

14 DECEMBER 2020

# SARS-CoV-2: Where do people acquire infection and ‘who infects whom’?

This rapid review of the science of infection and COVID-19 from the Royal Society is provided to assist SAGE in relation to COVID-19.

This paper is a pre-print and has been subject to formal peer-review.

## Executive summary

Where people acquire infection depends on many factors, including changes over time induced by social distancing measures (non-pharmaceutical interventions (NPIs)) in place at any point in time, prevailing compliance patterns to such measures, variation between geographical locations, and confounding variables such as age, gender, and cultural differences between social or ethnic groups within a defined population.

The available UK data suggests that most transmission occurs in household settings, with frequency related to social interactions within households. However, where most people acquire infection, may not be the most important issue in terms of controlling spread. Key factors in limiting onward transmission are who seeds the household setting, and where this index case acquired infection?

Data sources for gaining a better understanding of where people acquire SARS-CoV-2 infection are contact tracing, household studies and whole genome sequencing (WGS) of the virus, to ascertain who infects whom. Data sources and publications on these topics are limited at present in the UK.

Data are limited in the UK on where index cases in households acquire infection, but the information collated by Public Health England (PHE) suggest that much heterogeneity pertains, depending on a variety of social and demographic factors. Activity related to shopping and social gatherings in restaurants and pubs generate the most infections, although many other factors have influence. Given data limitations – this observation should be treated with caution.

## The role of test, trace and isolate in the SARS-CoV-2 epidemic

The test, trace and isolate (TTI) system as a tool to control spread, and as a source of epidemiological information, has limitations when viral spread is extensive given the enormous workload involved in tracing contacts. It is best employed to control transmission during phases when rates of infection are low, as well illustrated in many East Asian countries.

An illustration of the challenges facing TTI in the UK is provided by data collected in the first week of November 2020. Of those estimated by the Office of National Statistics (ONS) to have been infected that week, only 30% were reached to ask for information on close contacts, and only 18% of the estimated close contacts of those newly infected that week were asked to self-isolate. Perhaps only 9% of these contacts did so fully, due to reported poor compliance issues in certain demographic and social groups. These numbers are clearly far too low to have a major impact on viral transmission or to provide a representative national picture of spread and control measure impact.

The TTI system can play an important role as vaccination coverage rises in 2021 and attention is focused on eliminating smaller outbreaks.

Public Health England (PHE) has made TTI data available to researchers, but there is a need to focus more resources to help summarise tracing information on where people acquire infection week by week, so the impact of various social distancing measures region by region can be continuously monitored and appropriate changes to policy made if patterns change.

### Other routes to collect transmission information

A large amount of whole genome sequencing (WGS) data has been collected in the UK (over 100,000 COVID-19 genomes sequenced as recorded on COG UK [COVID-19 Genomics UK consortium] which serves as a key resource to monitor viral evolution relevant to diagnostics and vaccine development), but at present, with a few exceptions, little use has been made of this data base to ascertain who infects whom outside of settings such as hospitals or care homes.

Molecular epidemiological studies involving WGS should be commissioned in a variety of settings (e.g. household, on transport, workplace, regions, cities, villages), to use viral isolates collected at testing stations for diagnostic purposes, with appropriate completion of questionnaires to capture demographic, social and home setting information, to link with WGS information, and concomitantly assess who is infected whom.

The value of sequence data from care home staff and residents is shown by the identification of multiple independent introductions of SARS-CoV-2 into each care home by both family members, friends and staff, rather than a single introduction followed by within-care home transmission.

Ascertaining the settings that generate index cases with any certainty, requires both questionnaire studies associated with contact tracing, and linked WGS information.

Policy makers may wish to consider the respective value of information that is generated by national data collection (e.g. the TTI system), versus that arising from directed small-scale research studies in defined settings of who infects whom, supported by local authorities and research funding bodies. Funding is understandably mainly targeted at the former. Both are important, but too little attention has been directed to detail local, contact tracing, household and WGS studies within UK settings. Such studies should be encouraged as a matter of urgency.

The paucity of local contact tracing studies with clearly defined aims, is needed, working with research councils, in partnership with local public health authorities. Local knowledge is important, given great heterogeneity in where people acquire infection by region, social grouping, income level and ethnicity.

### Introduction

Where people acquire infection, and who infects whom, are important issues for policy makers when considering the design of social distancing measures (the so-called non-pharmaceutical interventions (NPIs)) to control the spread of SARS-CoV-2. This is particularly important when no other interventions to limit transmission are available as is the case at the date of this report. This short review examines publications and data from the UK, and other countries, that throw light on what is currently understood about these questions.

In the analysis of the available data, it is important to recognise that where people acquire infection depends on many factors. These include changes over time influenced by social distancing measures in place at any point in time, prevailing compliance patterns to such measures, variation between geographical locations, and confounding variables such as age, gender and cultural differences between social or ethnic groups within a defined population. As such, it may not be possible to say that a given fixed percentage of transmission events occur within households, or on transport, or in restaurants or pubs, since values will change over time given prevailing NPIs, public attitudes reflected in compliance to such measures, and geographical plus cultural variability. At any point in time a great deal of heterogeneity will therefore prevail in any measures of the frequency of transmission by setting.

A clear example of how changing rules and regulations about moving and mixing under various tiers of constraints and social distancing rules in the UK in different locations is provided by Google mobility data.<sup>1</sup> Data from Google's Community Mobility Reports demonstrate the heterogeneity in activities of people by region of the UK (although it should be noted that categorisation of places varies from region to region, which can limit comparability). Appendix 1 Figures A1 and A2 represent how visits and length of stay at different places change in the last three months up to November 2020, compared to a baseline defined as January to February 2020. Cornwall, as a popular tourist location, would be subject to high seasonal variation in these activities, but London, also popular with tourists, has a strikingly different pattern, with fewer people visiting central London during the first wave of infection. All regions show considerably and consistently more time at home and less time at workplaces than in January and the impacts of the second lockdown (from 5th November) on activities, particularly public transport and retail and recreation, are beginning to be detectable. These differences in people's activities: where they go, who they mix with and for how long, vary by location and over time. Region and date of data collection must therefore be borne in mind

when interpreting evidence from studies of SARS-CoV-2 transmission. For example, a well-designed study of where transmission takes place in Germany provides excellent insights at the time of data collection, reporting that little was happening within schools<sup>2</sup>. However, the study was conducted when schools were closed during a lockdown period.

Current understanding of COVID-19 transmission suggests that where most people acquire infection, such as, for example, within a household setting, may not be the most important issue in terms of control by social distancing measures. While guidelines have been produced recommending steps that everyone can take to reduce household transmission<sup>4</sup> who seeds the household setting, and where this index case acquired infection, is the more important questions. Two recently published reviews addressing transmission in households were released in November 2020. The first report details a meta-analysis of 45 peer reviewed publications from many different countries, deriving secondary attack rates (SAR) in household settings<sup>3</sup>. They find moderate evidence for less transmission both from and to individuals under 20 years of age in the household context, but this difference is less evident when examining all settings. They conclude that there were limited data to allow exploration of transmission patterns in settings other than households. The second report is from the Environmental and Modelling Group (EMG) and the Scientific Pandemic Influenza Group on Behaviours (SPI-B) and focuses on mitigating risks of COVID-19 transmission associated with household social interactions<sup>4</sup>. The report provides evidence that social interactions in the home increases the risk of infection, and discusses ways to reduce this risk.

This review examines epidemiological data on transmission and describes published studies, both in the UK and in other countries, where they have provided insights into where most transmission takes place. The concluding section focuses on what is known based on UK data and a discussion of future data collection priorities and research needs. An extensive literature search was conducted to form the template for this review, but it is not exhaustive given the volume of publications in this area and the large number of preprints on medRxiv.org. Selected publications that contain data, and brief comments on what the publication reports, are listed in a set of Tables placed in appendices.

## Methods

Epidemiologists employ three key methods in the study of transmission from person to person. They are household studies, contact tracing, and whole genome sequencing (WGS) of viral isolates (to ascertain who infects whom).

The first two approaches have been employed widely in attempts to study patterns of viral transmission, starting early in the 1940s with the work of Hope-Simpson who followed transmission in families of the common communicable diseases of children<sup>5,6</sup>. These studies sought to establish the incubation and infectious periods of defined viral infections, such as the measles virus<sup>5,7-9</sup>. Household studies can be particularly informative about transmission between different age classes (children and adults) and have been employed for many decades. Contact tracing in public health is the process of identification of persons who may have come into contact with an infected person ('contacts') and subsequent collection of further information about these contacts. Contact tracing has been a pillar of communicable disease control in public health for decades. The eradication of smallpox, for example, was achieved not only by universal immunisation, but also by exhaustive contact tracing to find all infected persons<sup>10,11</sup>. This was followed by isolation of infected individuals and immunisation of the surrounding community and contacts at-risk of contracting smallpox. Eradication programmes for smallpox in Pakistan and India also utilised paying bounties to people who reported on those who they thought might be infected. Contact tracing has been widely used in the control of sexually transmitted infections for many decades<sup>12</sup>. As vaccination is introduced to protect against COVID-19 infection, and coverage rises over the coming few years, mopping up pockets of infection or small outbreaks will depend on this approach.

Sequencing is employed to assess the degree or relatedness between viral isolates taken from infected people, such that various statistical tools can be employed to ascertain degrees of relatedness between sequences<sup>13-15</sup> and hence infer who has infected whom. Despite the huge volume of viral samples collected and the relatively small size of the viral RNA genome of SARS-CoV-2 compared to many infectious agents, surprisingly few such studies have taken place, either in the UK or elsewhere. It is important to note, however, that Coronaviruses are the largest single-stranded RNA viruses. Physical size and genome size are highly correlated for parasitic organisms. Large size may be associated with short lived immunity since many large parasitic organisms modulate the human immune system. Whether this is the case for Coronaviruses is unknown at present. WGS methods have been very effectively applied to other viruses, such as HIV-1, in charting both viral evolution and patterns of spread<sup>16,17</sup>.

In consideration of all three methods, it is important to note that the generation of secondary cases from one infectious person in virtually all settings is a very heterogeneous process. The distribution of the generation of secondary cases is typically highly overdispersed (= aggregated), with a variance much greater than the mean value. Most infect none, or a few, and a few infect many. The latter are sometimes referred to as superspreading events and the source or index infectious person as a super spreader<sup>18</sup>. Contact tracing data provide crucial insights about this distribution.

### Who acquires infection from whom (WAIFW matrices)

In the study of infectious disease spread it has long been realised that a quantitative understanding of the transmission dynamics of directly transmitted infections such as respiratory viral infections, depends on a detailed understanding of who mixes with whom. The so-called WAIFW (who acquires infection from whom) matrices of mixing by age have long been used to refine epidemiological studies of childhood viral infections and the impact of vaccination<sup>19</sup>. Such contact patterns in populations prior to the imposition of social distancing measures to control SARS-CoV-2, are a good baseline to understand how various stratifications of human communities interact. The usual variable examined based on contact tracing, contact diaries and other sources is age. Matrices can be constructed to illustrate how different age groups mix. A series of good examples from different countries are provided in the studies of Prem and colleagues<sup>20</sup>. The study used data from population-based contact diaries. Household level data from the Demographic and Health Surveys for nine lower-income countries and socio-demographic factors from several on-line databases for 152 countries were used to quantify similarity of countries to estimate contact patterns in the home, work, school and other locations for countries for which no contact data are available, accounting for demographic structure, household structure where known, and a variety of metrics including workforce participation and school enrolment. Contacts are greatest with age classes across all countries considered as recorded in Appendix 1 Figure A3. However, pronounced regional differences in the age-specific contacts at home were noticeable, with more inter-generational contacts in Asian countries than in other settings. In an ethnically diverse country such as the UK, it is to be expected that the different intergenerational mix will vary greatly between households depending in part on ethnic background and factors including metrics of wealth. Furthermore, it is not just co-residence but intergenerational proximity (households of extended families living near each other) that facilitate transmission. Household surveys of living arrangements and intergenerational contacts will provide further understanding of these dynamics<sup>21</sup>.

