



Original Article

HOSTED—England’s Household Transmission Evaluation Dataset: preliminary findings from a novel passive surveillance system of COVID-19

JA Hall,^{1,2} RJ Harris,³ A Zaidi,¹ SC Woodhall,¹ G Dabrera¹ and JK Dunbar^{1*}

¹COVID-19 Epidemiology Cell, Public Health England, London, United Kingdom, ²UCL Institute for Women’s Health, London, United Kingdom, and ³Statistics, Modelling and Economics Department, Public Health England, London, United Kingdom,

*Corresponding author. Public Health England, Wellington House, 133–155 Waterloo Road, London, SE1 8UG, UK.

E-mail: Feedback.C19epi@phe.gov.uk

Editorial decision 23 February 2021; Accepted 15 March 2021

Abstract

Background: Household transmission of SARS-CoV-2 is an important component of the community spread of the pandemic. Little is known about the factors associated with household transmission, at the level of the case, contact or household, or how these have varied over the course of the pandemic.

Methods: The Household Transmission Evaluation Dataset (HOSTED) is a passive surveillance system linking laboratory-confirmed COVID-19 cases to individuals living in the same household in England. We explored the risk of household transmission according to: age of case and contact, sex, region, deprivation, month and household composition between April and September 2020, building a multivariate model.

Results: In the period studied, on average, 5.5% of household contacts in England were diagnosed as cases. Household transmission was most common between adult cases and contacts of a similar age. There was some evidence of lower transmission rates to under-16s [adjusted odds ratios (aOR) 0.70, 95% confidence interval (CI) 0.66–0.74]. There were clear regional differences, with higher rates of household transmission in the north of England and the Midlands. Less deprived areas had a lower risk of household transmission. After controlling for region, there was no effect of deprivation, but houses of multiple occupancy had lower rates of household transmission [aOR 0.74 (0.66–0.83)].

Conclusions: Children are less likely to acquire SARS-CoV-2 via household transmission, and consequently there was no difference in the risk of transmission in households with children. Households in which cases could isolate effectively, such as houses of multiple occupancy, had lower rates of household transmission. Policies to support the effective isolation of cases from their household contacts could lower the level of household transmission.

Key words: SARS-CoV-2 COVID-19, household transmission, passive surveillance

Key Messages

- The Household Transmission Evaluation Dataset (HOSTED) is a passive surveillance system generating a novel data set enabling the exploration of household transmission across the SARS-CoV-2 epidemic and by characteristics of cases, contacts and households in England.
- Household transmission (based on diagnosed secondary cases) in England between June and September 2020 ranged between 4 and 6.4 per 100 household contacts.
- Children were less likely to be a secondary case and there were significant differences in the rates of transmission by region, even after controlling for household structure and property type.
- Household transmission was highest in households of over-65s and lowest in houses of multiple occupancy; households with children and multigenerational households did not have a different risk of transmission to households of two adults aged 17–64 years.
- Our findings indicate that household transmission is highest where distancing of the case from contacts is more difficult and suggest that the development of policies to support effective isolation within the home could reduce household transmission and have an important effect on the overall epidemic.

Introduction

Household transmission of SARS-CoV-2 is believed to have had an important role in the overall spread of COVID-19 cases in the community.¹ The few studies conducted show that the secondary-attack rate within the household can vary widely, from 11.8% in South Korea² and 10% in Germany³ to just 3.9% in a small study in Australia.⁴ In the UK, the secondary-attack rate in households studied under the World Health Organization FFX protocol was 37%.⁵ A recent systematic review and meta-analysis estimated the overall secondary-attack rate in households as 16.6%.⁶ The proportion of cases due to household transmission of SARS-CoV-2 will rise as effective control measures reduce community transmission, with one study showing that 70% of transmission was within households once effective community control measures were in place.² It has been argued that controlling household transmission is necessary to consistently suppress transmission rates.¹

SARS-CoV-2 has higher rates of household transmission than either SARS-CoV and Middle East respiratory syndrome coronavirus.^{6,7} Despite the demonstrated contribution of household transmission, there are no longitudinal studies examining how household transmission has varied during the pandemic and few analyses of the factors associated with household transmission, especially outside of China. Existing studies suggest that children are less likely to be secondary cases^{6,7,11,13} and that the risks are higher for spouses,^{6,13} from symptomatic cases⁶ or those with frequent daily close contact⁸ and in households with

one contact compared with those with three or more. Current household control measures for SARS-CoV-2 in England are limited to advice to minimize sharing spaces and to eat meals separately.⁹ Further evidence is required to better establish risk factors for transmission and what control measures may be effective.

