medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Association between SARS-CoV-2 Transmission Risk, Viral Load, and Age: A Nationwide Study in Danish Households*

Frederik Plesner Lyngse[1,2,3,†],

Kåre Mølbak[3,4],

Kristina Træholt Frank[3],

Claus Nielsen[3],

Robert Leo Skov[3],

Carsten Thure Kirkeby[4]

February 28, 2021

*We thank Statens Serum Institut and The Danish Health Data Authority for data access and helpful institutional knowledge. We also thank the rest of the Expert Group for Mathematical Modelling of COVID-19 at Statens Serum Institut.

†Correspondence to Frederik Plesner Lyngse, fpl@econ.ku.dk. Affiliations: [1]Department of Economics & Center for Economic Behaviour and Inequality, University of Copenhagen; [2]Danish Ministry of Health; [3]Statens Serum Institut; [4]Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1 Abstract

Aim The objective of this observational study was to investigate the association between SARS-CoV-2 transmission risk, RT-PCR Cycle threshold (Ct) values, and age of infected cases in Danish households.

Background The Covid-19 pandemic is one of the most serious global public health threats in recent times. Understanding transmission of SARS-CoV-2 is of utmost importance to be able to respond to outbreaks and take action against the spread of the disease. Viral load is generally thought to correlate with transmission risk.

Methods We used comprehensive administrative register data from Denmark, comprising the full population and all SARS-CoV-2 tests (August 25, 2020 to February 10, 2021), to estimate household transmission risk.

Results We found that the transmission risk was negatively associated—approximately linear—with the Ct values of the tested primary cases. Also, we found that even for relatively high Ct values, the risk of transmission was not negligible; e.g., for primary cases with a Ct value of 38, we found a transmission risk of 8%. This implies that there is no obvious cut-off for Ct values for risk of transmission. We estimated the transmission risk according to age and found an almost linearly increasing transmission risk with the age of the primary cases for adults ($\geq$20 years) and negatively for children ($<$20 years). Age had a higher impact than Ct value on the risk of transmission.

Conclusions Lower Ct values (indicating higher viral load) are associated with higher risk of SARS-CoV-2 transmission. However, even at high Ct values, transmission occurs. In addition, we found a strong association between age and transmission risk, and this dominated the Ct value association.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2 Introduction

The world is in the midst of a pandemic caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). It is essential to understand the SARS-CoV-2 transmission dynamics and associated factors in order to design interventions to control the spread of the disease, such as contact tracing efforts.

Real-time reverse transcription polymerase chain reaction (RT-PCR) is used worldwide to detect SARS-CoV-2 in swabs from nasopharynx and oropharynx (Corman et al., 2020). These tests are useful for fast and reliable detection of infected individuals who should then isolate themselves to prevent further spread of the disease (Cheng et al., 2020). Detection of SARS-CoV-2 by RT-PCR is—similar to other viruses—dependent on the viral load and it is reversely correlated with the Cycle threshold (Ct) value of a test (Singanayagam et al., 2020). The viral load changes during the course of disease and has been shown to be highest around the time of symptom onset (He et al., 2020).

Recently, an epidemiological study of 282 primary cases in Spain showed that higher viral load was associated with increased transmission risk of SARS-CoV-2 (Marks et al., 2021). The results were confirmed in a study of 219,722 primary cases in the UK (Lee et al., 2021). Thus, it is argued that contact tracing should be prioritized on cases with low Ct values (high viral load), as they will have a higher risk of generating secondary cases.

The aim of this observational study was to investigate the association between the transmission risk of SARS-CoV-2, RT-PCR Ct values, and age of infected cases in Danish households. In Denmark, all residents have access to tax-paid universal health insurance, and a test for SARS-CoV-2 is free of charge. The testing capacity is high, and testing is widespread. Denmark also has comprehensive social insurance, and SARS-CoV-2 sick leave is fully reimbursed by the state. Thus, neither financial reasons nor access to tests were a major obstacle to obtaining a test during our study.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

3 Data and Methods

During the study period, RT-PCR tests for SARS-CoV-2 could be obtained from either community testing facilities at TestCenter Denmark (TCDK) or from hospitals. Statens Serum Institut (SSI) analyses all tests from TCDK. Information on Ct values was only available for samples that tested positive at TCDK.

3.1 Register Data

In this study we used Danish administrative register data comprising the full population. All residents in Denmark have a unique personal identification number that allows a complete linkage of information across different registers at the individual level. All laboratory results from departments of clinical microbiology in Denmark are registered in the Danish Microbiology Database (MiBA), from which we obtained individual level data on all national tests for SARS-CoV-2 for the period August 25, 2020 (which was the first date with accessible Ct values) to February 10, 2021. Primary cases were only included until January 25, 2021 in order to allow for secondary cases to present within the following 14 days. Moreover, we only included primary cases identified by TCDK, as we only had Ct values on those case samples. For potential secondary cases, we include all tests, regardless of them being tested at TCDK or hospitals.

Information on the reason for being tested (e.g., symptoms, potential contact with infected persons etc.) was not available. From the Danish Civil Registry System, we obtained information about the sex, age, and home address for all individuals living in Denmark. People who tested positive for SARS-CoV-2 using antigen tests were not included, as these test results were not transferred to MiBA at the time of the study.