Poorer households may have more generations living together<sup>22</sup>, while lower-educated individuals have been shown to have closer intergenerational proximity<sup>23</sup>. There are large differences in intergenerational proximity in the UK by ethnicity and foreign- versus UK-born groups<sup>24</sup>.

We have focused on mixing matrices by age, but many more stratifications are possible and desirable in the sense that they do influence transmission. Further stratifications include social and cultural factors, ethnicity, sectors of the economy, and work occupation. To date little work has been published in this area in the context of SARS-CoV-2 transmission.

### Contact tracing – the ‘test, trace and isolate’ system in the UK

Contact tracing is a long-established methodology for the control of infectious disease spread<sup>25–27</sup>. In the past it was most widely employed in the control of sexually transmitted infections<sup>28,29</sup>. For COVID-19, contact tracing is an essential tool (along with social distancing) in trying to limit viral spread and bring the effective reproductive number to less than unity in value, given the absence of vaccines and effective treatments at present. However, it is important to note that some treatments (corticosteroids and biologicals<sup>30,31</sup>) are showing promise in reducing disease severity in those infected.

In the National Health Service (NHS), contact tracing in England has been labelled as the Test and Trace system<sup>32</sup> or by some as the ‘Test Trace and Isolate system’ (TTI), since advice is given to isolate if contact has been made with an infected person. Similar systems have been implemented in Wales (‘Test Trace Protect’<sup>33</sup>), Northern Ireland<sup>34</sup> and Scotland (‘Test and Protect’<sup>35</sup>). A summary of how the system is designed is given in Appendix 2.

### Performance of the testing system

The government has stated that the TTI system is one of the cornerstones of the effective management of the COVID-19 epidemic. The success of TTI depends on a number of key factors: namely, adherence to isolating if symptomatic, getting a test if symptomatic, passing on details of close contacts if infection is confirmed, and quarantining of contacts. Rates of adherence to TTI behaviours in the UK have frequently been questioned as the second wave develops<sup>36–39</sup>.

A recent publication on medRxiv investigates rates of adherence to the UK’s TTI system over time<sup>40</sup>. Data were collected by Smith and colleagues, a team at King’s College London, between 2 March and 5 August 2020. There were 42,127 responses from 31,787 people living in the UK, aged 16 years or over. This sample is not random, but self-selected. The main outcome measures of the overall

study are many. They include the identification of the key symptoms of COVID-19 (cough, high temperature/fever, and loss of sense of smell or taste), self-reported adherence to self-isolation if symptomatic, numbers requesting an antigen test if symptomatic, and numbers with the intention to share details of close contacts, and numbers self-reporting adherence to quarantine if informed that they had been in contact with a confirmed COVID-19 case.

Only 48.9% of participants (95% CI 48.2% to 49.7%) identified key symptoms of COVID-19. Self-reported adherence to TTI behaviours was low (self-isolation 18.2%, 95% CI 16.4% to 19.9%; requesting an antigen test 11.9%, 95% CI 10.1% to 13.8%; intention to share details of close contacts 76.1%, 95% CI 75.4% to 76.8%; quarantining 10.9%, 95% CI 7.8% to 13.9%) and largely stable over time. By contrast, intention to adhere to protective measures was much higher. Non-adherence was associated with gender (males with the poorest records), younger age groups, having a dependent child in the household, lower socio-economic group greater hardship during the pandemic, and working in a key sector. The authors conclude that government should take practical measures to encourage individuals to adhere much better to the guidelines and consider giving financial reimbursement to improve adherence. They also suggested that targeting messaging to men, younger age groups, and key workers may also be necessary to improve performance.

One key challenge in contact tracing for COVID-19 cases, is the very high fraction of asymptomatic individuals, especially in the young<sup>41, 42</sup>. Overall, it may be as high as 75% on the basis of some studies<sup>43</sup>. In these circumstances, tracing and isolating becomes even more important, since individuals have to be persuaded that they may infect others despite having no symptoms. This is, and has been, a considerable challenge in Europe and North America – but much less so in Asian societies, perhaps due to differences in culture resulting in better acceptance of guidance from government. Table A7 in Appendix 3 summarises the methods of epidemiological investigation via contact tracing undertaken in South Korea, which are unlikely to be enforceable in countries of Western Europe and North America.

Various modelling studies have been conducted on how efficient TTI needs to be to reduce the effective reproductive number at time  $t$ ,  $R_t$ , below unity in value. They all conclude that TTI has to be very efficient in terms of speed of response in giving positive test results to people. Such studies suggest it must have good adherence to isolation and effective tracing of contacts at a level well in excess of 80%<sup>44, 45</sup>. The study by Keeling and colleagues employing detailed survey information on social encounters coupled to predictive models, investigated the likely efficacy of the current UK definition of a close contact (within 2 metres for 15 minutes or more) and the distribution of secondary cases that may go untraced. Taking recent estimates for COVID-19 transmission ( $R_0$  values), they suggested that with good management less than 1 in 5 cases will generate any subsequent untraced cases, although this comes at a high logistical burden with an average of 36.1 individuals (95th percentiles 0-182) needing to be traced per case. As cases rise this imposes a burden on the TTI system that cannot be addressed – and clearly has not been addressed, for example in the UK, France and the USA during both first and second waves.

During the lull between the first and second waves of the epidemic, the performance of TTI in the UK was inadequate to keep the virus in check. The second lock down provided a further opportunity to rectify these shortcomings and move more resources to a local level where public health authorities can employ local knowledge to enhance performance.

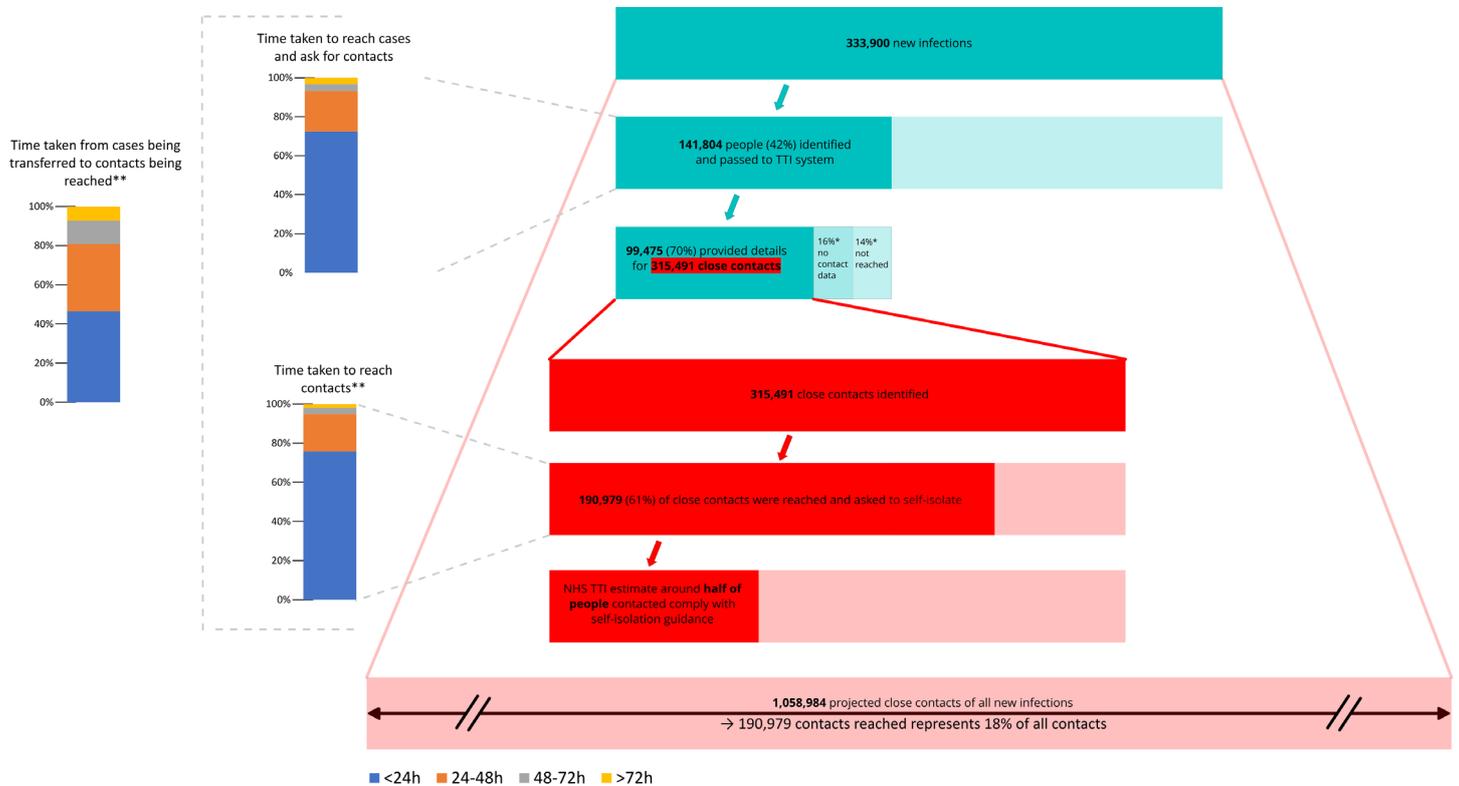
Over recent weeks, during the rising phase of the second wave in the UK, the performance of the contact tracing aspect of the system has been poor for very understandable reasons. During a period of extensive viral spread, the sheer volume of contacts requiring tracing presents huge and indeed unsurmountable challenges. Contact tracing can work well to restrict transmission when spread is low, as well illustrated in many South Asian countries Appendix 3 Table A3). An illustration of the current performance of TTI in the UK is presented in Appendix 1 Figure A4 which records number of SARS-Cov-2-infected individuals transferred to the contact tracing system in England and number of people transferred who were reached and asked to provide details of recent close contacts (includes cases managed and not managed by local protection teams (HPTs)). Appendix 1 Figure A5 provides associated information on the people transferred to the contact tracing system percentage of cases reached and asked to provide details of recent close contacts by UTLA since Test and Trace began in England stratified by region. What is less clear in much of the released information is, of the total number of those who tested positive region by region (based on ONS estimates), what fraction were reached by the TTI system?

The fraction of the cases that occur that are referred or picked up by TTI can be assessed by comparing the TTI and ONS data for a defined period of time. Such calculations are very sobering. As an example, we examined the data reported in the first week of November 2020 (Figure 1). ONS reports an estimate of 333,900 new infections in that week. Of these, 141,804 people newly infected were identified by the TTI system. Only 99,458 cases were reached and provided details of contacts (70% of cases referred), in total identifying 315,492 close contacts. This represents an average of just over 3 contacts per person. Of these identified close contacts, only 61% were reached and asked to self-isolate. If we assume that all 333,900 individuals infected during that week had the same average 3 close contacts each, then the percentage of this total who were asked to self-isolate is only 18%. This is clearly a very disturbing result. Even more so, if we take the estimate from NHS that only 50% of those asked to self-isolate actually do so. In other words, only 9% of close contacts of infected people fully self-isolated in that week of a rising second epidemic wave. This is clearly far too small a percentage to have a very significant impact on the epidemic. Furthermore, the estimate of 50% compliance may be too optimistic given data recorded by Smith and colleagues in a recent King's College study<sup>40</sup>.