Data routinely collected by diagnostic labs yield a wealth of epidemiological information that can be used to answer key questions about SARS-CoV-2 transmissions. The Household Transmission Evaluation Dataset (HOSTED) is a passive surveillance system that links laboratory-confirmed COVID-19 cases to individuals living in the same household in England. HOSTED represents a unique cohort of individuals who are assumed to share household-setting exposures. As household characteristics are included, the data can be used to explore associations between characteristics of both the household contacts and the index case, as well as the structure of the household, deprivation and region as predictors of household transmission and how these change over time. In this paper, we describe this new data set and the initial analyses of these data.

Methods

Data-set creation

HOSTED extracts data on individuals in England who share the same address as known cases of COVID-19. This is done by using a list of all laboratory-confirmed cases of COVID-19 in England from the Second-Generation

Surveillance System (SGSS). SGSS enables diagnostic laboratories in England to comply with the Health Protection (Notification) Regulations (2010) to notify Public Health England (PHE) of the identification of certain causative agents in human samples within 7 days, including SARS-CoV-2.¹⁰ National Health Service Digital (NHSD) links cases to the NHS Personal Demographics Service (PDS) in a secure data-access environment using the NHS number. The PDS contains the home address of all NHS patients in England registered with a GP and residences are indexed using the Unique Property Reference Number (UPRN). Individuals sharing the same UPRN are assumed to be sharing a household. Patient confidentiality is preserved, as NHSD provides data to PHE as a pseudonymized data set.

HOSTED is then created by excluding known institutions (using UPRN) such as care homes, prisons and households with >10 residents. This was taken as a pragmatic cut-off, with households of >10 considered to be less representative of residential households.

Data-set description

HOSTED includes data on individual-level socio-demographics of cases and household contacts including age,

sex and Index of Multiple Deprivation (IMD); information on property type, PCR-based SARS-CoV-2 testing through national reporting systems and linked information on hospitalization and mortality.

Data preparation

To generate household definitions, we classified ages according to ‘life-course events’, with children up to school age (0–16 years), young adults potentially in further education (17–24 years), working-age adults (25–64 years) and older individuals (65+ years) to broadly reflect exposures. Case and household definitions are shown in [Box 1](#).

Statistical analysis

To make formal comparisons between groups, we restricted analyses to data from 1 June 2020 onwards, when community testing, including for children, became widely available.

We explored the relationship between age of contact and age of index case through descriptive analyses, using 5-year bands (from 0–4 to 85+) to explore greater granularity than the life-course definitions used for household

Box 1. Case, contact and household definitions used in the Household Transmission Evaluation Dataset

Case definitions

Index case: The earliest case of laboratory-confirmed COVID-19, by diagnosis date, for a household

Co-primary cases: Cases of laboratory-confirmed COVID-19 occurring within 1 day of each other

Secondary case: A known household contact of an index or co-primary case with a positive SARS-CoV-2 test who has a specimen date between 2 and 14 days after the specimen date of the index or co-primary case

Contact definition

Household contact: All individuals with the same address, according to NHS registration records, as the index or co-primary cases of COVID-19

Household definitions

Household: A residential address that is not a known institution (e.g. care home, prison, university residence) and where a maximum of 10 people (including index case) are registered at the address

Index household: A household in which an index case has been identified

Adult pair: A household with two adults aged 17–64 years

Older pair: A household with two adults where at least one is aged ≥ 65 years

Household with children: Households with at least one child aged ≤ 16 years and at least one adult aged ≥ 17 years and not multigenerational

Multigenerational household: a household with at least three individuals, with at least one child aged ≤ 16 years and one adult aged ≥ 60 years AND with at least a 16-year gap in age

House of multiple occupancy: a household with three or more adults

categorization and analysis. Data were restricted to households with a single index case (and no co-primary cases) to clearly define the age of the index case. The proportion of secondary cases occurring in each index case/contact-age combination was then calculated and plotted as a ‘heat map’.

We estimated the odds of becoming a secondary case within 2 weeks of the index case in different groups via logistic-regression analysis, providing odds ratios (ORs) and 95% confidence intervals (CIs), with robust standard errors to account for clustering within households. Variables included: date of index case in half-month (roughly 15-day) bands; life-course age group; sex, government-office region, IMD quintile (1 = most deprived, 5 = least deprived), property type (semi-detached, terraced, flat, detached, other) and household type with adult pairs (‘couples’). We also explored the effect of multiple index cases within a household and the effect of household size in households consisting of three or more people. Univariable and multivariable models were fitted. All possible two-way interactions between variables were considered and compared with the base model via likelihood ratio tests. Any interactions found to be important were reported separately and simplified where possible.

To further assess regional differences, we conducted a nested case-control study, matching secondary cases to non-cases on: life-course age group, sex, month, IMD quintile and household composition. The latter included the number of people of different age groups living in the same household, and whether they were a contact or index/co-primary case. Due to the rarity of identical matching factors, in particular the age composition of larger households, this analysis was restricted to households of five or fewer. This analysis was intended to balance any confounding factors within the region.