3.2 Data Linkage

We constructed households by linking all individuals living at the same address, and we only considered households with two to six members, in order to exclude, e.g., institutions. Thus, six single apartments in the same building counted as six independent households.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Person-level data, which included information on the test results, dates and times of sampling as well as the times the results were available, were linked to individuals within households. For each household, we identified the first positive test for SARS-CoV-2; called the primary case throughout this paper. We considered all subsequent tests from other members in the same household as tests taken in response to the primary case. We defined secondary cases as those who had a positive test with sampling dates 1 to 14 days after the primary case tested positive. Thus, we defined cases that tested positive on the same day as co-primary cases, thereby excluding them as potential secondary cases. For a thorough discussion on co-primary cases, see Lyngse et al. (2020). In addition, we assumed that all identified secondary cases were infected by the primary case within the same household, because all other household members should isolate themselves from society once a primary case within a household is confirmed. Thus, the primary case is the most likely source of infection for any additional household members.

3.3 Laboratory Analyses

Analysis of tests of TCDK is performed using a set of primers that target the E-gene on SARS-CoV-2 (Corman et al., 2020), which is recommended by the WHO and the ECDC, and has a high sensitivity and specificity (Vogels et al., 2020). TCDK have used the same methodology and primers throughout the epidemic, making the Ct values comparable across the study period. An RT-PCR test for SARS-CoV-2 is defined as positive if the Ct value is $\leq$38 (SSI, 2021).

3.4 Statistical Analyses

The transmission risk of the primary case (p) is the outcome variable y_p; defined as the proportion of potential secondary cases that tested positive.

To estimate the association between Ct values and transmission risk, we estimated the non-parametric regression equation:

$$y_p = \beta \times Ct_{p,1} + \varepsilon_p \,, \tag{1}$$

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

where $Ct_{p,1}$ is the Ct value (rounded to the nearest integer) of the primary case. β measures the transmission risk for each Ct value. ε_p denotes the error term, clustered on the household (event) level.

To estimate the association between age and transmission risk, stratified by the median Ct value (28), we estimated the non-parametric regression equation:

$$y_p = \beta \times Age_{p,5} + \varepsilon_p \,, \tag{2}$$

where $Age_{p,5}$ is the age (in five-year groups) of the primary case. β measures the transmission risk for each five-year age group of the primary cases. ε_p denotes the error term, clustered on the household (event) level.

To estimate the association between Ct value, age, and transmission risk, we estimated the non-parametric regression equation:

$$y_p = \beta \times Ct_{p,2} \times Age_{p,10} + \alpha \times Age_s + \varepsilon_p \,, \tag{3}$$

where $Ct_{p,2}$ is the Ct value (in bi-value groups) and $Age_{p,10}$ is the age (in ten-year groups) of the primary case. β measures the transmission risk of the interaction between Ct value and age of the primary case. Age_s is the age of the potential secondary cases (s) and α measures the linear association with age of the potential secondary cases (Lyngse et al., 2020). ε_p denotes the error term, clustered on the household (event) level.

To quantify the increased transmission risk across different observable characteristics, we estimated a univariable and a multivariable logistic regression. In particular, to estimate the odds ratio, we estimated the logistic regression equation:

$$log\left(\frac{y_p}{1-y_p}\right) = \beta \times Ct_{p,2} + \gamma \times Age_{p,5} + \phi \times Female_p + \delta \times HouseholdSize_p + \varepsilon_p \,, \tag{4}$$

where β measures the non-parametric association with Ct values, γ measures the non-parametric association with age of the primary case, ϕ measures the association with sex,

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

and δ measures the association with the size of the household. ε_p denotes the error term, clustered on the household (event) level.

3.4.1 Sensitivity analyses

To investigate whether age had an impact on the result, we estimated the association between Ct value and transmission risk, controlling for age of the primary case. Furthermore, we estimated the age structured transmission risk stratified by sex to see whether there were different patterns across men and women. We then estimated the age structured transmission risk stratified by Ct value quartiles to see whether the pattern was independent of the median Ct value cut-off.

As Ct values were only available for the primary cases that were identified by being tested in TCDK, only these primary cases were included in the analyses. To address the potential bias from not including primary cases that were identified at hospitals, we performed sensitivity analyses by estimating transmission risk stratified by TCDK and hospitals.

We also estimated the transmission risk of the interaction between Ct value and age of the primary case without controlling for age of the potential secondary case to see whether our results were driven by the age of the potential secondary cases. Because people normally live with a partner around their own age and parents with their children, susceptibility correlation with age could drive an age structured transmission risk. To address this potential bias, we stratified our sample by Ct value quartiles and estimated the transmission risk with the interaction of the age of the primary and potential secondary case.

3.5 Ethical statement

This study was conducted on administrative register data. According to Danish law, ethics approval is not needed for such research. All data management and analyses were carried out on the Danish Health Data Authority's restricted research servers with project number FSEID-00004942. The project was notified to the Danish Data Protection Agency.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

4 Results

4.1 Descriptive Statistics

From August 25, 2020 to February 10, 2021, TCDK analyzed 9,416,298 samples for SARS-CoV-2 (74% of all tests in Denmark) and identified 66,602 primary cases up to January 25 (73% of all primary cases) (Table S1). Of these primary cases, a Ct value was available for 99.6%.