Further evidence of the performance of both England and Scotland contact tracing systems is shown in Figure 2. Conducting a thorough comparative analysis between the two regions is hampered by the lack of comparable performance measures which both systems record. An exemplar of this is demonstrated by the different measures taken to record the time taken for a contact tracing team member to reach out and interview a confirmed case and record their contact history. The UK and Wales data present this as the total number of people reached within <24 hours, 24 – 48 hours, 48 – 72 hours, and >72 hours. This was manually transformed into a percentage, using the total number of cases recorded. Comparatively, Scotland has broken down the system of reaching out to their cases into index created, test taken, interview completed, case created, and case closed. Definitions of each stage are not publicly available, and our attempts to reach out to gain clarity on this ambiguity has been fruitless thus far. To create the figure above, the Scottish measure of 'test to interview completed' was selected as the best comparative measure to the UK and Wales' 'total number of people reached and asked to provide details of close contacts. The synchronisation of data format reported on key measures such as contact tracing should be essential, especially between countries within the UK, for accurate comparability of programme progress and success.

**FIGURE 1**

Performance of the NHS Test, Trace and Isolate (TTI) system for England, for the week 29 October to 4 November 2020.



\* Of those passed to the TTI system that did not provide details for their contacts, 20,128 (14%) were not reached and 22,218 (16%) were reached but provided no data on their contacts.

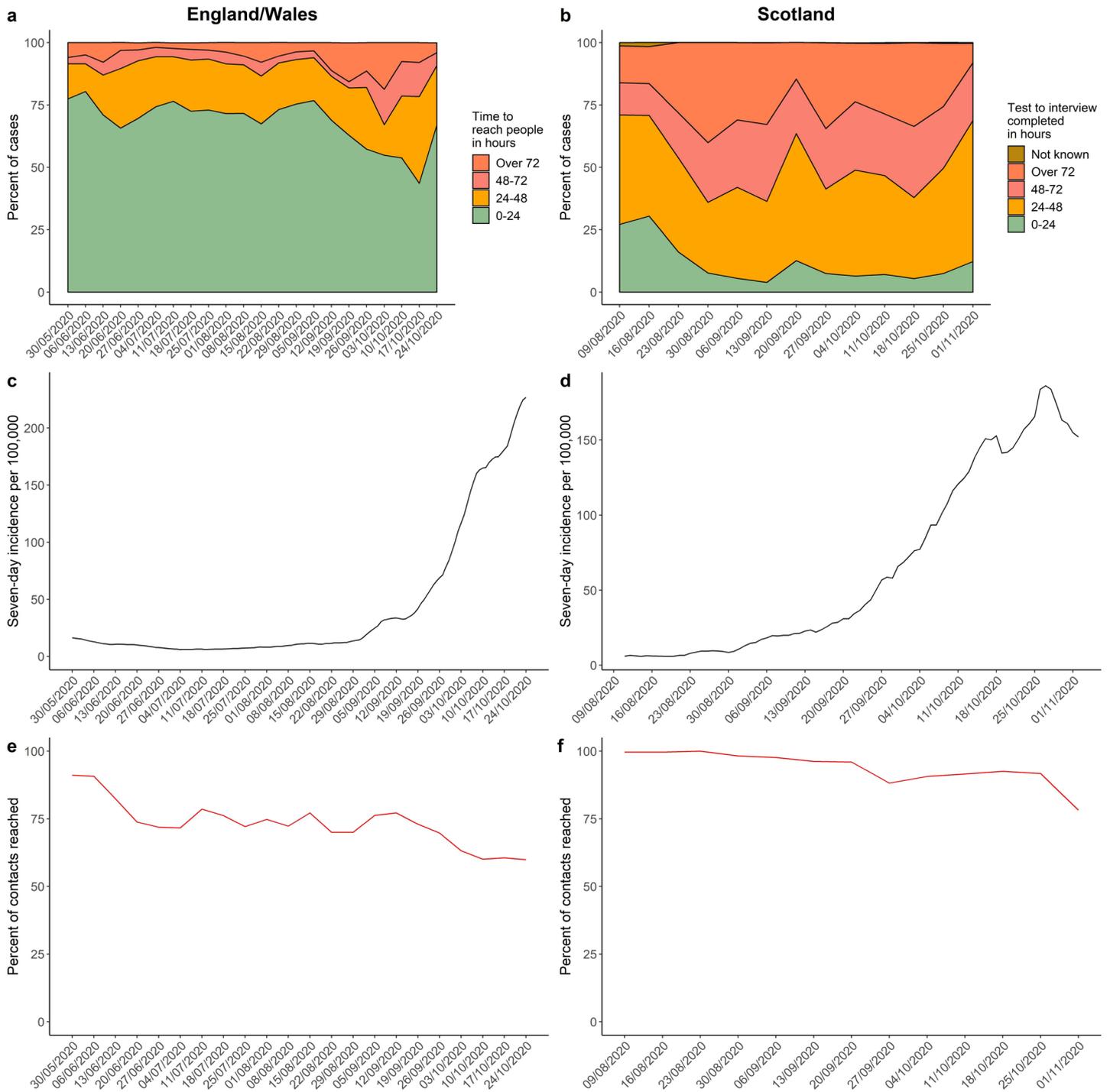
\*\* Data for time taken to reach close contacts and advise them to self-isolate are from 83% of contacts passed to the TTI system (contacts managed by local health protection teams as part of outbreak investigations are not included).

Source: Estimate of number of new infections is from the Office for National Statistics ([www.ons.gov.uk](http://www.ons.gov.uk)). Figure drawn by authors using data from NHS Test, Trace and Isolate, report published 10 December 2020<sup>46</sup>. 'Contacts' represent close contacts. Figures presented by TTI are based on a data cut several days after the end of the reporting period, to give time for cases reported at the end of the 7 day period to have an outcome. Some cases may continue to be in progress after this period, and therefore data are revised over time. Data for week 29 October to 4 November 2020 first published on 12 November were near 100% complete compared to update reported on 10th December (completion rates: number passed to TTI system=100%; number of cases providing details of contacts=99.8%; number of close contacts identified=99.8%; close contacts reached and asked to self-isolate=99.6%).

**FIGURE 2**

Comparison of England and Wales (upper row) and Scotland (lower row) national contact tracing efforts.

The time taken to contact cases in both scenarios has increased over time (a, d), as the seven-day incidence per 100,000 cases increases (b, e), highlighting the pressure placed on the contact-tracing systems (c, f). Figure drawn by authors using data from GOV.UK COVID-19 dashboard<sup>47</sup> and Public Health Scotland COVID-19 dashboard<sup>48</sup>.



## Where do people acquire infection based on contact tracing studies

This section examines the data and reports available to assess where people acquire infection based on the current TTI system managed by Public Health England. In assessing this information, it is important to keep in mind the comments made earlier in the Introduction about granularity and changes over time, in different regions and different social and cultural settings. We first focus on UK data and then turn to comparisons with other countries.

The England data recorded week by week since the start of the second wave provide information on where contacts were made with a person with a positive test for viral infection. An example of this information is presented in Figure 2 for a series of weeks running through the beginning of the second wave up to November 2020. Most contacts were in household settings, with household visitors being the second most frequent contact, closely followed by contacts made during leisure activities.

A useful comparison with this pattern is provided by a study in Germany carried out by the Robert Koch Institute, the results of which are presented in Appendix 1 Figures A6 and A7 for up to August 2020.<sup>2</sup> Note that this study was conducted during the first wave in Germany, a setting where relatively few cases were reported by comparison with other European countries.

The household was one of the two most common settings for contacts along with care homes. Schools were not important, but it is important to note that school closures were in place during the period of this study in Germany. As recorded in Table A6 in Appendix 3, other countries tend on average to record similar patterns.

We now turn to the equally important issue of – where did the index case introducing infection into a household acquire infection?

## Who seeds households (the index case) and where do they get infected?

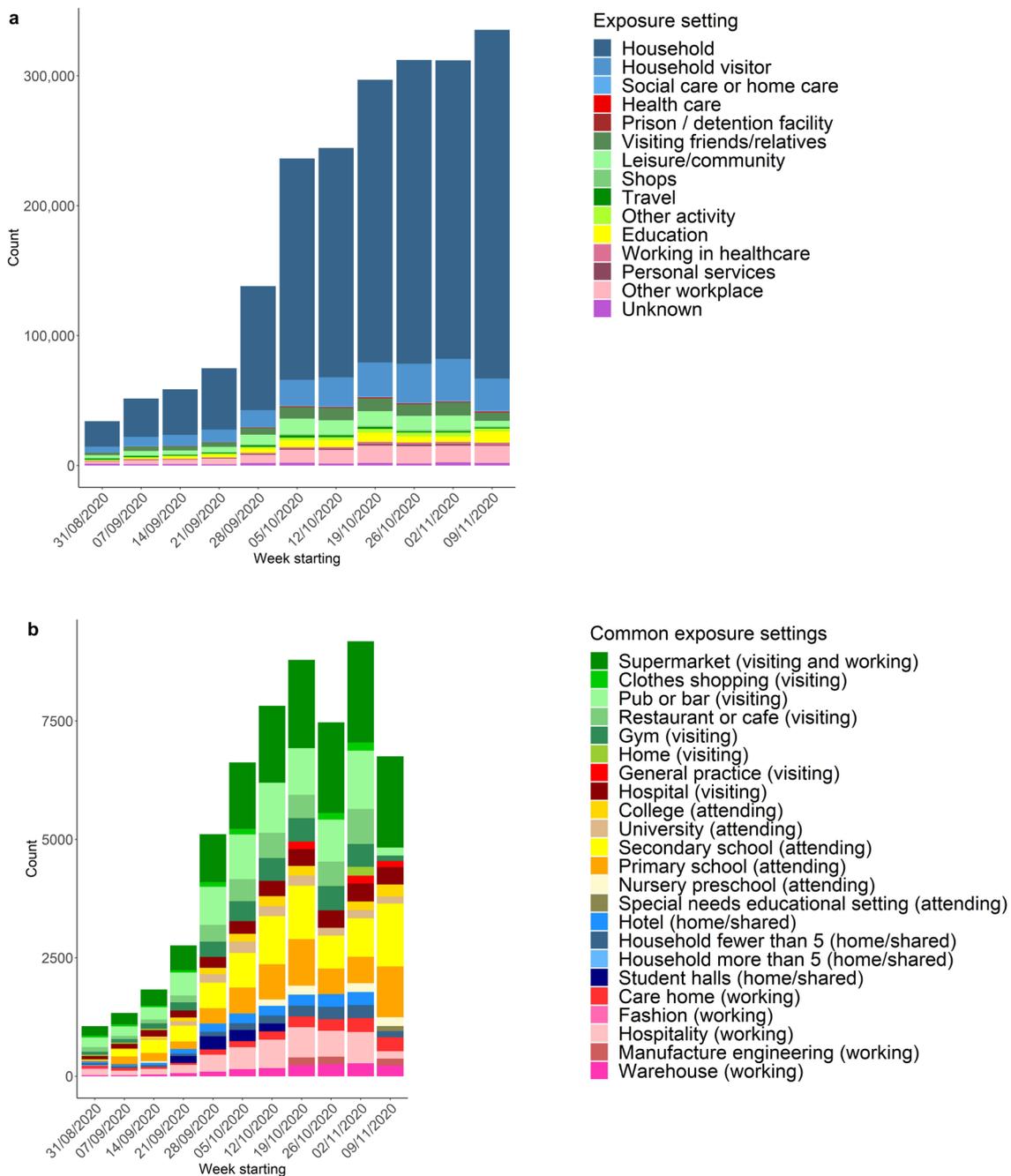
The seeding of infection in a household is typically more difficult to ascertain in the absence of molecular epidemiological studies involving whole genome sequencing (WGS), since the index case may have been in many settings in which infection could have been transmitted including transport, supermarket, work-place and leisure facility. The England TTI system does record information on locations and activities reported by people testing positive for COVID-19 and as shown by Figure 3, showing common exposure settings reported by cases over time in the England.

