Proportion of subclinical cases	0.27 for A in [0,1] 0.48 for A in [2,6] 0.57 for A in [7,19] 0.43 for A in [20,29] 0.38 for A in [30,39] 0.30 for A in [40,49] 0.24 for A in [50,59] 0.15 for A in [60,69] 0.11 for A in [70,79] 0.12 for A in [80,89] 0.26 for A in 90+	[10]

Supplementary Material

Additional methods

Algorithms for the generation of the synthetic population

The generation of the synthetic population is a stochastic process resulting in contact networks being slightly different in each run. Thanks to this mechanism, we can account for some of the uncertainty concerning the input data. It also allows for population scaling (with scaling factor of 10) to reduce the population by respecting its composition and spatial distribution, thus increasing computational efficiency. Specifically, the scaling decreases the

number of individuals in municipalities and the fluxes of commuters between them, but it does not impact the number of municipalities nor the number of schools and workplaces. The smaller population has a smaller number of households, but it maintains the statistics regarding family size and age structure given by the INSEE data.

Households

Census data on age structure and household type and size are used to randomly assign age and locate individuals in households. Five different types of household are considered: single person, single with children, couple without children, couple with children, other household groups; some of the household types may also contain an additional adult member (usually an elderly person or a relative: if the number of additional adults is greater than one, the household falls in the "other" category). For each municipality *m* with population size pop_m, we generate new households until the size of the virtual population of the municipality vir_m reaches the real size of population pop_m. For each household, we determine its type, its size and the presence of an additional member (if the household type is not single-person or couple without children, and in case of household with children if the size is greater than the number of adults plus one): then according to the role of each individual (adult, child or other) we randomly extract his/her age, with some additional conditions:

C1: the age of any child is between 15 and 45 years less than that of the youngest parent;

C2: spouses' ages differ by no more than 15 years.

The detailed procedure is summarised in Algorithm 1.

Algorithm S1: Creation of households in municipalities

for *each municipality* **do**

> $pop_m \leftarrow$ number of individuals in *m;*
>
> $vir_m \leftarrow 0;$
>
> ***while*** $vir_m < pop_m$ ***do***

determine household type $t \sim M(p_T)$, where $M(p_T)$ is a multinomial distribution with probabilities p_T given in the first column of Table 1;

determine age class of household head $c \sim M(p_C(t))$, where $M(p_C(t))$ is a multinomial distribution with probabilities $p_C(t)$ given in Table 2;

determine if the household contains an additional member according to probabilities $p_E(t)$ given in the second column of Table 1;

determine age of household head $a \sim M(p_A(c))$, where $M(p_A(c))$ is the multinomial distribution of the French age structure in the interval c;

for *each other member of household* **do**

determine the role r of the member (other adult in couple, children of couple; children of single, other);

determine age class $c_m \sim M(p_{Cm}(r))$, where $M(p_{Cm}(r))$ is a multinomial distribution with probabilities $p_{Cm}(r)$ given in Table 4;

determine the exact age $a_m \sim M(p_A(c_m))$ with the additional constraints C1 and C2;

end

$vir_m \leftarrow vir_m + s$

end

End

Table S1 For each household type, the fist column shows its frequency; the second column the probability that the household type contains an additional member; the third column the frequency of individuals in each household type

Type of household	Frequency	Fr. of add. member	Fr. of individuals

Single without children	0.338	0.0	0.149
Couple without children	0.268	0.0204	0.243
Couple with children	0.277	0.0096	0.475
Single with children	0.09	0.0123	0.106
Other	0.027	0.0	0.027

Table S2 For each household type, the frequency of the age class of the household head

Type of household	0 - 14	15 - 19	20 - 24	25 - 39	40 - 54	55 - 64	65 - 79	80 - 100
Single without children	0.0	0.019	0.075	0.191	0.186	0.162	0.212	0.155
Couple without children	0.0	0.029	0.132	0.263	0.264	0.133	0.094	0.085
Couple with children	0.0	0.002	0.023	0.271	0.47	0.116	0.073	0.045
Single with children	0.0	0.001	0.021	0.251	0.313	0.197	0.17	0.047
Other	0.0	0.008	0.042	0.233	0.283	0.176	0.173	0.085