Results

Description of the data set

Data were extracted on 30 September 2020 and included individuals with ≥ 14 days of observable follow-up from the date of specimen collection from the index case (i.e. index cases between 1 June and 14 September 2020, secondary cases up to 28 September 2020). As shown in [Figure 1](#), HOSTED contains 225 254 people living households of 2–10 people where there was at least one case. There were 70 835 cases (index, co-primary and secondary) and 164 169 contacts.

From 1 June to 28 September 2020, 8957 of 164 169 household contacts became secondary cases. There were 4632 households that had one secondary case (8.5%),

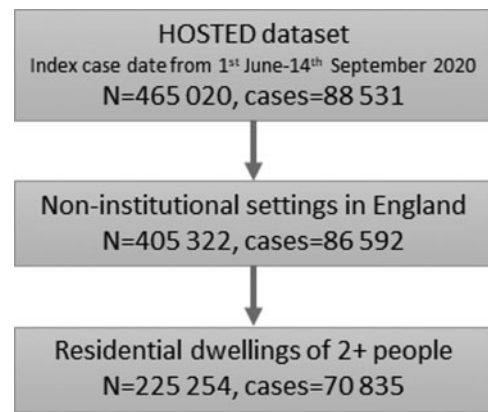


Figure 1 Flow chart of total linked individuals and confirmed COVID-19 cases in HOSTED

N, number of people in households with at least one case; cases, all cases: index, co-primary and secondary.

1235 had two secondary cases (2.3%) and 551 had three or more (1.0%). The median time from the specimen date of the index case to the specimen date of the secondary case was 4 days (IQR 2–7 days). There were 70 835 cases in total, so 12.6% of cases in HOSTED were secondary cases. Of the index/co-primary cases, 6096 (8.6%) were co-primary cases.

Overall, households had a median of four people (IQR 3–6); families with children and multigenerational households were slightly larger. Households in less deprived areas tended to be slightly smaller (median 3.8) than those in the most deprived areas (median 4.5). London had the highest average household size (median 4.5) and the largest proportion of multigenerational households or houses of multiple occupancy. More detail is shown in [Supplementary Table S1](#), available as [Supplementary data](#) at *IJE* online.

Descriptive epidemiology of household transmission

Household transmission was highest in the week of 8 June at 6.4 secondary cases per 100 household contacts, as shown in [Figure 2](#). Household transmission fell during June, was stable at around five throughout July and August, and then appears to have increased from mid-August. The ratio of cases to contacts was stable at 2.5–3.0 contacts per case.

The proportion of contacts who become secondary cases was highest where the age of the contact was close to the age of the index case, particularly in those over 30 years old, as shown in [Figure 3](#). There is a ‘hotspot’ of transmission between older parents (40–59 years) and older children (15–24 years) and a small number of cases,