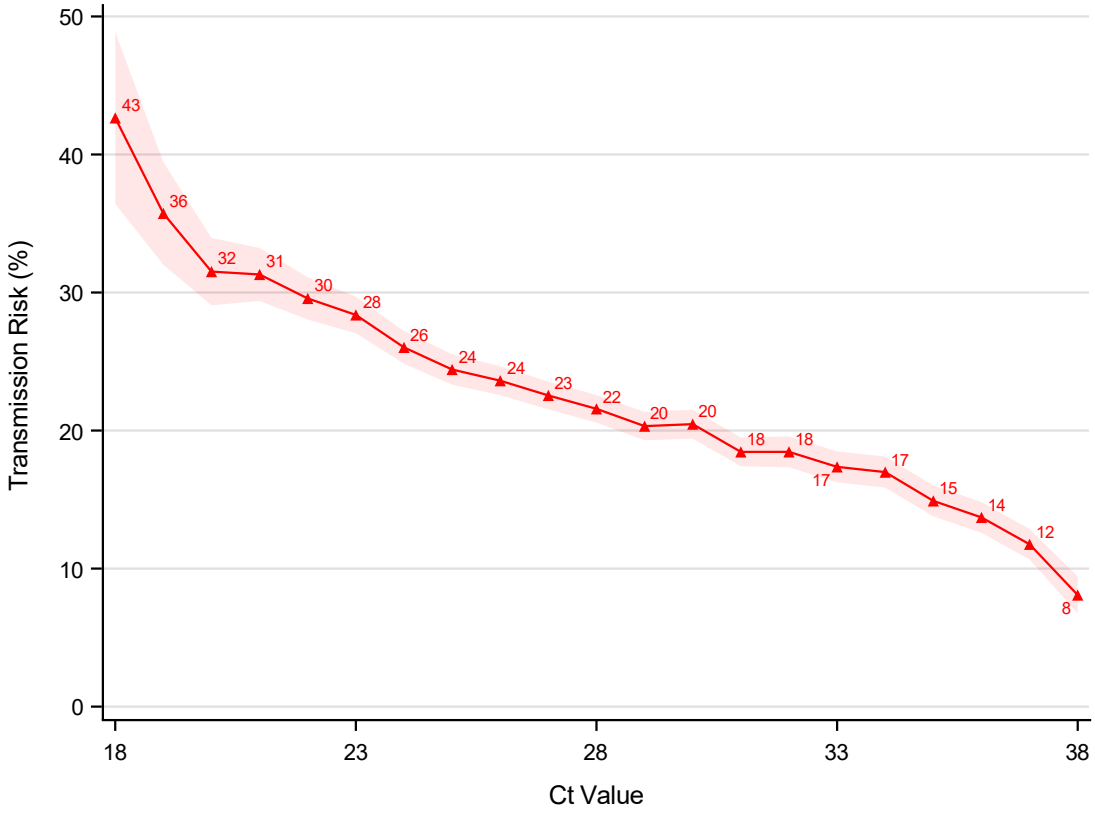
In this study, we had 66,311 primary cases living with 213,576 potential secondary cases, of which 103,389 (48%) tested positive for SARS-CoV-2 and were thus actual secondary cases (Table S2). Approximately half were men and half women. 25% of primary cases had a Ct value $\leq$25, 50% had a Ct value $\leq$28, and 75% had a Ct value $\leq$32 (Figure S1). The distribution of Ct values was relatively similar across age groups, suggesting that differences in test strategy across age were not driving our results (Figure S2 and Table S4).

4.2 Associations with Transmission Risk

Figure 1 shows the association between Ct values and transmission risk. There is an approximately linear decreasing relationship between Ct values and transmission risk.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 1: Association between Ct Values and Transmission Risk