Table S3 Probability distributions of the size of the household (except singles and couples without children, having size 1 and 2, respectively) for each age class of the household head. Rows are the age class of the household head, while columns are the size of the household

Age class	size: 1	size: 2	size: 3	size: 4	size: 5	size: 6
0 - 19	0.0	0.3386	0.4432	0.1169	0.0479	0.0534
20 - 24	0.0	0.1247	0.6325	0.1732	0.044	0.0256
25 - 29	0.0	0.169	0.5217	0.2291	0.0585	0.0217
30 - 34	0.0	0.1327	0.3919	0.343	0.1008	0.0316
35 - 39	0.0	0.106	0.273	0.411	0.1594	0.0506

40 - 44	0.0	0.1087	0.2546	0.3984	0.1767	0.0616
45 - 49	0.0	0.1123	0.3323	0.3568	0.1434	0.0552
50 - 54	0.0	0.1021	0.4621	0.2887	0.1005	0.0466
55 - 59	0.0	0.0997	0.5513	0.2269	0.0763	0.0458
60 - 64	0.0	0.1119	0.5888	0.1884	0.0657	0.0452
65 - 69	0.0	0.2312	0.5307	0.1469	0.0526	0.0386
70 - 74	0.0	0.3896	0.4509	0.1026	0.0349	0.022
75 - 79	0.0	0.5679	0.3392	0.0643	0.0183	0.0103
80 - 100	0.0	0.7572	0.1956	0.0335	0.0087	0.005

Table S4 Probability of age class of individuals depending on his role in the household, excluding the household head

Age class	Child of couple	Child of a single	Adult of couple with children	Adult of couple without children	Other
0 - 4	0.2384	0.1288	0.0	0.0	0.0161
5 - 9	0.2264	0.181	0.0	0.0	0.0176
10 - 14	0.212	0.2125	0.0	0.0	0.0226
15 - 19	0.1777	0.2131	0.0035	0.0008	0.0825
20 - 24	0.0886	0.1144	0.0445	0.0164	0.1587
25 - 29	0.0305	0.0448	0.071	0.0718	0.1051
30 - 34	0.0105	0.0204	0.0414	0.1392	0.066
35 - 39	0.0064	0.0182	0.0244	0.1848	0.0576
40 - 44	0.0045	0.0181	0.0224	0.1873	0.0577
45 - 49	0.0028	0.0172	0.0434	0.1642	0.0578

50 - 54	0.0014	0.0139	0.0902	0.111	0.057
55 - 59	0.0006	0.0101	0.1379	0.0621	0.057
60 - 64	0.0002	0.0053	0.1537	0.0306	0.0506
65 - 69	0.0	0.0016	0.1126	0.0135	0.0364
70 - 74	0.0	0.0005	0.0993	0.0086	0.0356
75 - 79	0.0	0.0001	0.0812	0.0056	0.0382
80 - 100	0.0	0.0	0.0745	0.0041	0.0835

Employment

School and industry census data are used to randomly assign an employment category to each individual on the basis of age: the probabilities are reported in Table S. The table assumes that all the children attend school from elementary to high school, while the attendance of universities is taken from census data. Census data contains the number of individuals commuting from one municipality to another one, specifying if they are students or workers, and the same for the number of individuals that study or work in the same municipality where they live. Such information is used to randomly assign students and workers to a municipality of work, which may be different from the one where they live, and then to a random school building or workplace inside this municipality. Students are assigned to a random school building, while workers are assigned to a random workplace type (5 types, depending on the workplace size, i.e., the number of employees) and then to a random workplace building. Students are not grouped in classes.