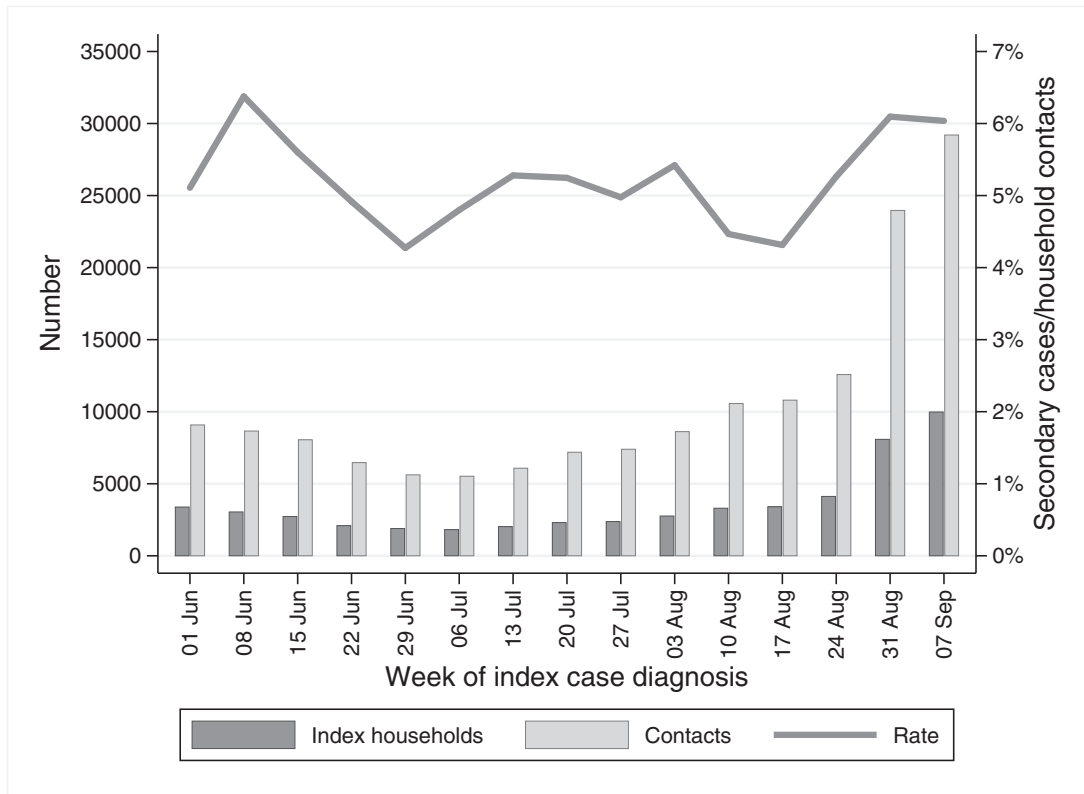


Figure 2 Number of index cases, household contacts and secondary-transmission rate for index cases in HOSTED from 1 June to 14 September 2020.

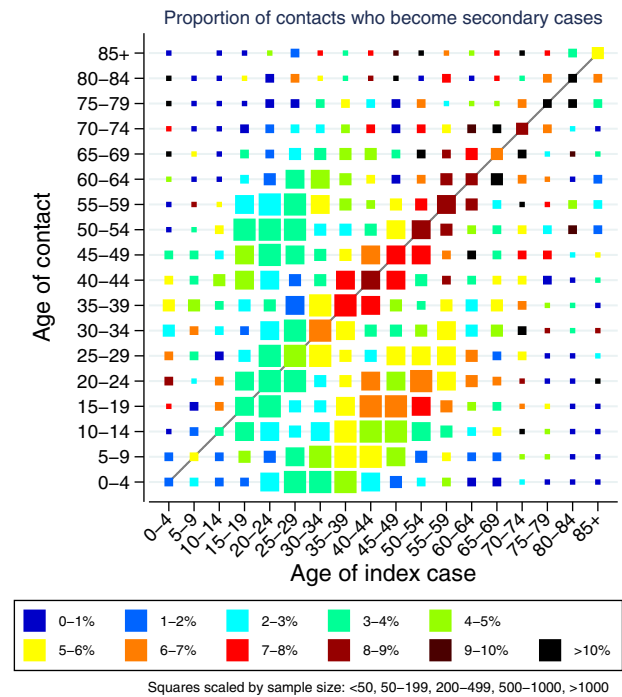


Figure 3 Heat map of the proportion of contacts who become secondary cases according to both the age of index case and age of contact. Red to black shading indicates higher proportions, greens and blue low proportions. Squares are scaled according to sample size.

but high levels of transmission, between adults (40–49 years) and older people (80+ years), which may reflect adults living with elderly parents. Children, particularly the under-fives, have a low risk of being a secondary case, regardless of the age of the index case. The OR for adult-to-child transmission vs child-to-adult transmission in households with children was 1.01 (95% CI 0.85–1.19), indicating no evidence of a difference in the direction of transmission.

There has been a consistent difference between regions, with the Yorkshire and the Humber, the North West, the Midlands and, aside from a small dip in June, the North East having higher rates of secondary transmission, as shown in Figure 4.

Logistic regression

The univariable and multivariable results were generally similar, as shown in Table 1. Odds of being a secondary case were lower in the 0–16 age group (adjusted odds ratios (aOR) 0.70, 95% CI 0.66–0.74) and females had higher odds of being a secondary case (aOR 1.32, 95% CI 1.26–1.38). The differences by region remained. There were lower rates of household transmission in less deprived areas on univariate analysis, but no effect was observed