Figure 3a shows that the most important exposure setting is within households. Outside of households, Figure 3b and Figure 4 demonstrate that non-household exposure to SARS-CoV-2 occurs in a wide variety of settings, although these data are dependent on accurate self-reporting of cases. Phenomena such as desirability bias, and universal usage of tracking QR codes both by hospitality establishments and customers alike will impact the reliability and scope of these data. During the second UK wave, which coincided with return to school in September after the summer holidays, the distribution of exposure settings has not changed substantially over time until lockdown. Figure 3b shows very little contribution of schools as exposures in the last week of August, as expected, but relatively greater impact of education settings and essential shops (supermarkets) in the most recent week (commencing 9th November 2020), as the impact of the second lockdown is felt and other exposure settings were closed. The contribution of hospitality venues such as bars, pubs and cafes to transmission has been widely discussed, given the financial impact of COVID-19-related restrictions on the hospitality sector. Figure 3b and Figure 4a demonstrate that these venues contribute a substantial amount to exposures, but whether transmission occurs here has not been proven. A great reassurance is the small and unchanging contribution of care homes and healthcare settings to exposures.

**FIGURE 3**

Exposure settings as documented in contact tracing data shared by Public Health England, August – November 2020.

All data show categories where exposure to infected contacts occurred in England by week and do not infer transmission. a) Contacts by exposure setting. These are the settings where a person who tested positive for SARS-CoV-2 reported that they met with and potentially exposed their contacts (forward contact tracing). Work is ongoing to link contacts to future cases and to determine where transmission occurs. b) Common locations reported by people who tested positive for SARS-CoV-2. Two or more individuals who tested positive reported the same location, defined by the same postcode, as possible exposure setting in the 3 – 7 days prior to symptom onset, or test date if asymptomatic (backward contact tracing). Information on this type of event and the location are recorded but not information on contacts.



Source: Figure drawn by authors using data from Public Health England.

The prominent category 'other' in Figure 4a includes many ill-defined locations and activities. The two dominant locations are shops and restaurants (recorded as 'Shopping' and 'Eating out' in the legend of Figure 4a). Figure 4b shows that similar numbers of cases report exposure in both leisure and work settings. Figure 5 shows the proportion of common exposure settings relative to all reported common exposures over time. A common exposure occurs when two or more individuals who tested positive report the same location, defined by the same postcode, as possible exposure setting in the period three to seven days prior to symptom onset (or test date if no symptom onset is observed). Prior to the November lockdown, visiting shops and hospitality venues together constitute up to half of reported common exposure settings. Schools and education settings also make up a significant proportion. However, these data should be interpreted with caution as transmission following exposure has not been demonstrated. Precise who infects whom data based on virus whole genome sequencing are not available to confirm people's responses to a set of questions about where they have been and where they think they may have acquired infection. What is striking about the data is the wide range of possible activities and settings in which transmission could have occurred. Given transmission can take place via airborne droplets and contaminated surfaces, this is hardly surprising. Future work should ideally focus on refining the analyses of the data collected by TTI week by week, and further exploration and refinement of ways to collect data from contacts and index cases to better identify risk settings.

The results of the contact tracing data are supported by an independent data source, the Office for National Statistics Opinions and Lifestyle Survey for 11th to 15th November 2020<sup>49</sup>. The survey reported that nearly 2 in 10 (18%) adults in England stated that they were in direct physical contact with at least one other person indoors, including settings such as the home, cafés, pubs or restaurants in the last 24 hours, excluding those in their household or support bubble; a similar percentage (19%) was reported by those in tier 3, and a higher percentage was reported by those living under fewer restrictions in tier 1 (27%) and tier 2 (25%) from two weeks ago<sup>49</sup>. In addition, older age groups were more likely to have avoided physical contact, washed their hands and maintained social distancing, while nearly a quarter of young people had physical contact indoors with someone who was not part of their household or support bubble.

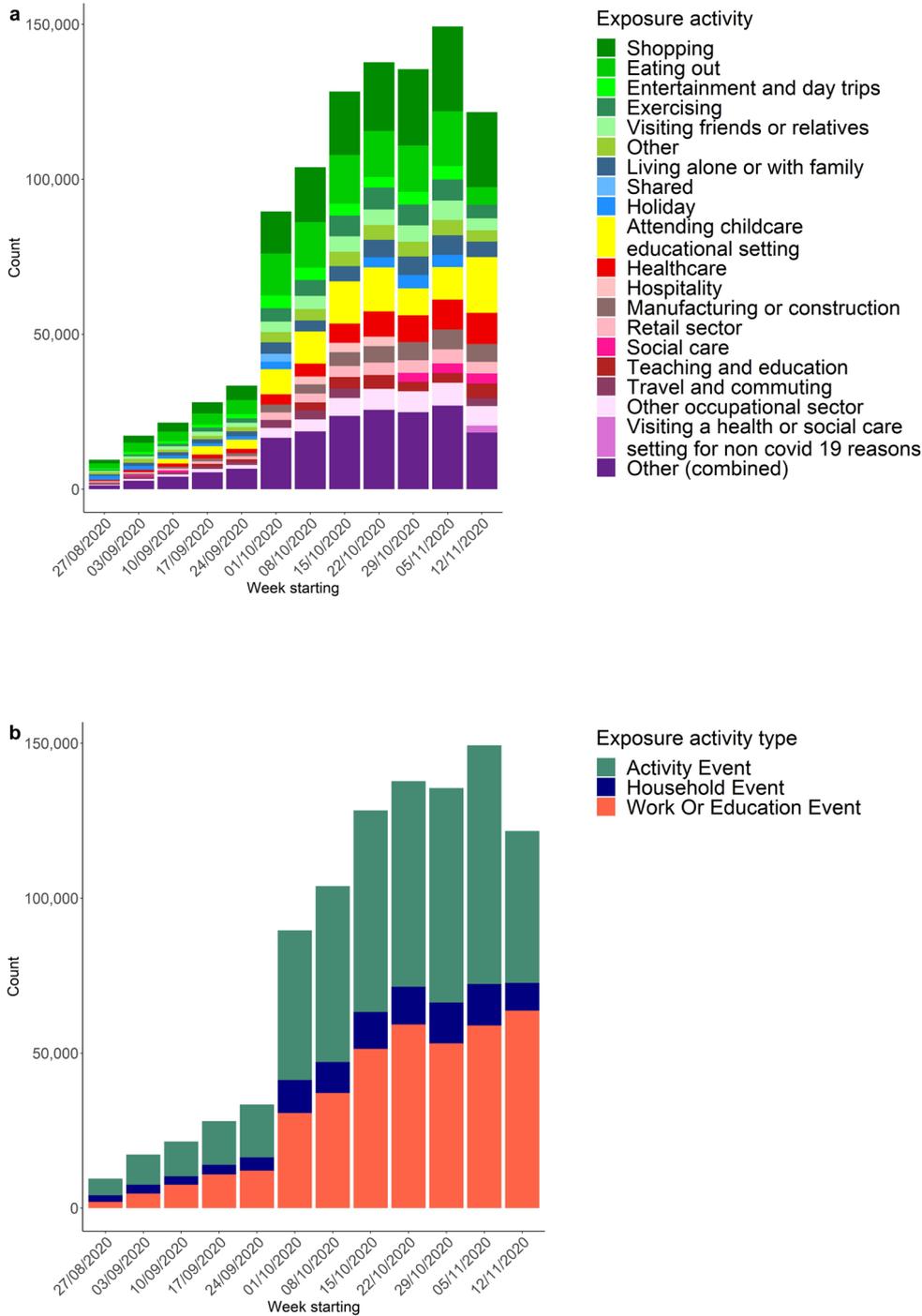
Table A3 in Appendix 3 documents the main published studies on contact tracing. It demonstrates the paucity of published evaluations of major contact tracing interventions in Western European and North American countries. Much of the information is from informal reports rather than formal evaluations based on large numbers traced and follow up of their contacts. At present, evaluation of the potential effectiveness of contact tracing for limiting COVID-19 transmission is primarily from mathematical modelling studies<sup>50</sup>. Programme outcomes vary considerably, primarily focussing on coverage and number of cases and contacts identified. Contact tracing programmes that have been regarded as successful have tended to focus on small populations and/or been evaluated during periods of low viral transmission, so tracing systems have not become overloaded. For example, the success of Northern Ireland's contact tracing initiative, as reported in June 2020, was attributed to Northern Ireland's small population<sup>51</sup>. Furthermore, the evaluation was undertaken while the country was in lockdown, so number of reported contacts was small.

One published report, of the contact tracing programme implemented in Germany, provided data on clusters of transmission and the environments where transmission took place (Appendix 1 Figures A6 and A7)<sup>2</sup>. The dominant transmission environments were private households, care homes, workplaces and hospitals. While schools and universities appeared to play virtually no part in transmission, the reporting period was up to 11th August and so did not capture the effect of educational institutes since they were only fully reopening in September. A further contact tracing study from Trento province, Italy, reported that the majority of transmissions identified were in the households (500 of 3546 contacts becoming infected, 14%), followed by non-cohabiting family and friends (206/1596, 13%), followed by the workplace (79/499, 16%, further details in Appendix 3, Table A3), but this study was conducted in March and April when the province was largely in lockdown, and so contacts outside the household would be expected to be low.<sup>52</sup> Authors presented estimates of contagiousness of cases by age and gender (proportion of contacts becoming infected) and found that children were as infectious, if not more so, than adults: for example, transmission rate 0 – 14 years: 22% (11/49) versus 30-49 years: 11% (250/2361), further details in Appendix 3, Table A3.

**FIGURE 4**

Exposure activities as documented in contact tracing data shared by Public Health England, August – November 2020.

All data show categories where exposure to infected contacts occurred in England by week and do not infer transmission. a) Events/activities reported by individuals who tested positive for SARS-CoV-2 as possible exposure settings in the 3 – 7 days prior to symptom onset, or test date if asymptomatic (backward contact tracing). Information on this type of event and the location are recorded but not information on contacts. b) As for a) but looking at broader categories of exposure settings.