Table S5 Probability of workplace kind by age: individuals in the last column stay at home

Age class	creche	maternel	elementaire	college	lycee	universite	work	neet
0 - 3	0.6	-	-	-	-	-	-	0.4

4 - 6	-	1.0	-	-	-	-	-	-
7 - 11	-	-	1.0	-	-	-	-	-
12 - 15	-	-	-	1.0	-	-	-	-
16 - 18	-	-	-	-	1.0	-	-	-
19 - 21	-	-	-	-	-	0.365	0.448	0.187
22 - 23	-	-	-	-	-	0.332	0.471	0.197
24 - 26	-	-	-	-	-	0.026	0.687	0.287
27 - 62	-	-	-	-	-	-	0.705	0.295
63 - 100	-	-	-	-	-	-	-	1.0

Face-to-face contact network

Acquaintance network

For each place, i – a workplace (W layer), a school (S layer) and a municipality (C and T layers) – we build an Erdős–Rényi network, with average degree χ_i. The latter is a stochastic variable and depends on place layer, s_i, and size, n_i. We draw it from a gamma distribution with average $\overline{\chi(s,n)}$ and coefficient of variation CV.

We expect that when the size of a place is small each individual enters in contact with everybody. As the size increases the number of contacts saturates. We model this by assuming $\overline{\chi(s,n)} = \frac{w_s\,(n-1)}{w_s+(n-1)}$. The function approaches $(n-1)$ for small n and saturates to w_s as n increases (Figure S6).

For each setting the parameter w_s is tuned to reproduce the overall proportion of contacts occurring in the layer. CV rules the level of heterogeneity among places of the same kind and size. For simplicity we assume it to be the same for all settings. Additional details on the parametrization are provided below.

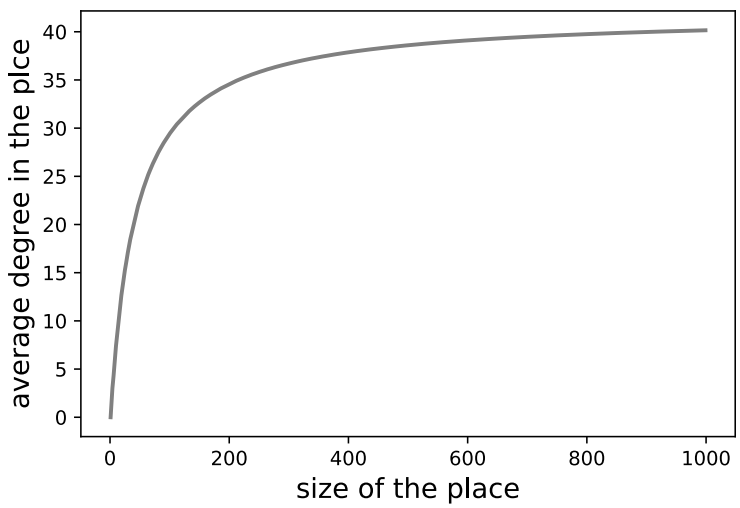

Figure S6 Average degree of the acquaintance network $\overline{\chi(s,n)}$ as a function of the size of the place. As the size goes to infinity the degree saturates to w_s that depends on the setting. Here we show as an example the curve for the workplace ($w_W = 41.8$). The other parameters estimated are $w_S = 18{,}23$, $w_C = 4{,}3$, $w_T = 20{,}9$.

Daily contact network

Once the acquaintance network is built a daily activation rate x is assigned to each link according to a cumulative distribution $F_s(x)$ that depends on the layer s. We model $F_s(x)$ with a sigmoid function with two parameters, A_s and B_s. For simplicity we assume $F_s(x)$ to be the same for $s = W, S, C$, while we allow it to be different in households (where contacts are more frequent) and in transports (where contacts are sporadic). On average, a fraction $\langle x \rangle_s$ of the links of the acquaintance network is active each day. By indicating with K_s and k_s the average degree of a layer in the acquaintance and daily networks, respectively, we have $\langle x \rangle_s = \frac{k_s}{K_s}$.