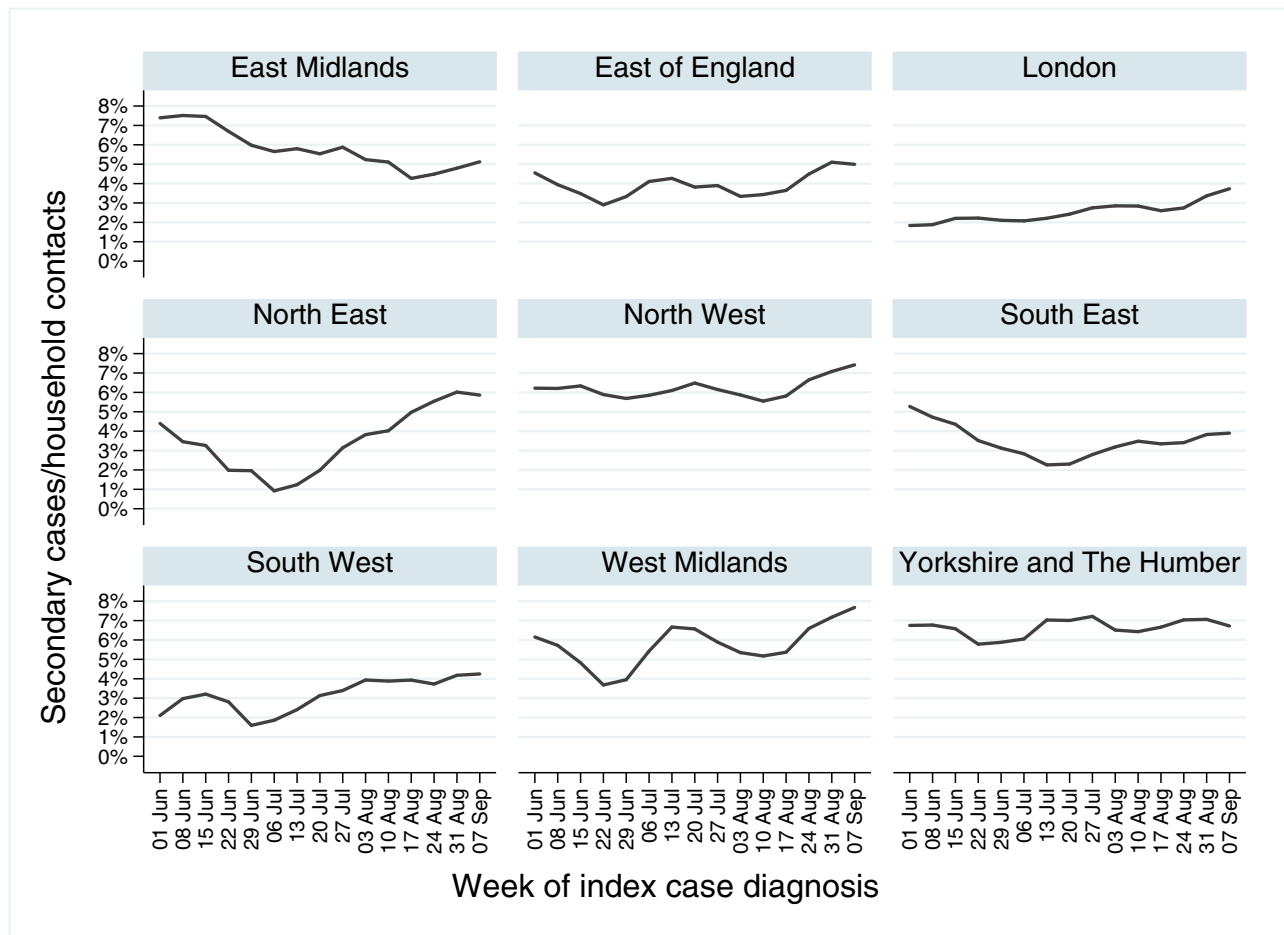


Figure 4 Proportion of household contacts becoming secondary cases by region for index cases in HOSTED from 1 June to 14 September 2020. Three-week rolling averages are shown to smooth the effects of variation due to small numbers.

after adjusting for other variables, in particular, region. Older-pair households had a 25% increase in the odds of household transmission (after adjusting for age) vs adult-pair households (aOR 1.25, 95% CI 1.04–1.49) and households of multiple occupancy had lower odds of household transmission (aOR 0.74, 95% CI 0.66–0.83). The risk in households with children was not different to that in adult-pair households. Contacts living in flats had lower odds of being a case (aOR 0.62, 95% CI 0.54–0.72), but there were no differences between other defined property types.

There were interactions between several variables, e.g. the increased odds for being female were highest in the 17–24 and 25–64 age groups, and no different to males in the 65+ group. As seen in [Figure 4](#), the effect of region varied over time.

In households with two index cases, the adjusted OR for contacts becoming a secondary case was 2.11 (95% CI 1.93–2.31) compared with those with one index case and 2.23 (95% CI 1.91–2.61) for households with three or

more index cases. In households with three or more people, there were lower odds of being a secondary case in larger households. ORs decreased monotonically: for households of five vs three people, the adjusted OR was 0.92 (95% CI 0.85–0.99); for seven vs three, the OR was 0.79 (95% CI 0.72–0.88); and, for nine vs three, the OR was 0.71 (95% CI 0.63–0.81). The reduction was more pronounced in households with children and less so in multigenerational households. Whereas larger households were more likely to have secondary cases (17.6% of households with 5–10 people had a secondary case compared with 9.3% of households with 2–4 people), the risk to each individual was lower (5.2% vs 6.1% of contacts became secondary cases; see [Supplementary Table S3](#), available as [Supplementary data at IJE online](#)).