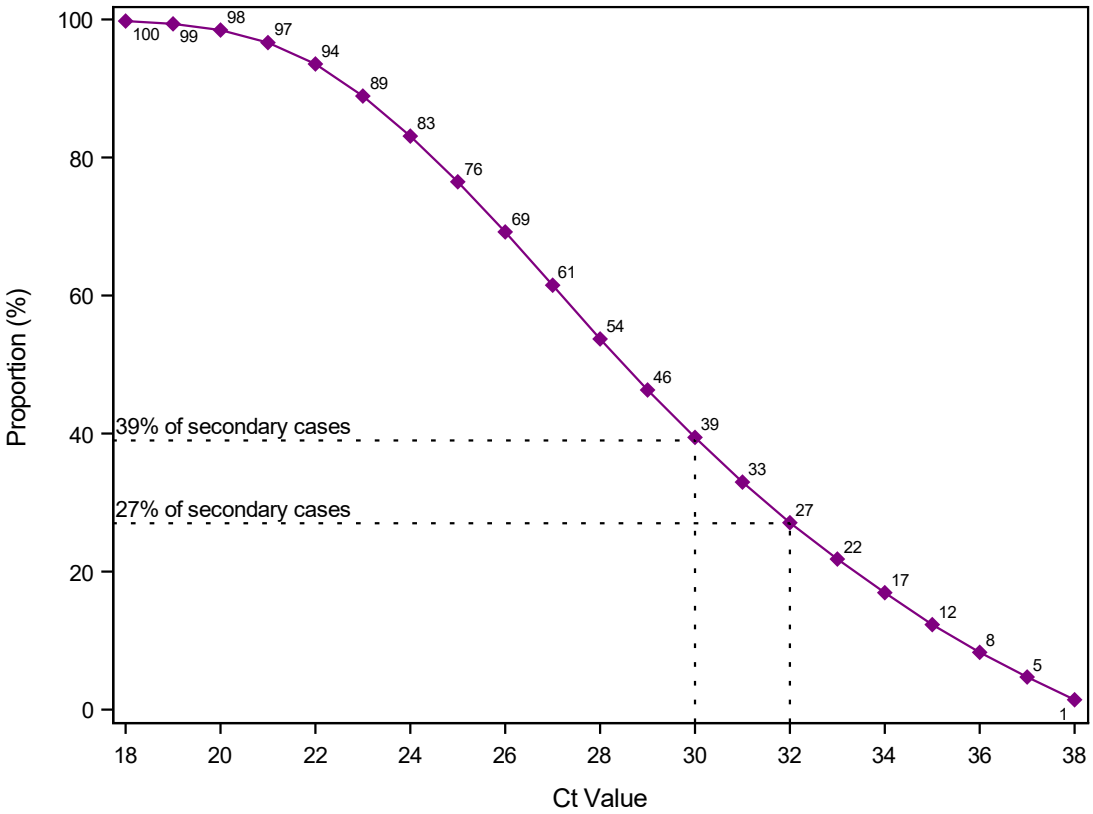


Notes: A primary case with a Ct value of 18 has a transmission risk of 43%, and a primary case with a Ct value of 38 has a transmission risk of 8%. The estimates come from estimating regression equation 1. The shaded area shows the 95% confidence bands clustered on the household level.

Figure 2 shows the cumulative distribution of secondary cases by Ct values, e.g., primary cases with Ct values $\geq$30 account for 39% secondary cases.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 2: Cumulative Distribution of Secondary Cases by Ct Values

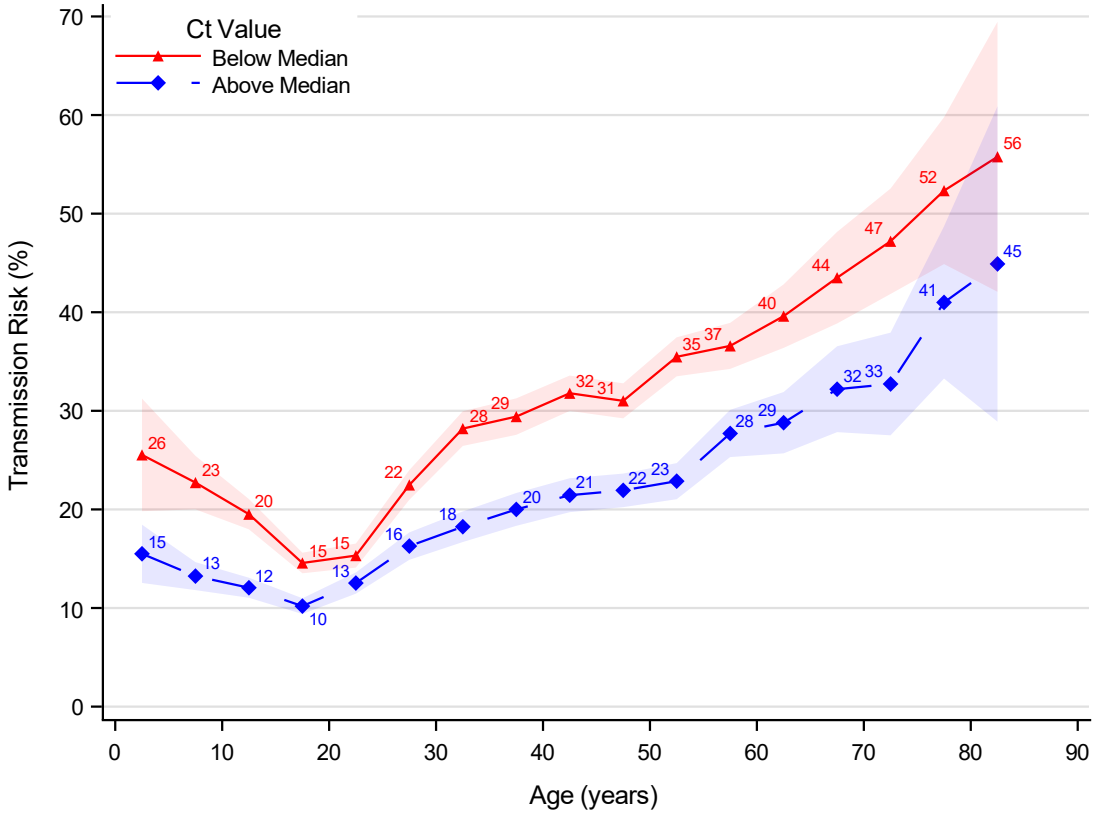


Notes: This figure illustrates the proportion (%) of positive secondary cases from primary cases with a Ct value ≥X. E.g., primary cases with a Ct value ≥30 account for 39% of total secondary cases; primary cases with a Ct value ≥32 account for 27% of total secondary cases. An RT-PCR test is positive if the Ct value is ≤38.

Figure 3 shows the age structured transmission risk stratified by the median Ct value. There is an overall positive association between age and transmission risk for adults (≥20 years) and a negative association for children (<20 years), i.e., the transmission risk increased with younger age for children. Across all age groups, we found that primary cases with a lower Ct value (<28, red) had a significantly higher transmission risk compared to primary cases with a higher Ct value (≥28, blue).