Parametrization

We tuned parameters w_s, A_s, B_s and CV to reproduce the contact statistics in [25], namely:

- the average daily number of contacts is 10;
- the contact distribution is skewed with mode 3;

- being k_s the average daily number of contacts in setting s and $r_s = \frac{k_s}{k_H}$, the survey reports $r_W = 3.17$, $r_S = 1.55$, $r_C = 1.86$, and $r_T = 0.23$.

35% of the registered contacts were with people met every day, while the rest with people met less frequently.

Specifically, these data provide the following constraints:

- Combining point 1) and 2) above we get $k_H(1 + r_W + r_S + r_C) = 10$, meaning $k_H = 1.28$. The household statistics used for our synthetic population reconstruction yields $K_H = 1.97$. This implies $\langle x \rangle_H = 0.65$.
- We assume that the daily contact network has 35% of links with activity rate >0.95. In order to do so we must account for the fact that being $f(x)$ the distribution of activation rate values assigned to links of the acquaintance network, the distribution of links sampled in the daily network is biased toward higher rates, i.e. it is given by $\frac{x}{\langle x \rangle} f(x)$.

We assume uncommon but still possible to meet more than once people in transports within a time frame of one or few months. Thus, we assume an activation rate for links in the transport layer as high as few percent or lower. Based on these constraints we design the frequency distribution as in Figure S2.

Once $F_s(x)$ is parametrized we tune the parameters w_s of the acquaintance network to reproduce the proportion r_s of daily contacts in different settings. We then fix $CV = 0.2$ to reproduce the mode of the distribution. The main properties of the network (contact distribution, link activation frequency, and age contact matrix) are shown in Figure S1 of the main paper. Other features are summaries in table S1.

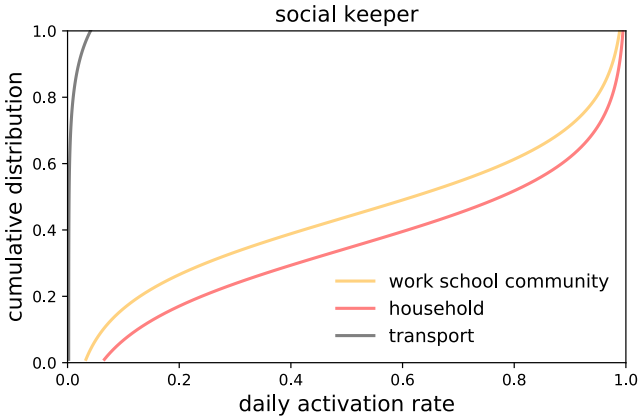

Figure S2 Cumulative distribution of the daily activation rate of contacts. We model it with the function $F(x) = A\tanh^{-1}(1-2x)+B$. Parameters are the following: $A = 0.25$ for all settings; $B_H = 0.65$,$B_T = -0.40$, $B_s = 0.56$ $(s = S, W, C)$.

Table S6 Main network features

Parameter	Description	Value
$\langle x\rangle_s$, $s = S, W, C$	Average activation rate of contacts for the school, workplace, community layer	0.56
$\langle x\rangle_H$	Average activation rate of contacts for the household layer	0.65
$\langle x\rangle_T$	Average frequency of contacts for the transport layer	0.007
k_H (K_H)	Average degree of the daily network (acquaintance network) in a household	1.79 (2.31)
k_W (K_W)	Average degree of the daily network (acquaintance network) in a workplace	12.9 (22)
k_S (K_S)	Average degree of the daily network (acquaintance network) in a school	7.9 (14)

k_C (K_C)	Average degree of the daily network (acquaintance network) in a community	2.6 (4.3)
k_T (K_T)	Average degree of the daily network (acquaintance network) in transport	1.1 (21)

Details of the numerical simulations

Simulations are discrete time and stochastic. At each time step, corresponding to one day, three processes occur: *(i)* the contact network is sampled according to the activation rate of each link; *(ii)* for each node, the infectious status is updated; and *(iii)* cases and contacts are isolated, or get out from isolation.