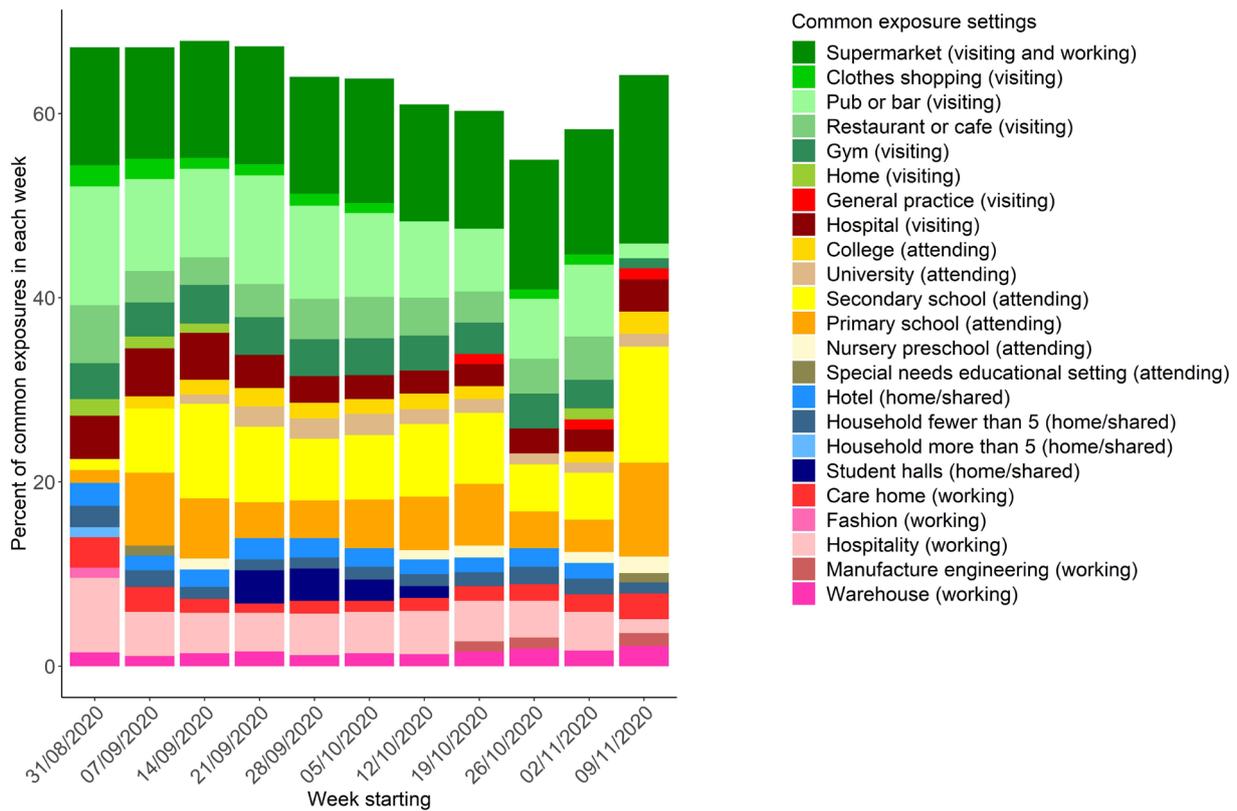


Source: Figure drawn by authors using data from Public Health England.

**FIGURE 5**

Proportion of common exposure settings reported by people who tested positive for SARS-CoV-2 in England relative to all common exposures reported by week.

Contact tracing data shared by Public Health England, August – November 2020. Two or more individuals who tested positive reported the same location, defined by the same postcode, as possible exposure setting in the 3 – 7 days prior to symptom onset, or test date if asymptomatic (backward contact tracing). Information on this type of event and the location are recorded but not information on contacts.



A further study in Spain from 29 May to 2 August 2020 examined outbreaks and cases by setting<sup>53</sup>. The results are presented in Appendix 1 Table A1. This study examined 8390 cases of infection in a wide variety of outbreaks and recorded occupational (work) setting and social setting (including family gatherings) as accounting for over 50% of transmission events. Transmission at social events was the dominant setting for outbreaks (defined as a cluster of cases).

Other regions in the world have had different experiences with contact tracing interventions, collecting detailed data from which key epidemiological and natural history information can be obtained. For example, Hu *et al* reported on intensive contact tracing in Hunan, China, where detailed data collection of the characteristics of cases and their close contacts allowed the estimation of risk factors for susceptibility and transmissibility, including serial interval, generation time, peak of infectiousness, proportion of

transmission that is pre-symptomatic and proportion of infections that are asymptomatic<sup>54</sup>. Authors identified that SARS-CoV-2 transmissibility did not significantly differ between working-age adults (15 – 59 years old) and other age groups (0 – 14 and ≥60 year olds), whilst susceptibility was estimated to increase with age ( $p=0.03$ ). Transmission risk was found to be higher for household contacts ( $p<0.001$ ), decreased for higher generations within a cluster (second generation: odds ratio (OR)=0.13,  $p<0.001$ ; generations 3-4: OR=0.05,  $p<0.001$ , relative to generation 1), and decreased for infectors with a larger number of contacts ( $p=0.04$ ). However, results from these studies very early in the epidemic from Asian countries such as China and South Korea are not generalisable since infection numbers were small, and contact tracing efforts and public compliance with intervention measures are not comparable with the current situation under the second wave in the UK.

Taiwan's quick response to the pandemic, and subsequent prompt activation of their contact tracing response honed during the SARS-CoV-1 epidemic in 2002 – 2003) is clearly demonstrated in their case statistics. To date, 578 cases and 7 deaths have been recorded from a population of 23.8 million (one third of the population of the UK). Comparatively, the UK has seen 1.43 million cases and 53,274 deaths as of the 19th November 2020. Taiwan's relative success may be due in part to the launch of their national contact tracing app, TRACE, inspired by lessons learnt from managing SARS-CoV-1 locally and from experiences in Africa with Ebola<sup>55</sup>. Launched in 2017, the contact tracing app is applicable for all notifiable diseases, and has been used for measles, rubella and avian flu. Adding a COVID module to the existing platform took 48 hours, and has the functionality to add cases, contact information, monitor health status and can produce active maps of cases and contacts. Suspected COVID-19 cases isolate until they receive their results, and if positive will be followed up for contact registration. Movement data collection is aided by police and telecommunication companies where necessary. Patients received twice daily monitoring, either in the form of in-person or two-way text messaging check-ups<sup>56</sup>.

A further study in Hong Kong employing contact tracing data from 1038 cases confirmed between 23 January and 28 April 2020 identified 4 – 7 super spreading events across 51 cluster of infection<sup>57</sup>. Transmission in social settings was associated with more secondary cases than households when controlling for age. This result is more akin to the study cited earlier in Spain, where social mixing was a key determinant of spread.

Recorded outbreaks or super spreading events can also give valuable information on where transmission takes place and who might have seeded the outbreak. Many different types of COVID-19 outbreaks have been documented in the literature, including school-<sup>58, 59</sup> healthcare-<sup>60</sup> ship-<sup>41, 61, 62</sup> and care home-<sup>63</sup> associated outbreaks. Sequence data from care home staff and residents identified multiple independent introductions of SARS-CoV-2 into each care home by both family members, friends and staff, rather than a single introduction followed by within-care home transmission. Several introduction events were followed by considerable within-care home transmission, although there were numerous instances where an introduction event was not followed by any detected forward transmission. Transmission clusters often contained isolates from both residents and staff members, although it was not possible to infer directionality. No clusters contained isolates from different care homes, challenging the suggestion that staff working in multiple care homes had transmitted SARS-CoV-2 between different homes.

For work places the evidence varies greatly between countries. Some common threads include farms who employ migrant farm workers perhaps sharing group housing that share amenities, slaughter-houses, meat packing factories, and factories in general where people work in close proximity<sup>64, 65</sup>. In military settings such as warships a few very detailed studies have charted spread, and settings where this is more frequent<sup>41</sup>. Those who work in enclosed environments with others were typically most likely to acquire infection.

Closed environments such as gyms and restaurants are also identified as important transmission sites in many international studies<sup>66</sup>. Nishiura et al. investigated 110 cases among eleven clusters in Japan, and suggest that closed indoor environments, including fitness gyms, a restaurant boat on a river, hospitals, and a snow festival in tents, are 18.7 times more likely to facilitate primary case transmission, compared to open-air environments.

Public transport may also be an important setting. Zheng et al. evaluated the frequency of trains, buses and planes in China, and concluded that public transportation is an important factor in spreading the virus, with particular emphasis on the connectivity and distance between the epicenter and the destination<sup>67</sup>.

Restaurants are a common theme in many countries<sup>68</sup>. A case study of 154 COVID-19 outpatients found that they were about twice as likely to have eaten at restaurants (indoor, patio, and outdoor seating) than the control participants who tested negatively. A case study in China showed that transmission occurred between non-associated families in a restaurant with poor ventilation. No close-contact or fomite contact was observed, and transmission was likely via virus-laden aerosols<sup>69</sup>.

### Household studies

Contact tracing in the UK suggests households are the most important setting for contacts and subsequent transmission. Table A6 in Appendix 3 lists a wide variety of such studies and records the main findings from published studies and reports.

To elucidate risk factors for household transmission, household studies provide an alternative to studies based on routine data collection from contact tracing. For example, Grijalva et al have reported results from the first seven days of an ongoing prospective case-ascertained household study using intensive daily observation, conducted in Tennessee and Wisconsin (enrolment April-September 2020), which characterised transmission of SARS-CoV-2 in US households (Appendix 1 Figure A8<sup>70</sup>). Index patients

were defined as the first household member with confirmed COVID-19 and who lived with at least one other household member. Household members were trained to complete symptom diaries and obtain self-collected specimens daily for 14 days. Authors' analysis of specimens from the first 7 days found that 191 enrolled household contacts of 101 index patients reported having no symptoms on the day of the associated index patient's illness onset, and among these 191 contacts, 102 had SARS-CoV-2 detected in either nasal or saliva specimens during follow-up, for a secondary infection rate of 53% (95%CI 46%–60%). Excluding 54 household members who had SARS-CoV-2 detected in specimens taken at enrolment, the secondary infection rate was 35% (95%CI 28%–43%). Secondary infections occurred rapidly, with approximately 75% of infections identified within 5 days of the index patient's illness onset. Secondary infection rates were high across all racial/ethnic groups. Substantial transmission occurred whether the index patient was an adult or a child (Appendix 3, Table A6).

Following 233 UK households (>2 members inclusive) and their 472 members, Lopez Bernal<sup>71</sup> conducted a prospective case ascertained study in late January this year (Appendix 1 Figure A8). At this time point, the studied household R0 was 1.85, and a household reproduction number of 2.33. Following the households elucidated the fact secondary attack rates (SAR) were higher when the primary case was a child.

Fung and colleagues<sup>72</sup> reviewed 22 published and pre-published studies from 10 countries (20,291 household contacts) that were available through September 2nd 2020. Household SAR tended to be higher among older adult contacts. Half the included studies were from China, with the remainder from South Korea (n=2), Taiwan (n=1), Singapore (n=1), Brunei (n=1), Israel (n=1), Germany (n=1), the Netherlands (n=1), the United States (n=2), and Australia (n=1). As evident from Appendix 3, Table A5, data for Western Europe are lacking by comparison with South East Asia and China.

The household secondary attack rate was higher among adults and older adults than among children, and higher among female contacts. SARs were also elevated among spouses or significant others of index cases relative to non-spouse household members (Appendix 3 Table A6).