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 3: Age Structured Transmission Risk stratified by Median Ct Value

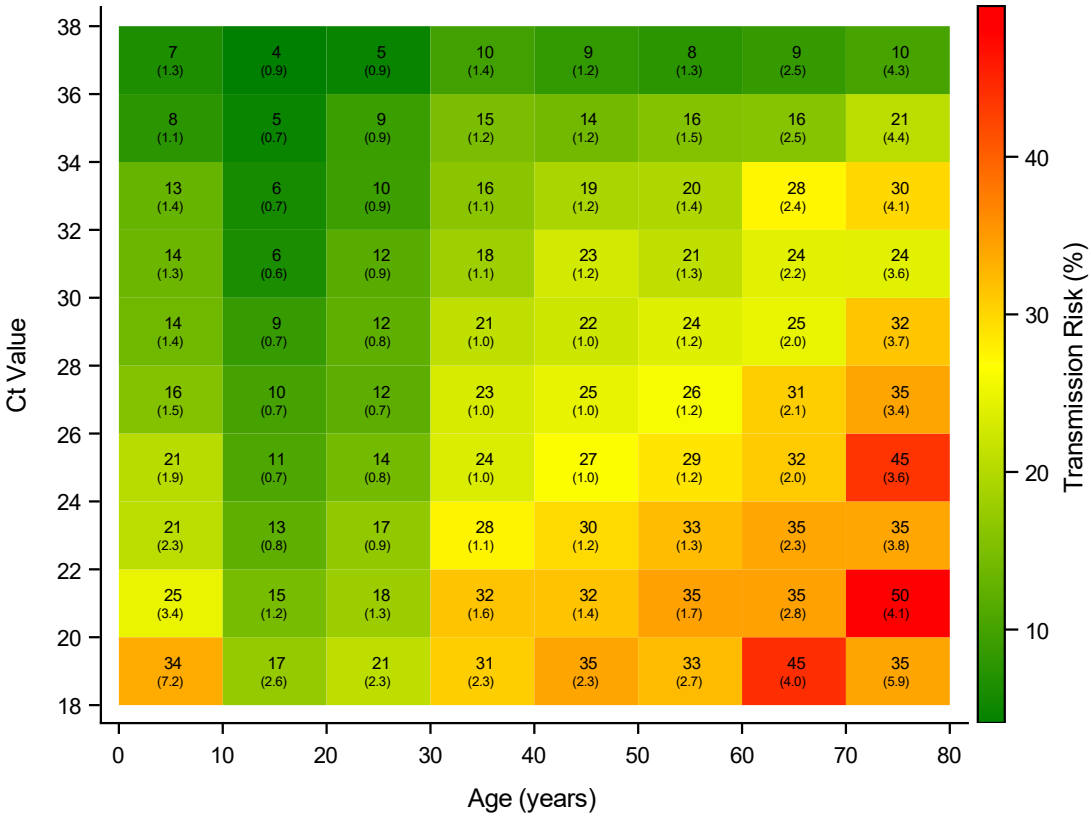


Notes: A primary case aged 0-5 years with a Ct value <28 has a transmission risk of 26%, while a primary case aged 0-5 years with a Ct value ≥28 has a transmission risk of 15%. The estimates come from estimating regression equation 2. The shaded areas show the 95% confidence bands clustered on the household level. The median Ct value for primary cases is 28.

Figure 4 shows the association between age, Ct value and transmission risk. The transmission risk generally increases with higher age and lower Ct value.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 4: Association between Age, Ct Value, and Transmission Risk



Notes: A primary case aged 30-40 years with a Ct value of 26-28 has a transmission risk of 23%, with a standard error of 1.0. The estimates control linearly for age of potential secondary cases. The estimates come from estimating regression equation 3. Standard errors clustered on the household level in parentheses.

Table 1 provides univariable and multivariable regression estimates of the odds ratio for the transmission risk across different observational characteristics. The transmission risk increases with a lower Ct value, i.e., with a higher viral load. For example, a primary case with a Ct value of 18-20 has a transmission risk 1.89 higher than a primary case with a Ct value of 36-38. The transmission risk increases with age for adults, such that a primary case aged 75-80 has a 1.98 times greater transmission risk than a primary case aged 30-35 year.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1: Odds Ratio for Transmission Risk