Nested case-control study

The matched case-control analysis included 4555 secondary cases and the same number of non-cases. The results of

Table 1 Adjusted (multivariable model) and results (odds ratios with 95% confidence intervals)

	Unadjusted	Adjusted
Age (years)		
0–16	0.81 (0.76–0.86)	0.70 (0.66–0.74)
17–24	1.06 (0.99–1.13)	1.04 (0.97–1.11)
25–64	1 (base)	1 (base)
65+	1.12 (1.02–1.22)	1.02 (0.92–1.14)
Gender		
Male	1 (base)	1 (base)
Female	1.33 (1.27–1.38)	1.32 (1.26–1.38)
Region		
East Midlands	1.67 (1.45–1.92)	1.64 (1.43–1.89)
East of England	1.15 (0.97–1.36)	1.15 (0.97–1.36)
London	0.79 (0.68–0.93)	0.87 (0.74–1.03)
North East	1.44 (1.21–1.71)	1.34 (1.12–1.59)
North West	1.86 (1.63–2.11)	1.80 (1.58–2.05)
South East	1 (base)	1 (base)
South West	0.98 (0.79–1.21)	0.98 (0.79–1.22)
West Midlands	1.80 (1.57–2.07)	1.76 (1.53–2.03)
Yorkshire & Humberside	1.90 (1.66–2.17)	1.86 (1.63–2.13)
Index of Multiple Deprivation quintile		
1	1 (base)	1 (base)
2	0.94 (0.87–1.01)	1.08 (0.99–1.16)
3	0.86 (0.79–0.94)	1.01 (0.92–1.11)
4	0.83 (0.76–0.91)	0.92 (0.84–1.02)
5	0.88 (0.80–0.96)	0.98 (0.88–1.08)
Household type		
Adult pair/couple	1 (base)	1 (base)
Older pair/couple	1.30 (1.11–1.53)	1.25 (1.04–1.49)
HH with children	0.88 (0.79–0.99)	0.96 (0.86–1.08)
Multigenerational household	0.93 (0.81–1.06)	0.97 (0.85–1.11)
Adult household of 3 or more	0.76 (0.68–0.85)	0.74 (0.66–0.83)
Property type		
Semi-detached	1 (base)	1 (base)
Terraced	0.95 (0.89–1.02)	0.96 (0.90–1.03)
Flat	0.49 (0.43–0.56)	0.62 (0.54–0.72)
Detached	1.00 (0.92–1.09)	1.05 (0.96–1.15)
Other	0.63 (0.54–0.74)	0.70 (0.60–0.82)
Index case date—2020		
01 Jun–15 Jun	1 (base)	1 (base)
16 Jun–30 Jun	0.87 (0.77–0.99)	0.85 (0.75–0.96)
01 Jul–15 Jul	0.85 (0.75–0.98)	0.87 (0.76–0.99)
16 Jul–31 Jul	0.85 (0.75–0.96)	0.87 (0.77–0.98)
01 Aug–15 Aug	0.86 (0.76–0.96)	0.86 (0.77–0.97)
16 Aug–31 Aug	0.87 (0.78–0.97)	0.96 (0.86–1.07)
01 Sep–13 Sep	1.05 (0.96–1.15)	1.07 (0.98–1.17)

the analysis showed the same pattern as the multivariable model; the higher risk in the East Midlands, North West, North East and West Midlands persisted ([Supplementary Table S2](#), available as [Supplementary data](#) at *IJE* online).

Discussion

Overall, the rate of household transmission in this passive surveillance system was 5.5 secondary cases per 100 household contacts. Household transmission was highest in the week of 8 June 2020 at 6.4, falling to just over 4% in mid-August and then appearing to start increasing. This is lower than that seen in studies of household transmission and likely reflects a difference in methodology, i.e. passive surveillance vs whole-household testing vs antibody testing.

There was evidence that children were less likely to be a secondary case (aOR 0.70, 95% CI 0.66–0.74), as other studies have found.^{6,13} Consequently, there was no evidence that households with children were more likely to experience household transmission. There also appeared to be no difference in the odds of a child index case passing it to an adult secondary case or vice versa. However, in addition to the implications of the testing strategy, which limited access to tests for children, children may also be under-represented in HOSTED if they are more likely to be asymptomatic¹¹ or test false negative on PCR¹² and so could have been index cases that were not detected and had no risk of becoming a secondary case.

Household transmission appeared to be most likely among co-habiting adults in the same or neighbouring age bands. Houses of multiple occupancy had the lowest risk of household transmission (aOR 0.74, 95% CI 0.66–0.83). This may be because adults living in these circumstances have less close contact with other household members than either a couple or a family has with each other. For example, adults sleeping in separate bedrooms would have a lower risk of transmission than those sharing a bed. This is in keeping with other studies, which found higher rates of transmission between spouses^{6,13} and those with regular contact.⁸ The highest risk of transmission was in older-pair households, i.e. two people aged over 65 years, which may be related to the increased susceptibility of this age group;¹⁴ there was no evidence that transmission was more likely in multigenerational households.