Thirteen studies reported the household SAR by age of the index case (Appendix 3 Table A6). In the study from Qingdao, China<sup>73</sup>, the household SAR was higher when index cases were 55 years or older than when cases were younger than 55 years. In South Korea<sup>74</sup> the household SAR was highest for index cases aged 10 – 19 years (18.6%; 95% CI: 14.0 – 24.0%) and lowest for those younger than 9 (5.3%; 95% CI: 1.3 – 13.7%); by comparison, the SAR estimates for adult age groups ranged from 7.0 to 18.0%.

Detailed reviews have been published by Lei *et al*<sup>75</sup> and Shah *et al*<sup>76</sup>. In 6 of the 24 studies, secondary attack rates were stratified by age, yielding a range of SARs from 15.7% to 47.6% in adults and from 5.2% to 26.9% in children<sup>75</sup>. This meta-analysis indicates that the risk of household transmission in adults is about three times higher than that in children (odds ratio (OR) = 3.67, 95% CI: 2.7<sup>6</sup> – 4.8<sup>7</sup>, p < 0.001). In 10 studies, the SAR among other contacts (not household contacts) was also reported and ranged from 0.1% to 28.8%. The meta-analysis indicates the risk of household transmission is about 10 times higher than that from other contacts (OR = 10.72, 95% CI: 5.70–20.17, p < 0.001) (Appendix 3 Table A6). Shah in another extensive review suggested greater vulnerability of spouse and elderly population for secondary transmission than other household members<sup>76</sup>.

### Viral genome sequencing studies

Whole genome sequencing (WGS) can be used to augment public health responses to infectious pathogens on broadly speaking three different organisational and geographic levels<sup>77</sup>. At the local level, e.g. hospital or care home outbreaks, WGS can be used to reconstruct transmission chains at the individual level, sources of outbreaks and reveal inadequacies in infection control measures. At the national level, WGS can retrospectively identify the geographic sources of introduction of SARS-CoV-2 and the number of independent introduction or 'seeding' events into previously unaffected countries. Phylogenomic analyses can be used to estimate the timing of introductions, to estimate in which regions the virus spread first and how fast the spread occurred. At the international level, WGS can reveal the geographic distribution of viral lineages and how SARS-CoV-2 spread between countries.

In the UK the COVID-19 Genomics Consortium has sequenced more than 100,000 SARS-CoV-2 genomes by 10 November 2020 and is currently sequencing 8,000 – 10,000 genomes per week with a plan to double this number over the coming months<sup>78, 79</sup>. The first 26,181 sequenced genomes from the UK were used to infer countries of origin from where SARS-CoV-2 was introduced into the UK and the geographic spread of the virus in the early phase of the epidemic in the UK<sup>80</sup>. The study identified 1179 UK transmission lineages. The earliest entries were from Italy, Spain and France. Lineages that were imported early spread further and are longer-lived than lineages that were imported later and whose spread was curbed by transmission control measures.

Richmond *et al*<sup>12</sup> report WGS of SARS-CoV-2 cases in La Crosse County, Wisconsin in September 2020. Although the majority of cases were among college-age individuals, from a total of 111 genomes sequenced the authors identified rapid transmission of the virus into more vulnerable populations. Eight sampled genomes represented two independent transmission events into two skilled nursing facilities, resulting in two fatalities.

A study from Iceland published by Gudbjartsson *et al* linked SARS-CoV-2 genome sequence data to contact tracing data to confirm transmission networks and settings where infection occurred (see example below)<sup>14</sup>. In this study, as in other studies that linked SARS-CoV-2 genomic data to contact tracing data, the genome sequences were mainly used to distinguish between imported cases and community transmission and to identify local transmission clusters, regional transmission networks and super spreading events<sup>81–86</sup>. High within-host genetic diversity may indicate high-dose transmission events<sup>86</sup>. Another useful application of SARS-CoV-2 genomic data is the identification of transmission routes in healthcare settings<sup>60, 87</sup>. Where genome sequencing studies have helped to confirm where transmission occurs, at-risk settings were mainly households, healthcare settings, social and religious gatherings<sup>14, 60, 84, 85</sup>. For example, Meredith *et al*'s epidemiological investigation and sequencing of transmission clusters in healthcare settings in Cambridge, England, found that a number of cases previously thought to have been community acquired were linked to an outpatient dialysis unit where shared transport to and from the clinic was identified as a likely contributing factor<sup>60</sup>.

Figure 6 demonstrates the mid-March peak in importation events, because moderate-to-high levels of inbound travel coincided with growing or peak case numbers in several European countries<sup>80</sup>. It shows that the rate and source of introduction of SARS-CoV-2 lineages into the UK changed

substantially and rapidly through time. The increasing rates and shifting source locations of SARS-CoV-2 importation were not fully captured by early contact tracing.

A further example of how the evolution of the virus has progressed over time in given settings is well illustrated by the work of Korber and colleagues on the spread of the G614 strain in Europe and its replacement of the D614 strain<sup>88</sup>. Clearly, G614 has a higher  $R_0$  than the D614 strain since it is this value that will determine who wins in terms of dominance in spread patterns.

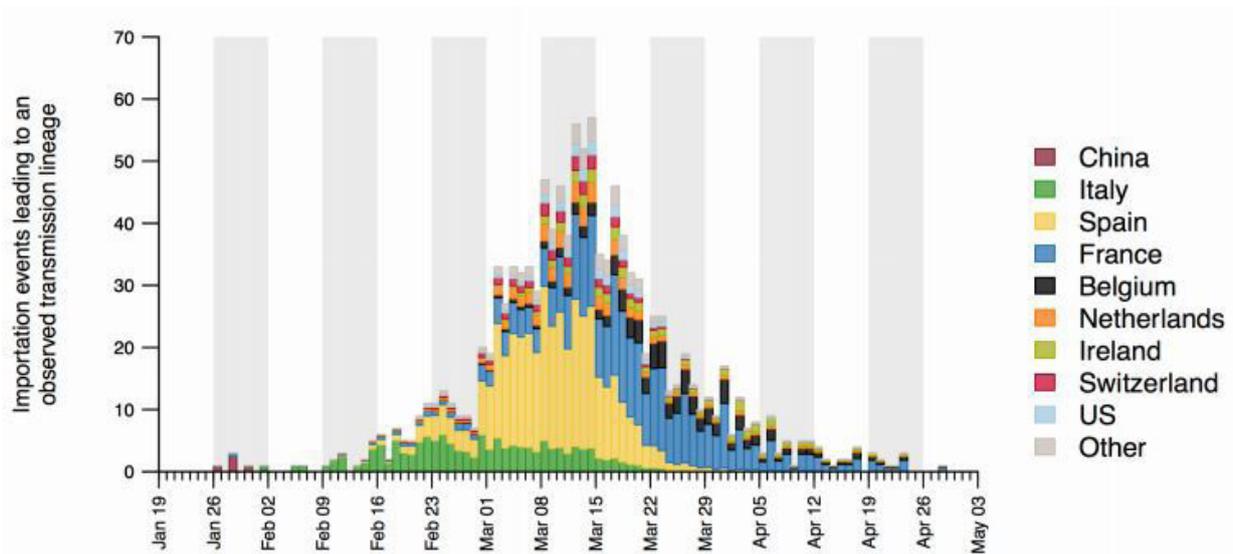
Gudbjartsson and colleagues report how WGS was used to augment the tracing of infected contacts in Iceland<sup>14</sup>. In an illustrated example (Appendix 1 Figure A10), they first show how SARS-CoV-2 sequence data from suspected infected contacts was first used to identify the source of infection as travellers returning from Italy. Prior to WGS it was unclear if the source of infection were travellers returning from Italy or Austria. WGS then revealed that two infected contacts were in fact not linked to the infection cluster but had acquired the infection from a different source. WGS also discovered a new mutation that arose in one of the cases in the transmission network and that this person passed on the mutated strain to new contacts. Finally, a search for similar SARS-CoV-2 genomes in the Icelandic population linked two more individuals to the same transmission network, although the route of transmission for these two cases could not be identified.

The importance of continued WGS studies of SARS-CoV-2 isolates taken day by day, and week by week through the various waves of infection by PHE, the Sanger Centre and university-based research groups, has been highlighted recently by the identification of a new strain spreading rapidly in the UK announced on the 14 December 2020 by the Secretary of State for Health and Social Care.

Genotypic changes in the RNA virus, SARS-CoV-2, have occurred continuously since the start of the pandemic. Charting these changes via GWS is straightforward. Associating genotypic changes, with changes in the phenotype of the virus, is more challenging since epidemiological information is required on both the respective rates of spread of different genotypes (=strains) and the associated morbidity and mortality. However, charting evolution in association with phenotypic properties must be a high priority as mutations will continue to occur, and mass vaccination will impose strong evolutionary pressures on the virus.

**FIGURE 6**

Estimated number of COVID-19 importation events attributed to inbound travellers to the UK from different countries of embarkation.



Source: Figure taken from du Plessis *et al*<sup>60</sup>.

### Data needs in the UK

Government and research funding bodies have the difficult task of distributing effort between national data collection infrastructure and small-scale research studies of specific issues. Understandably, current priorities are centred on the national data collection programmes.

Before turning to the case for more attention to be focused on the smaller scale projects, it is important to note that data collection and presentation at a national or regional scale is only part of the task to better understand and control viral spread. TTI is a good example of problems that need resolving. It is not just seeking enhanced tracing systems, encouraging those in contact with positive test individuals to isolate, and timely reporting of test results to those seeking tests. It is also timely analysis of the TTI data to better understand where most people acquire infection and where index cases introducing infection into settings such as the household, acquire their infection. At present PHE has provided very limited open access to the TTI information for detailed analysis by epidemiologists which would help guide policy formulation on how best to design social distancing measures to limit transmission in given regions. Furthermore, careful thought should be given to what information should be acquired from those contacted and how should this be summarized week by week.

Government should address this information and analysis gap, given the very large investment being made in TTI and the urgent need to understand better where most people get infected to inform policy formulation.

Turning to smaller scale, more focused studies, these are urgently needed not just in the UK but internationally. Each country is different due to prevailing cultural norms and diversity. As such where most transmission takes place will differ, as it will over time as social distancing measures are imposed and relaxed during and between the different waves of infection. Research funding agencies and local health authorities in collaboration with university based epidemiological research groups should be the focus for the more targeted and detailed studies involving contact tracing, household studies and perhaps most importantly WGS to determine who infects whom in different settings (small communities, work-place public transport and specialised care settings, including care homes and hospitals).

Little work in this area has been published or released to date from UK sources and funding should be increased in this area. Such studies are difficult to design and implement since they must be conducted ethically (under GDPR regulations) and within existing public health structures while the epidemic is progressing. For example, TTI in principle provides the opportunity to acquire viral DNA, but for WGS it must be preserved appropriately before transporting to sequencing centres (such as the Wellcome Trust Genome Campus in Hinxton). In addition, to get benefit in terms of understanding transmission better by setting and in terms of who infects whom, the viral isolate must be accompanied by, or linked to appropriate epidemiological, demographic and social information in a confidential data base (protecting privacy and following GDPR regulations). It would be impossible to do this for all tests (even taking out rapid test procedures where whole genomes may not be collected), so a sampling regime would need to be designed to try and get representative groups of those tested and to encourage their participation in information provision to better understand where transmission may have taken place. This is a complex task, but studies of this nature are essential to further understanding of SARS-CoV-2 transmission in the UK.