	Univariable		Multivariable	
	Odds Ratio	**95% CI**	**Odds Ratio**	**95% CI**
Ct Value				
18-20	1.00	(.)	1.00	(.)
20-22	0.92	(0.86-0.99)	0.96	(0.90-1.03)
22-24	0.84	(0.79-0.90)	0.89	(0.83-0.95)
24-26	0.77	(0.72-0.82)	0.82	(0.77-0.87)
26-28	0.73	(0.68-0.78)	0.78	(0.74-0.83)
28-30	0.69	(0.65-0.74)	0.74	(0.69-0.79)
30-32	0.65	(0.61-0.69)	0.71	(0.66-0.75)
32-34	0.62	(0.58-0.66)	0.68	(0.64-0.72)
34-36	0.57	(0.53-0.61)	0.62	(0.58-0.66)
36-38	0.50	(0.47-0.54)	0.53	(0.50-0.57)
Age				
0 - 5	0.87	(0.81-0.94)	1.11	(1.03-1.19)
5 - 10	0.75	(0.71-0.79)	0.95	(0.90-1.00)
10 - 15	0.67	(0.65-0.70)	0.82	(0.78-0.85)
15 - 20	0.64	(0.62-0.66)	0.70	(0.67-0.72)
20 - 25	0.90	(0.87-0.94)	0.73	(0.70-0.76)
25 - 30	1.09	(1.04-1.13)	0.87	(0.83-0.90)
30 - 35	1.00	(.)	1.00	(.)
35 - 40	0.96	(0.92-1.00)	1.07	(1.02-1.12)
40 - 45	1.01	(0.97-1.06)	1.13	(1.08-1.18)
45 - 50	1.09	(1.04-1.13)	1.11	(1.06-1.16)
50 - 55	1.36	(1.30-1.42)	1.15	(1.11-1.21)
55 - 60	1.74	(1.65-1.82)	1.20	(1.14-1.25)
60 - 65	2.10	(1.98-2.23)	1.27	(1.20-1.34)
65 - 70	2.69	(2.48-2.92)	1.49	(1.39-1.61)
70 - 75	3.08	(2.81-3.39)	1.63	(1.50-1.78)
75 - 80	3.88	(3.40-4.42)	1.98	(1.75-2.23)
80 - 85	4.67	(3.68-5.91)	2.45	(1.95-3.09)
85 - 90	3.53	(1.88-6.64)	1.97	(1.19-3.27)
Sex				
Male	1.00	(.)	1.00	(.)
Female	1.04	(1.02-1.06)	1.02	(1.01-1.04)
Household Size				
2	1.00	(.)	1.00	(.)
3	0.45	(0.45-0.46)	0.51	(0.50-0.52)
4	0.35	(0.34-0.35)	0.38	(0.38-0.39)
5	0.29	(0.28-0.29)	0.32	(0.31-0.33)
6	0.25	(0.24-0.26)	0.28	(0.27-0.29)
Number of observations		213,576		
Number of households		66,311		

Notes: 95% CI = 95% confidence intervals clustered on the household level. The estimates come from estimating regression equation 4.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

4.3 Sensitivity analyses

We found an approximately linear negative association between Ct value and transmission risk that was also present when controlling for age of the primary case (Figure S3). We found similar transmission risks for men and women (Figure S4). The age structured transmission risk stratified by Ct value quartiles showed the same picture as when stratifying by the median Ct value (Figure S5).

We also found that primary cases tested at hospital test facilities had a slightly higher transmission risk—across five-year age groups (Figure S6).

When excluding the age of the potential secondary case as a control variable, we found a slightly reduced association between the transmission risk and the interaction between the Ct value age of the primary case (Figure S7). Furthermore, when stratifying by Ct value quartiles and estimating the transmission risk by the interaction of the age of the primary case and the age of potential secondary cases, we still found the same overall pattern (Figure S8).

5 Discussion and Conclusion

To our knowledge, this is the first nationwide study investigating the association between transmission risk, age and Ct values. We here exploited the detailed Danish register data comprising the full population and all RT-PCR tests for SARS-CoV-2.

We found an approximately linear association between Ct value and transmission risk, implying that cases with a higher viral load are more infectious than cases with a lower viral load (Figure 1). This association was expected, because a high viral load implies that there are more virus particles in the sample, and hence the possibility of a larger inoculum.

However, this result also highlights that there is no obvious cut-off of Ct values to eliminate transmission risk. Importantly, we found a considerable transmission risk for cases with high Ct values: Primary cases with a Ct value of 38 had a transmission risk of 8% within the household. This result contradicts previous studies that have argued that

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

cases with a Ct value above a certain cut-off are not contagious. A Ct value cut-off of 30, or even lower, has been suggested (La Scola et al., 2020; Bullard et al., 2020; Brown et al., 2020; Prince-Guerra et al., 2021). These Ct values correspond to the limit of detection of virus cultures in Vero cells and antigen tests. However, choosing a cut-off of Ct$\leq$30 for infectiousness instead of Ct$\leq$38 would have missed transmission to 39% of the secondary cases in this study of household transmission (Figure 2, Table S3).

Rapid antigen tests are proposed for fast detection of cases as they provide a test result on the spot. Due to the lower sensitivity, they can only identify cases with a relatively high viral load, which implies that they can find the most infectious cases and allow for fast contact tracing. However, as they do not detect cases with a Ct value $>$30 (Jaafar et al., 2020; Bullard et al., 2020), they may miss a considerable share of infectious cases that would be detected with an RT-PCR test—42% of the primary cases in this study. This example assumes that only one test is performed—at least for a period of time. Nevertheless, rapid antigen tests could potentially increase detection of cases with repeated testing, conditional on the cases developing a viral load exceeding the detection threshold.