The transmission process is modelled through the links of the multi-layer network as follows. At each time step, a susceptible node i gets infected with the following probability

$$\Lambda_{\mathrm{i}} = 1 - \left(\prod_s \prod_{j \in v_s} (1 - \sigma_{A,i} \beta \, \beta_{I,j} \, \omega_s \, \delta_j \,) \right),$$

where j is a node belonging to the neighborhood v_s of i on layer s, and δ_j is 1 if j belongs to any infectious stage $(I_{p,sc}, I_{p,c}, I_{sc}, I_c)$ and 0 otherwise. Links of layer s have weight ω_s that represent the average level of risk associated to contacts occurring in the setting s. We assume that individuals in the I_c state stay at home due to illness, therefore they can infect only through the links of the household layer.

Simulations are run on the synthetic municipality of Strasbourg. The population is reduced of a factor 3 to reduce computational time to feasible levels, yielding to a population of 92,423 individuals.

A single-run simulation is executed with no modelled intervention, until the desired immunity level is reached. This guarantees that immune individuals are realistically clustered in space.

Then the simulations of the epidemic with contact tracing are started, considering 15 individuals initially infected randomly assigned in the population.

We vary COVID-19 transmission potential by tuning the daily transmission rate per contact β. The reproductive number R is computed numerically as the average number of infections each infected individual generates throughout its infectious period. To do so, population Immunity at the beginning of the simulation is set to 0 and R is computed considering the infections generated by individuals who get infected the first two time steps of the simulations to guarantee that the whole population is susceptible. We find that $\beta = 0.1, 0.125, 0.15, \ldots, 0.25$ corresponds to the following R values: 1.47 95% CI [0.0, 4.86], 1.75 95% CI [0.0, 5.78], 2.05 95% CI [0.0, 5.31], 2.25 95% CI [0.0, 5.31], 2.61 95% CI [0.0, 6.13], 2.95 95% CI [0.06, 7.29], 3.09 95% CI [0.68, 8.11]. We also compute numerically the generation time from the infector-infected pairs.

Additional results

Uncontrolled epidemic

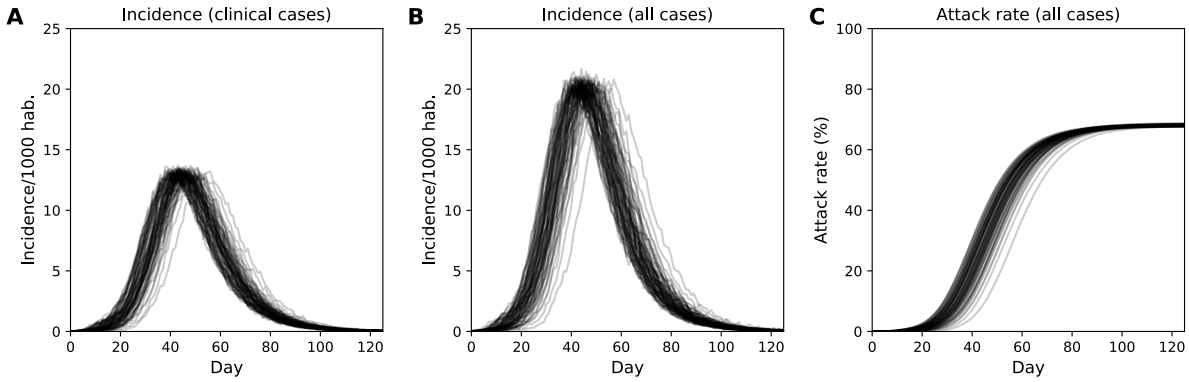

Figure S7 Epidemic in an uncontrolled scenario. **A** Incidence of clinical cases. **B** Incidence of all cases. **C** Attack rate. The bundle of curves shows 70 stochastic realizations. The epidemic is obtained with transmission rate $\beta = 0.25$ corresponding to R_0 = 3.1 (see **Error! Reference source not found.**).