## Discussion

The available data from UK sources highlights both the importance of the household as a site of secondary transmission and the dynamic nature of where people get infected as season and social distancing measures change over time and space (see Figure 2 for data from TTI collated by PHE). Season is important, as it is for the transmission of influenza A, where summertime in the Northern Hemisphere countries can take  $R_0$  to less than unity in value as people spend more time in well ventilated spaces and outside of the household. For COVID-19, in the absence of social distancing measures,  $R_0$  is likely to remain above unity in value irrespective of changes in behaviour over the summer months due to the relatively high non-summer  $R_0$  values (2 – 4) by comparison with influenza A (1.1 – 1.5), as well illustrated by the rise in case number prior to the onset of autumn in the UK in 2020.

How index cases introducing infection into households acquire their COVID-19 infection is less certain. Very limited data from the UK suggests social gatherings in restaurant and bars are important, as are shops and supermarkets. However, the robustness of this conclusion is weak given limited data. There is a need for more and better small scale, highly focused studies and timely, detailed analyses of what is revealed in the national TTI systems. Test Trace and Isolation has huge potential in the era of cell phones and genomics to assess who infects whom but requires some refining to be fully effective when case numbers

are low or in defined local settings. This hopefully will be important as mass vaccination reduces transmission via the impact of herd immunity on the effective reproductive number  $R$ . However, when transmission is extensive in the population it will always tend to fail to provide information in a timely manner, due to the extensive workload in tracing. Recent experiences in most countries, including South Korea and Japan, reflect this reality. The relatively short incubation period before people are infectious to others (perhaps around 3 days in symptomatic individuals,) plus the long infectious period of the virus (10 days or so), also mitigate against the effectiveness of TTI, since tracing and concomitant isolation must occur rapidly to have any impact on onward transmission.

The main message from the literature review and comparative study across countries, is the paucity of good data from European countries on contact tracing, household studies and WGS to ascertain who infects whom. Suggestions are made in the proceeding section on what is required. They involve both open access via PHE to TTI data for detailed analysis, and smaller, focused research studies in defined populations and settings.

The COG-UK updates timeline reports the consortium has sequenced more than 100,000 SARS-CoV-2 genomes by 10 November 2020, and the most recent number on the Microreact page is 104,568. More analysis should be conducted on what this database tells us about, not just viral evolution (critically important for the development of both diagnostic test and vaccines), but also who infects whom in defined settings. In the COG-UK blog a new funding announcement was made on the 18th November 2020 in which they state that the near-term aim is to sequence 8,000 – 10,000 SARS-CoV-2 genomes per week<sup>79</sup>. They also plan to double that number over the coming months. It is vitally important that, where possible, good demographic and epidemiological information is linked to these genome sequences and timely analyses of the data is released. If that transpires, then the UK will have a database to refine our understanding of who infects whom and concomitantly, where transmission takes place most frequently. In turn, this will enable policy makers to better tune social distancing measures.

## Authors

This review was compiled by Professor Sir Roy Anderson FMedSci FRS, Dr Carolin Vegvari and Rosie Maddren from Imperial College London and Dr Rebecca Baggaley from the University of Leicester for SET-C of The Royal Society.

## Acknowledgements

The authors of this report would like to thank Professor Melinda Mills FBA for helpful comments on the manuscript and Charlotte Anderson and Lina Hiironen from Public Health England for sharing contact tracing data and explaining the definitions of variables. We would also like to thank Professor Mark Woolhouse FMedSci for help with the Scottish contact tracing data.

### DISCLAIMER

This paper has drawn on the most recent evidence up to 14 December 2020 and has been subject to formal peer-review. Further evidence on this topic is constantly published and the Royal Society may return to this topic in the future. This independent overview of the science has been provided in good faith by subject experts and the Royal Society and paper authors accept no legal liability for decisions made based on this evidence.

### THANKS

The Royal Society is grateful to the Leverhulme Trust for its support for the Society's pandemic response work.

---

The text of this work is licensed under the terms of the Creative Commons Attribution License which permits unrestricted use, provided the original author and source are credited. The license is available at: [creativecommons.org/licenses/by/4.0](https://creativecommons.org/licenses/by/4.0)

Issued: December 2020 DES7296 © The Royal Society

# References

1. Google. 2020 COVID-19 Community Mobility Reports. Available from: <https://www.google.com/covid19/mobility/>
2. Robert Koch Institute. 2020 Epidemiologisches Bulletin: Characteristics of confirmed COVID-19 outbreaks in Germany. Available from: [https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/38\\_20.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/38_20.pdf?__blob=publicationFile)
3. Thompson H *et al.* 2020 Report 38: SARS-CoV-2 setting-specific transmission rates: a systematic review and meta-analysis. (doi:10.25561/84270)
4. Environmental Modelling Group, Scientific Pandemic Influenza Group on Behaviours. 2020 EMG/SPI-B: Mitigating risks of SARS-CoV-2 transmission associated with household social interactions, published 26th November 2020. Available from: [https://www.gov.uk/government/publications/emgspl-b-mitigating-risks-of-sars-cov-2-transmission-associated-with-household-social-interactions-26-november-2020?fbclid=IwAR3TPsNxUcx\\_ulx20mFFgLEyxCb2kIMfKgMGRhEWUS8mLS-vJf41KNSPn1](https://www.gov.uk/government/publications/emgspl-b-mitigating-risks-of-sars-cov-2-transmission-associated-with-household-social-interactions-26-november-2020?fbclid=IwAR3TPsNxUcx_ulx20mFFgLEyxCb2kIMfKgMGRhEWUS8mLS-vJf41KNSPn1)
5. Simpson REH. 1948 The period of transmission in certain epidemic diseases an observational method for its discovery. *The Lancet* 252, 755–760. (doi:10.1016/s0140-6736(48)91328-2)
6. Simpson REH. 1952 Infectiousness of communicable diseases in the household. *The Lancet* 260, 549–554. (doi:10.1016/s0140-6736(52)91357-3)
7. Sartwell PE. 1966 The incubation period and the dynamics of infectious disease. *American Journal of Epidemiology* 83, 204–216. (doi:10.1093/oxfordjournals.aje.a120576)
8. Bailey NTJ. 1957 *The mathematical theory of epidemics*. New York Hafner Publishing Co. (ISBN: 0852641133)
9. Wharton ME. 2004 Measles Elimination in the United States. *The Journal of Infectious Diseases* 189, S1–S3. (doi:10.1086/377693)
10. Lane JM. 2006 Mass Vaccination and Surveillance/Containment in the Eradication of Smallpox. In *Current Topics in Microbiology and Immunology*, pp. 17–29. Springer Berlin Heidelberg. (doi:10.1007/3-540-36583-4\_2)
11. Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyi ID. 1988 Smallpox and its eradication. Available from: <https://apps.who.int/iris/handle/10665/39485>
12. Richmond CS, Sabin AP, Jobe DA, Lovrich SD, Kenny PA. 2020 SARS-CoV-2 sequencing reveals rapid transmission from college student clusters resulting in morbidity and deaths in vulnerable populations. *MedRxiv*. (doi: 10.1101/2020.10.12.20210294)
13. The Royal Society. 2020 The SARS-CoV-2 genome: variation, implication and application. Available from: <https://royalsociety.org/-/media/policy/projects/set-c/set-c-genome-analysis.pdf?la=en-GB&hash=CF1883F618E851FF269487B02AB19CF8>
14. Gudbjartsson DF *et al.* 2020 Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med* 382, 2302–2315. (doi:10.1056/nejmoa2006100)
15. Voeten HACM *et al.* 2020 Unravelling the modes of transmission of SARS-CoV-2 during a nursing home outbreak: looking beyond the church super-spread event. *Clinical Infectious Diseases* (doi:10.1093/cid/ciaa1664)
16. Korber B. 2000 Timing the Ancestor of the HIV-1 Pandemic Strains. *Science* 288, 1789–1796. (doi:10.1126/science.288.5472.1789)
17. Gifford RJ *et al.* 2007 Phylogenetic Surveillance of Viral Genetic Diversity and the Evolving Molecular Epidemiology of Human Immunodeficiency Virus Type 1. *JVI* 81, 13050–13056. (doi:10.1128/jvi.00889-07)
18. Anderson RM, Hollingsworth TD, Baggaley RF, Maddren R, Vegvari C. 2020 COVID-19 spread in the UK: the end of the beginning? *The Lancet* 396, 587–90. (doi: 10.1016/S0140-6736(20)31689-5).
19. Anderson RM, May RM. 1985 Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *J. Hyg.* 94, 365–436. (doi:10.1017/s002217240006160x)
20. Prem K, Cook AR, Jit M. 2017 Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol* 13, e1005697. (doi:10.1371/journal.pcbi.1005697)
21. Dowd JB, Block P, Rotondi V, Mills MC. 2020 Dangerous to claim “no clear association” between intergenerational relationships and COVID-19. *Proc Natl Acad Sci USA*, 117, 25975–6. (doi: 10.1073/pnas.2016831117).
22. Keohane N. Longer lives, stronger families: The changing nature of intergenerational support. The Social Market Foundation, February 2016. ISBN: 978-1-910683-08-8 Available at: <https://www.smf.co.uk/publications/longer-lives-stronger-families-the-changing-nature-of-intergenerational-support/> Accessed 19th November 2020. 2016.
23. Chan TW, Ermisch J. 2015 Proximity of Couples to Parents: Influences of Gender, Labor Market, and Family. *Demography* 52, 379–399. (doi:10.1007/s13524-015-0379-0)
24. Chan TW, Ermisch J. 2015 Residential proximity of parents and their adult offspring in the United Kingdom, 2009–10. *Population Studies* 69, 355–372. (doi:10.1080/00324728.2015.1107126)
25. Mooney G. 2020 “A Menace to the Public Health” — Contact Tracing and the Limits of Persuasion. *N Engl J Med* 383, 1806–1808. (doi:10.1056/nejmp2021887)
26. Fraser C, Riley S, Anderson RM, Ferguson NM. 2004 Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences* 101, 6146–6151. (doi:10.1073/pnas.0307506101)
27. Klinkenberg D, Fraser C, Heesterbeek H. 2006 The Effectiveness of Contact Tracing in Emerging Epidemics. *PLoS ONE* 1, e12. (doi:10.1371/journal.pone.0000012)
28. Wykoff RF. 1988 Contact Tracing to Identify Human Immunodeficiency Virus Infection in a Rural Community. *JAMA* 259, 3563. (doi:10.1001/jama.1988.03720240025029)
29. Vernon TM, Hoffman RE. 1988 Contact Tracing to Control the Spread of HIV. *JAMA* 260, 3274. (doi:10.1001/jama.1988.03410220058014)
30. Group WHOREAfC-TW *et al.* 2020 Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19. *JAMA* 324, 1330. (doi:10.1001/jama.2020.17023)
31. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. 2020 Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19). *JAMA* (doi:10.1001/jama.2020.6019)
32. Dept. of Health and Social Care. 2020 Guidance: NHS Test and Trace: how it works. Available from: <https://www.gov.uk/guidance/nhs-test-and-trace-how-it-works>
33. Llywodraeth Cymru Welsh Government. 2020 Policy and Strategy: Test Trace Protect. Available from: <https://gov.wales/test-trace-protect>
34. Northern Ireland government services. 2020 Coronavirus (COVID-19): testing and contact tracing. Available from: <https://www.nidirect.gov.uk/>