We saw higher rates of household transmission in the North West, North East, Yorkshire and the Humber and the Midlands, even after adjustment for other factors in the multivariable model and in the regional case-control study. This suggests that there is an effect of region beyond the observed small differences in IMD, household structure or property type and could be related to the higher incidence rates seen in these areas over the course of the pandemic if some community acquisition is incorrectly being assigned to household transmission.

With regard to sex, women were only more likely to be secondary cases if aged from 17 to 65 years, which may be

because they are more likely to be caring for children or their older parents. People aged over 65 years had lower risks if they were in what was defined as a house of multiple occupancy compared with those living with one other >65 years old (aOR 0.57, 95% CI 0.47–0.71). This may suggest that they were living in some form of sheltered accommodation that had not been correctly identified in the UPRN and were essentially living alone. The secondary-attack rate was higher the higher the number of co-primary cases, as the contacts would have had increased risk of exposure.

Given that 12.6% of all cases in HOSTED were secondary cases acquired by household transmission, this is an important dimension of the epidemic to address. Our findings suggest that individuals in household settings where it is possible to maintain social distancing within the household, such as houses of multiple occupancy, do have lower risks of household transmission. Advice to households could aim to replicate these conditions, e.g. the case should sleep in a separate room. However, maintaining social distancing within the household is not always practical for families, particularly those with members who require care, and this has important mental health implications for all ages. There is also the risk that exposure will have occurred prior to diagnosis and so it may be too late to have an impact. However, one study showed that, where cases successfully quarantined from symptom onset (including wearing a mask, eating separately and sleeping alone), the risk of household transmission was reduced to zero.¹³ Another policy, pursued in China, was the relocation of cases to isolation facilities,¹⁵ but this would likely be unacceptable to the UK population. Therefore, limiting the spread of COVID within households with the case(s) *in situ*, where this is possible, is an important policy avenue to explore.

Given the likelihood of household transmission, consideration could be given to taking a household approach to cases. This would reduce the burden on the household and on contact tracers, by managing the case and all household contacts at once rather than on a contact-by-contact basis.

Limitations

The main limitation of HOSTED is that, as a passive surveillance system, we cannot calculate a true secondary-attack rate equivalent to an intensive testing protocol such as the WHO FFX studies. We only have information on diagnosed secondary cases, which will be an underestimate as not everyone develops symptoms and, of those that do, not everyone can or will access a test. However, as testing strategies were national and free at the point of access, there is no reason to believe that there are systematic

differences in the uptake of testing by region or other factors or how these are recorded in this novel data set. Therefore, we are able to use these data as a stable representative indicator of household transmission to make comparisons of the risk of household transmission by these factors and over time. However, this analysis has been conducted on data from when the daily case numbers in England were relatively low: after the first wave and before the second or third, and when ‘stay-at-home’ orders were not in effect. The lack of widespread testing in the first wave (March–April 2020) means that we are not able to estimate a household transmission rate for this time. The implications of using these data may be an underestimate of the household SAR if stay-at-home orders increase household transmission. Conversely, conducting the analysis at a time when the levels of SARS-CoV-2 were low in the community reduces the risk of misclassifying two separate community acquisitions of SARS-CoV-2 as household transmission. Analysis of household transmission in the second wave and subsequent lockdown, and of the impact of new variants, is underway.

Our definition of ‘secondary’ cases may have led to the inclusion of some cases that were in fact co-primary. As the incubation period for SARS-CoV-2 is thought to be 2–14 days,^{16,17} we chose a threshold of 2 days between the specimen dates of the index and secondary cases as a pragmatic decision to offset the risks of misclassifying cases as either co-primary or as secondary. A sensitivity analysis using a cut-off of 4 days found to have little effect on any of the results or on the multivariable model. We have also made the implicit assumption that two (or more) cases occurring in a household within 2–14 days represents household transmission when it also plausible that they are two independent community-acquired infections. However, given the timing of this study, when the number of cases was at its lowest, this risk is probably small.

The definitions of household structure that we have used are broad and cannot realize the entire the multitude of living arrangements of households in England. However, we did explore several definitions of multigenerational households and our findings were robust to these. Further, definitions are based only on the observed ages of those in the households. We make assumptions about the relationships of two adults with children, two adults of a similar age, etc. living in a household; the exact living arrangements and familial relationships are not known. We are also limited to those individuals that are on the PDS database and can be linked by NHS number; therefore, some misclassification of household types, cases and contacts is possible. Our choice of a limit to household sizes of 10 was somewhat arbitrary but, with increasing household size, the risk of misclassification of a residence

as an institute becomes more likely. Overall, <2% of 'households' contained 11–20 people: these were most common in London. Larger households tended to have lower IMD scores, but this trend flattened in households of >10 people. A sensitivity analysis confirmed that our findings were the same if we did include households of ≤ 20 .