App adoption variable by age

The model accounted for age-varying smartphone penetration. However, we assumed that the probability of downloading the app, provided an individual owns a smartphone, is uniform and independent on age – uniform app adoption scenario, U. This hypothesis is simplified. Indeed, elderly people may be less incline to use the app even when they own a smartphone. We also tested the extreme case scenario in which no individual in the 70+ age cohort adopts the app – non-uniform app adoption scenario, NU. We consider a 32% app coverage over the whole population, and we compared U and NU scenarios, assuming the same number of apps are downloaded in the two cases. Figure S4 shows the attack rate relative reduction $ARR = \frac{AR_U^A - AR_{NU}^A}{AR_U^A}$ by age group, where AR_{sce}^A is the attack rate of the epidemic for the age group, A, and the scenario, $sce = U, NU$. We found that ARR is close to zero, meaning that $AR_{NU}^A \simeq AR_U^A$, for all age groups. This means that distributing the app only to individuals younger than 70 years would not reduce the protection in the 70+ age group.

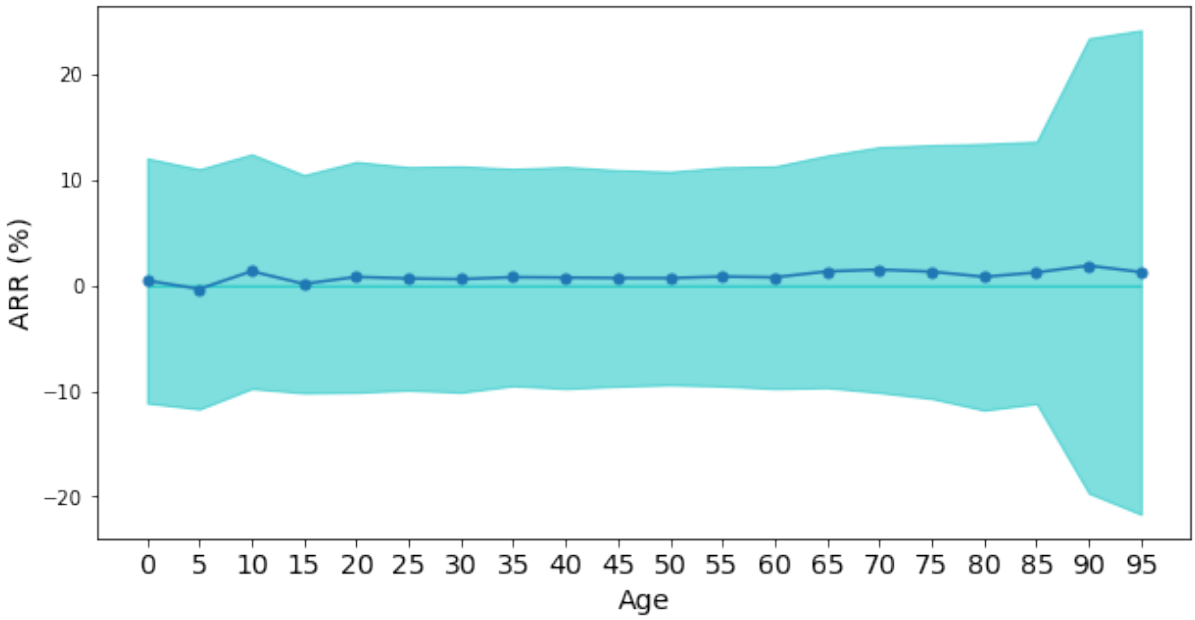

Figure S4 Comparison between the U and the NU contact tracing scenario. Here $R = 2.6$, Immunity is 10%, detection probability is 50% and app penetration is 32%. The line shows the average attack rate relative reduction and the shaded area is standard deviation.