35. Scottish Government. 2020 Coronavirus (COVID-19): Test and Protect. Available from: <https://www.gov.scot/publications/coronavirus-covid-19-test-and-protect/>
36. Rubin GJ, Smith LE, Melendez-Torres G, Yardley L. 2020 Improving adherence to 'test, trace and isolate'. *J R Soc Med* 113, 335–338. (doi:10.1177/0141076820956824)
37. Reynolds M. 2020 It's not just testing that is failing, contact tracing is broken too. Published 18th September 2020. Available from: <https://www.wired.co.uk/article/coronavirus-nhs-contact-tracing-testing>
38. Financial Times. 2020 Low UK isolation rates raise worries of virus spreading. Available from: <https://www.ft.com/content/d214ef31-7a5c-4bb5-894f-17c7057ab999>
39. Smout A. 2020 Over 80% of Britons not heeding COVID-19 self-isolation rules, study finds. Available from: <https://uk.reuters.com/article/us-health-coronavirus-britain-trace/over-80-of-britons-not-heeding-covid-19-self-isolation-rules-study-finds-idUKKCN26G14F>
40. Smith LE *et al.* 2020 Adherence to the test, trace and isolate system: results from a time series of 21 nationally representative surveys in the UK (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] study). *medRxiv*. (doi: 10.1101/2020.09.15.20191957).
41. Kasper MR *et al.* 2020 An Outbreak of Covid-19 on an Aircraft Carrier. *N Engl J Med* (doi:10.1056/nejmoa2019375)
42. Poletti P *et al.* 2020 Probability of symptoms and critical disease after SARS-CoV-2 infection. *arXiv*. (arXiv: 2006.08471)
43. The Royal Society. 2020 Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. Available from: <https://royalsocietypublishing.org/~/media/policy/projects/set-c/set-covid-19-R-estimates.pdf>
44. Kucharski AJ *et al.* 2020 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* 20, 1151–1160. (doi:10.1016/s1473-3099(20)30457-6)
45. Keeling MJ, Hollingsworth TD, Read JM. 2020 Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *J Epidemiol Community Health*, jech-2020-214051. (doi:10.1136/jech-2020-214051)
46. Dept. of Health and Social Care. 2020 Guidance: NHS Test and Trace statistics (England): methodology. Available from: <https://www.gov.uk/government/publications/nhs-test-and-trace-statistics-england-methodology/nhs-test-and-trace-statistics-england-methodology>
47. UK Government. 2020 Coronavirus (COVID-19) in the UK. Available from: <https://coronavirus.data.gov.uk/details/cases>
48. Public Health Scotland. 2020 Data and Intelligence. Available from: <https://beta.isdscotland.org/find-publications-and-data/population-health/covid-19/covid-19-statistical-report/28-october-2020/dashboard/>
49. Office for National Statistics. 2020 Coronavirus and the social impacts on Great Britain: 20 November 2020. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandthesocialimpactsongreatbritain/20november2020>
50. Braithwaite I, Callender T, Bullock M, Aldridge RW. 2020 Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *The Lancet Digital Health* 2, e607–e621. (doi:10.1016/s2589-7500(20)30184-9)
51. Baraniuk C. 2020 How Northern Ireland relaunched contact tracing for covid-19 a month before the rest of the UK. *BMJ*, m2373. (doi:10.1136/bmj.m2373)
52. Fateh-Moghadam P *et al.* 2020 Contact tracing during Phase I of the COVID-19 pandemic in the Province of Trento, Italy: key findings and recommendations. *medRxiv*. (doi: 10.1101/2020.07.16.20127357)
53. The national COVID-19 outbreak monitoring group. 2020 COVID-19 outbreaks in a transmission control scenario: challenges posed by social and leisure activities, and for workers in vulnerable conditions, Spain, early summer 2020. *Eurosurveillance* 25. (doi:10.2807/1560-7917.es.2020.25.35.2001545)
54. Hu S *et al.* 2020 Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. *medRxiv*. (doi: 10.1101/2020.07.23.20160317)
55. Polonsky JA *et al.* 2019 Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Phil. Trans. R. Soc. B* 374, 20180276. (doi:10.1098/rstb.2018.0276)
56. Jian S-W, Cheng H-Y, Huang X-T, Liu D-P. 2020 Contact tracing with digital assistance in Taiwan's COVID-19 outbreak response. *International Journal of Infectious Diseases* 101, 348–352. (doi:10.1016/j.ijid.2020.09.1483)
57. Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, Leung GM, Cowling BJ. 2020 Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med* 26, 1714–1719. (doi:10.1038/s41591-020-1092-0)
58. Stein-Zamir C *et al.* 2020 A large COVID-19 outbreak in a high school 10 days after schools' reopening, Israel, May 2020. *Eurosurveillance* 25. (doi:10.2807/1560-7917.es.2020.25.29.2001352)
59. Fontanet A *et al.* 2020 Cluster of COVID-19 in northern France: A retrospective closed cohort study. *medRxiv* 2020. (doi: 10.1101/2020.04.18.20071134)
60. Meredith LW *et al.* 2020 Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *The Lancet Infectious Diseases* 20, 1263–1272. (doi:10.1016/s1473-3099(20)30562-4)
61. Payne DC *et al.* 2020 SARS-CoV-2 Infections and Serologic Responses from a Sample of U.S. Navy Service Members — USS Theodore Roosevelt, April 2020. *MMWR Morb. Mortal. Wkly. Rep.* 69, 714–721. (doi:10.15585/mmwr.mm6923e4)
62. Mizumoto K, Kagaya K, Zarebski A, Chowell G. 2020 Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* 25. (doi:10.2807/1560-7917.es.2020.25.10.2000180)
63. Ladhani SN *et al.* 2020 Investigation of SARS-CoV-2 outbreaks in six care homes in London, April 2020. *EClinicalMedicine* 26, 100533. (doi:10.1016/j.eclinm.2020.100533)
64. Haley E, Caxaj S, George G, Hennebray J, Martell E, McLaughlin J. 2020 Migrant Farmworkers Face Heightened Vulnerabilities During COVID-19. *J. Agric. Food Syst. Community Dev.*, 1 – 5. (doi:10.5304/jafscd.2020.093.016)
65. Günther T *et al.* 2020 SARS-CoV-2 outbreak investigation in a German meat processing plant. *EMBO Mol Med* (doi:10.15252/emmm.202013296)
66. Nishiura H *et al.* 2020 Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). *medRxiv*. (doi: 10.1101/2020.02.28.20029272)

67. Zheng R, Xu Y, Wang W, Ning G, Bi Y. 2020 Spatial transmission of COVID-19 via public and private transportation in China. *Travel Medicine and Infectious Disease* 34, 101626. (doi:10.1016/j.tmaid.2020.101626)
68. Fisher KA *et al.* 2020 Community and Close Contact Exposures Associated with COVID-19 Among Symptomatic Adults  $\geq 18$  Years in 11 Outpatient Health Care Facilities — United States, July 2020. *MMWR Morb. Mortal. Wkly. Rep.* 69, 1258–1264. (doi:10.15585/mmwr.mm6936a5)
69. Li Y *et al.* 2020 Evidence for probable aerosol transmission of SARS-CoV-2 in a poorly ventilated restaurant. medRxiv. (doi: 10.1101/2020.04.16.20067728)
70. Grijalva CG *et al.* 2020 Transmission of SARS-COV-2 Infections in Households — Tennessee and Wisconsin, April–September 2020. *MMWR Morb. Mortal. Wkly. Rep.* 69. (doi:10.15585/mmwr.mm6944e1)
71. Lopez Bernal J, Panagiotopoulos N, Byers C. 2020 Transmission dynamics of COVID-19 in household and community settings in the United Kingdom. medRxiv. (doi: 10.1101/2020.08.19.20177188)
72. Fung HF, Martinez L, Alarid-Escudero F, Salomon JA, Studdert DM, Andrews JR, Goldhaber-Fiebert JD. 2020 The household secondary attack rate of SARS-CoV-2: A rapid review. *Clinical Infectious Diseases* (doi:10.1093/cid/ciaa1558)
73. Xin H *et al.* 2020 Risk factors associated with occurrence of COVID-19 among household persons exposed to patients with confirmed COVID-19 in Qingdao Municipal, China. *Transbound Emerg Dis* (doi:10.1111/tbed.13743)
74. Park YJ *et al.* 2020 Contact Tracing during Coronavirus Disease Outbreak, South Korea, 2020. *Emerg. Infect. Dis.* 26, 2465–2468. (doi:10.3201/eid2610.201315)
75. Lei H, Xu X, Xiao S, Wu X, Shu Y. 2020 Household transmission of COVID-19—a systematic review and meta-analysis. *Journal of Infection* (doi:10.1016/j.jinf.2020.08.033)
76. Shah K, Saxena D, Mavalankar D. 2020 Secondary attack rate of COVID-19 in household contacts: a systematic review. *QJM: An International Journal of Medicine* (doi:10.1093/qjmed/hcaa232)
77. World Health Organisation. 2020 Global Antimicrobial Resistance and Use Surveillance System (GLASS): GLASS whole-genome sequencing for surveillance of antimicrobial resistance. Available from: <https://www.who.int/publications/i/item/9789240011007>
78. COG-UK COVID-19 Genomics UK Consortium. 2020 Blog: COG-UK passes 100K genomes. Available from: [https://www.cogconsortium.uk/news\\_item/cog-uk-passes-100k-genomes/](https://www.cogconsortium.uk/news_item/cog-uk-passes-100k-genomes/)
79. COG-UK COVID-19 Genomics UK Consortium. 2020 Blog: What next for COG-UK following new funding of £12.2 million? Available from: [https://www.cogconsortium.uk/news\\_item/what-next-for-cog-uk-following-new-funding-of-12-2-million/](https://www.cogconsortium.uk/news_item/what-next-for-cog-uk-following-new-funding-of-12-2-million/)
80. du Plessis L *et al.* 2020 Establishment & lineage dynamics of the SARS-CoV-2 epidemic in the UK. medRxiv. (doi: 10.1101/2020.10.23.20218446)
81. Watson C. 2020 How countries are using genomics to help avoid a second coronavirus wave. *Nature*, 582, 19. (doi: 10.1038/d41586-020-01573-5)
82. Oude Munnink BB *et al.* 2020 Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med*, 26, 1405-1410. (doi: HYPERLINK "https://doi.org/10.1038/s41591-020-0997-y" 10.1038/s41591-020-0997-y)
83. Rockett RJ *et al.* 2020 Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modelling. *Nat Med*, 26, 1398-1404. (doi:10.1038/s41591-020-1000-7)
84. Zhang W *et al.* 2020 Analysis of Genomic Characteristics and Transmission Routes of Patients With Confirmed SARS-CoV-2 in Southern California During the Early Stage of the US COVID-19 Pandemic. *JAMA Netw Open* 3, e2024191. (doi:10.1001/jamanetworkopen.2020.24191)
85. Walker A *et al.* 2020 Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020. *Eurosurveillance* 25. (doi:10.2807/1560-7917.es.2020.25.22.2000746)
86. Lythgoe KA *et al.* 2020 Shared SARS-CoV-2 diversity suggests localised transmission of minority variants. bioRxiv. (doi: 10.1101/2020.05.28.118992)
87. Takenouchi T *et al.* 2020 Clinical Utility of SARS-CoV-2 Whole Genome Sequencing in Deciphering Source of Infection (in press). *The Journal of Hospital Infection.* (doi: 10.1016/j.jhin.2020.10.014)
88. Korber B *et al.* 2020 Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182, 812-827.e19. (doi:10.1016/j.cell.2020.06.043)