It has been clear since early in the pandemic that there are inequalities in the burden of COVID-19 by Black, Asian and Minority Ethnic (BAME) groups and there is a great need to understand patterns of transmission by ethnicity. We do not have information on the ethnicity of cases and contacts in HOSTED currently, but we are exploring ways to address this. We were also not able to examine differences by symptom status, as this was not collected initially and has yet to be added to HOSTED.

Conclusion

There has been considerable interest in the role of household transmission in relation to other modes of transmission particularly in the context of stay-at-home orders. Understanding transmission trends and patterns in household settings is an essential component of the current epidemiology of and response to the pandemic. HOSTED is a unique data set; no other country has created anything like it. These data have shown, by linking cases to household contacts, that 4.3–6.4% of contacts become confirmed cases during a time of relatively low activity in the COVID-19 epidemic in England. This stable electronic health records data set can be used to compare rates of laboratory-confirmed household transmission across several other factors and to use household transmission as an indicator of trends in population-infection activity more generally. From this, we can see that transmission is highest in adults between cases and contacts of similar ages, and that children are less likely to be secondary cases. On a positive note, our findings demonstrate that household transmission is not inevitable. Households in which the isolation of cases is possible had lower rates of household transmission. Given the consistent rate of household transmission observed over this time period, there is a need to consider further interventions to modify this route of transmission.

Supplementary data

Supplementary data are available at *IJE* online.

Ethics approval

The HOSTED surveillance system was reviewed and approved by the PHE Research Ethics Governance Group. The data were collected and linked by NHS Digital. The

data were processed lawfully under GDPR Article 6(1)e and 9(2)i, and shared under Regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002.

Funding

This work was undertaken as part of the core functions of Public Health England in relation to the surveillance of communicable diseases and outbreak response. Funding: None.

Data availability

The data underlying this article cannot be shared publicly due to the legal and policy controls placed on data used as part of the government's response to the COVID-19 pandemic.

Acknowledgements

The authors wish to thank Heather Pinches, Phillip Bowker and Paul Ellingham as well as the rest of the NHS Digital team whose hard work and dedication enabled the creation of these data during the UK COVID-19 pandemic response.

Conflict of interest

None declared.

References

1. Haroon S, Chandan JS, Middleton J, Cheng KK. Covid-19: breaking the chain of household transmission. *Bmj* 2020;370: m3181.
2. Shen M, Peng Z, Guo Y *et al.* Assessing the effects of metropolitan-wide quarantine on the spread of COVID-19 in public space and households. *Int J Infect Dis* 2020;96:503–05.
3. Böhmer MM, Buchholz U, Corman VM *et al.* Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect Dis* 2020; 20:920–28.
4. Draper AD, Dempsey KE, Boyd RH *et al.* The first 2 months of COVID-19 contact tracing in the Northern Territory of Australia, March–April 2020. *Commun Dis Intell* 2020;44:1–8.
5. Boddington NL, Charlett A, Elgohari S *et al.* COVID-19 in Great Britain: epidemiological and clinical characteristics of the first few hundred (FF100) cases: a descriptive case series and case control analysis. *Bull World Health Organ* 2021;99:178–89.
6. Madewell ZJ, Yang Y, Longini IM, Jr *et al.* Household transmission of SARS-CoV-2: A systematic review and meta-analysis. *JAMA Netw Open* 2020;3:e2031756.
7. Jing Q-L, Liu M-J, Zhang Z-B *et al.* Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect Dis* 2020;20:1141–50.
8. Wang Y, Tian H, Zhang L *et al.* Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. *BMJ Glob Health* 2020;5:e002794.

9. UK Government. *COVID-19: Guidance for Households with Possible Coronavirus Infection*. 2020. <https://www.gov.uk/government/publications/covid-19-stay-at-home-guidance> (12 November 2020, date last accessed).
10. PHE Laboratory Reporting Guidelines. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/739854/PHE_Laboratory_Reporting_Guidelines.pdf (10 October 2020, date last accessed).
11. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019: a systematic review and meta-analysis. *J Med Virol* 2021;93:820–30.
12. Dattner I, Goldberg Y, Katriel G *et al*. The role of children in the spread of COVID-19: using household data from Bnei Brak, Israel, to Estimate the Relative Susceptibility and Infectivity of Children. *MedRxiv*. doi:10.1101/2020.06.03.20121145, 11th October 2020, pre-print: not peer-reviewed.
13. Li W, Zhang B, Lu J *et al*. Characteristics of household transmission of COVID-19. *Clin Infect Dis* 2020;71:1943–46.
14. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. *Bmj* 2020;368:m1198.
15. Chen S, Zhang Z, Yang J *et al*. Fangcang shelter hospitals: a novel concept for responding to public health emergencies. *Lancet* 2020;395:1305–14.
16. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China. *Euro Surveill* 2020;25:20.
17. Linton NM, Kobayashi T, Yang Y *et al*. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med* 2020;9:538.