During an infection, the viral load is low shortly after exposure, increases over the infection, peaks around the onset of symptoms, and decreases later on (He et al., 2020). Hence, the Ct value is sensitive to the timing of the test, e.g., when a case is pre-symptomatic compared to later being symptomatic. Furthermore, there is a large variation in both severity of symptoms and infectivity across persons. Some studies found an indication of lower viral loads in asymptomatic persons (Zhou et al., 2020), while others found no differences across symptomatic and asymptomatic cases (Long et al., 2020). Additionally, the test results depend on the quality of the sampling as well as the assay, i.e., the chosen primers, probe and other reagents, which determines the accuracy of the test, making comparisons across laboratories difficult. In this study, the same methodology, set of primers, and probe were used throughout the whole study period, making the Ct values comparable. Despite the variations mentioned above, we found an association between Ct values and transmission risk, emphasizing the importance of this association.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Two other studies have also investigated the association between SARS-CoV-2 transmission risk and Ct values (Marks et al., 2021; Lee et al., 2021). Both also found a negative association. Lee et al. (2021) estimated the odds ratio of transmission to be 0.93 (95%-CI: 0.92-0.93), which is close to our estimate of 0.97 (95%-CI: 0.97-0.97) (Table S5). Our estimates here benefit from a large sample size and objective selection of potential secondary cases. We included all household members as potential secondary cases (unconditional on them being contacted by the official contact tracing system), of which 88% were tested within 1-14 days of the primary case. 48% of the potential secondary cases tested positive, indicating a high degree of transmission in the household domain. This is in contrast to Lee et al. (2021), who found that 40% of contacts were tested and 6% tested positive within 2-7 days.

Another main result from this study is that age was strongly associated with transmission risk—even when controlling for viral load (Figure 3 and S5). This result was not driven by the age of the potential secondary cases, as we found roughly the same pattern when adjusting for this (Figure S7). We found an overall positive association between age and transmission risk for adults ($\geq$20 years), whereas the transmission risk decreased with age for children ($<$20 years) (Figure 3). This pattern was found independently of the Ct value of the primary case, but the overall the transmission risk increased with lower Ct values (Figure S8).

It is noteworthy, that we found that age dominated viral load in predicting transmission risk. For instance, the transmission risk doubles when the Ct value decreases from 36-38 to 18-20; similarly, the risk doubled when the age of the primary case increases from 20-25 years to 65-70 years, and triples, when the case is 80-85 years (Table 1). This pattern could be driven by the susceptibility of the potential secondary cases, as people tend to live with their partner, who is around their own age, and parents live with their children (Lyngse et al., 2020; Madewell et al., 2020). We investigated this and did find that primary cases tend to infect other persons around the same age within the household (Figure S8). However, when we control for age of the potential secondary cases, we still found an association between age and transmission risk (Figure 4).

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Possible explanations for this finding could be that age is associated with increased viral exhalation (Edwards et al., 2021) or that the immunological response is associated with age (Long et al., 2020). The findings in children could be associated with younger children having closer contact with their parents than adolescents. Further clinical research is needed to clarify this.

Heald-Sargent et al. (2020) studied the viral load in 46 cases aged 0-5 years, 51 cases aged 5-17 years, and 48 cases aged 18-65 years. They found indications of high viral loads in children and speculated that they could be a main driver of the epidemic. Our results contradict this, as we did not find any association between the distribution of Ct values and age (Table S4). To calculate the distribution of Ct values across age groups, it is necessary to include a relatively large sample comprising all age groups.

As Ct values were only available for the primary cases identified by being tested in TCDK, only these primary cases were included in the analyses. We addressed the potential bias from not including primary cases that had been identified at hospitals (27%) by estimating the transmission risk stratified by primary cases identified at TCDK and hospitals (Figure S6). Primary cases identified in hospitals generally had a higher transmission risk, possibly because they were symptomatic. However, the trend in the age structured transmission risk was approximately the same across TCDK and hospitals, indicating that sample selection did not affect the general results.

We defined the primary cases as the first positive test within a household and all other persons living in the same household as potential secondary cases. We defined all secondary cases as those testing positive 1-14 days after the primary case. However, some of these co-primary and secondary cases may be misclassified, e.g., if they were infected earlier but not diagnosed, because they were pre- or asymptomatic. Including secondary cases found >14 days after the primary case could result in misclassification of secondary cases being either tertiary cases or having somewhere else as the source of secondary infections.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Optimal contact tracing naturally has to prioritize the order of new cases and their contacts. Our results suggest that contact tracing should prioritize cases according to Ct values, but more so, according to age.

In conclusion, we found that lower Ct values (indicating higher viral load) is associated with higher risk of SARS-CoV-2 transmission. However, even at high Ct values, transmission occurs. In addition, we found a strong association between age and transmission risk that dominated the Ct value association.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

Brown, C. S., Clare, K., Chand, M., Andrews, J., Auckland, C., Beshir, S., Choudhry, S., Davies, K., Freeman, J., Gallini, A., et al. (2020). Snapshot PCR surveillance for SARS-CoV-2 in hospital staff in England. *Journal of Infection*, 81(3), 427–434.

Bullard, J., Dust, K., Funk, D., Strong, J. E., Alexander, D., Garnett, L., Boodman, C., Bello, A., Hedley, A., Schiffman, Z., et al. (2020). Predicting Infectious Severe Acute Respiratory Syndrome Coronavirus 2 From Diagnostic Samples. *Clinical Infectious Diseases*, 71(10), 2663–2666.

Cheng, H.-Y., Jian, S.-W., Liu, D.-P., Ng, T.-C., Huang, W.-T., Lin, H.-H., et al. (2020). Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA Internal Medicine*, 180(9), 1156–1163.

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L., et al. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*, 25(3), 2000045.

Edwards, D. A., Ausiello, D., Salzman, J., Devlin, T., Langer, R., Beddingfield, B. J., Fears, A. C., Doyle-Meyers, L. A., Redmann, R. K., Killeen, S. Z., et al. (2021). Exhaled aerosol increases with COVID-19 infection, age, and obesity. *Proceedings of the National Academy of Sciences*, 118(8).

He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5), 672–675.

Heald-Sargent, T., Muller, W. J., Zheng, X., Rippe, J., Patel, A. B., & Kociolek, L. K. (2020). Age-related differences in nasopharyngeal severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) levels in patients with mild to moderate coronavirus disease 2019 (COVID-19). *JAMA Pediatrics*, 174(9), 902–903.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Jaafar, R., Aherfi, S., Wurtz, N., Grimaldier, C., Hoang, V. T., Colson, P., Raoult, D., & La Scola, B. (2020). Correlation between 3790 qPCR positives samples and positive cell cultures including 1941 SARS-CoV-2 isolates. *Clinical Infectious Diseases*, (pp. ciaa1491).

La Scola, B., Le Bideau, M., Andreani, J., Grimaldier, C., Colson, P., Gautret, P., Raoult, D., et al. (2020). Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. *European Journal of Clinical Microbiology & Infectious Diseases*, 39(6), 1059–1061.

Lee, L. Y., Rozmanowski, S., Pang, M., Charlett, A., Anderson, C., Hughes, G. J., Barnard, M., Peto, L., Vipond, R., Sienkiewicz, A., Hopkins, S., Bell, J., Crook, D. W., Gent, N., Walker, A. S., Eyre, D. W., & Peto, T. E. (2021). An observational study of SARS-CoV-2 infectivity by viral load and demographic factors and the utility lateral flow devices to prevent transmission. *Modernising Medical Microbiology*.

Long, Q.-X., Tang, X.-J., Shi, Q.-L., Li, Q., Deng, H.-J., Yuan, J., Hu, J.-L., Xu, W., Zhang, Y., Lv, F.-J., et al. (2020). Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Medicine*, 26(8), 1200–1204.

Lyngse, F. P., Kirkeby, C. T., Halasa, T., Andreasen, V., Skov, R. L., Møller, F. T., Krause, T. G., & Mølbak, K. (2020). COVID-19 Transmission Within Danish Households: A Nationwide Study from Lockdown to Reopening. *medRxiv*.

Madewell, Z. J., Yang, Y., Longini, I. M., Halloran, M. E., & Dean, N. E. (2020). Household Transmission of SARS-CoV-2: A Systematic Review and Meta-analysis. *JAMA Network Open*, 3(12), e2031756–e2031756.

Marks, M., Millat-Martinez, P., Ouchi, D., h Roberts, C., Alemany, A., Corbacho-Monne, M., Ubals, M., Tobias, A., Tebé, C., Ballana, E., et al. (2021). Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. *The Lancet Infectious Diseases*.

Prince-Guerra, J. L., Almendares, O., Nolen, L. D., Gunn, J. K., Dale, A. P., Buono, S. A., Deutsch-Feldman, M., Suppiah, S., Hao, L., Zeng, Y., et al. (2021). Evaluation of

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abbott BinaxNOW Rapid Antigen Test for SARS-CoV-2 Infection at Two Community-Based Testing Sites—Pima County, Arizona, November 3–17, 2020. *Morbidity and Mortality Weekly Report*, 70(3), 100–105.

Singanayagam, A., Patel, M., Charlett, A., Bernal, J. L., Saliba, V., Ellis, J., Ladhani, S., Zambon, M., & Gopal, R. (2020). Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Eurosurveillance*, 25(32), 2001483.

SSI (2021). Statens Serum Institut. COVID-19: PCR-test. `https://covid19.ssi.dk/diagnostik/pcr-test`. Accessed: 2021-02-12.

Vogels, C. B., Brito, A. F., Wyllie, A. L., Fauver, J. R., Ott, I. M., Kalinich, C. C., Petrone, M. E., Casanovas-Massana, A., Muenker, M. C., Moore, A. J., et al. (2020). Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. *Nature Microbiology*, 5(10), 1299–1305.

Zhou, R., Li, F., Chen, F., Liu, H., Zheng, J., Lei, C., & Wu, X. (2020). Viral dynamics in asymptomatic patients with COVID-19. *International Journal of Infectious Diseases*, 96, 288–290.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

6 Appendix A: Summary statistics

Table S1: Summary Statistics, Testing

	TCDK			Hospitals			Total		
	N Obs.	% Tot.	% Ind.	N Obs.	% Tot.	% Ind.	N Obs.	% Tot.	% Ind.
All tests									
Tests	9,416,298	74		3,223,544	26		12,639,842	100	
Positive tests	146,708	69		67,168	31		213,876	100	
Individuals	3,329,115	84		1,601,804	40		3,968,337	100	
Primary Cases									
Individuals	66,602	73		24,842	27		91,444	100	
With Ct value	66,311		100	0		0	66,311		73
Potential Secondary Cases									
Individuals	213,576	73		77,421	27		290,997	100	
Tested	188,807		88	68,738		89	257,545		89
Tested positive	103,389		48	39,274		51	142,663		49

Notes: N Obs. = number of observations. % Tot. = percentage of total. % Ind. = percentage of individuals.

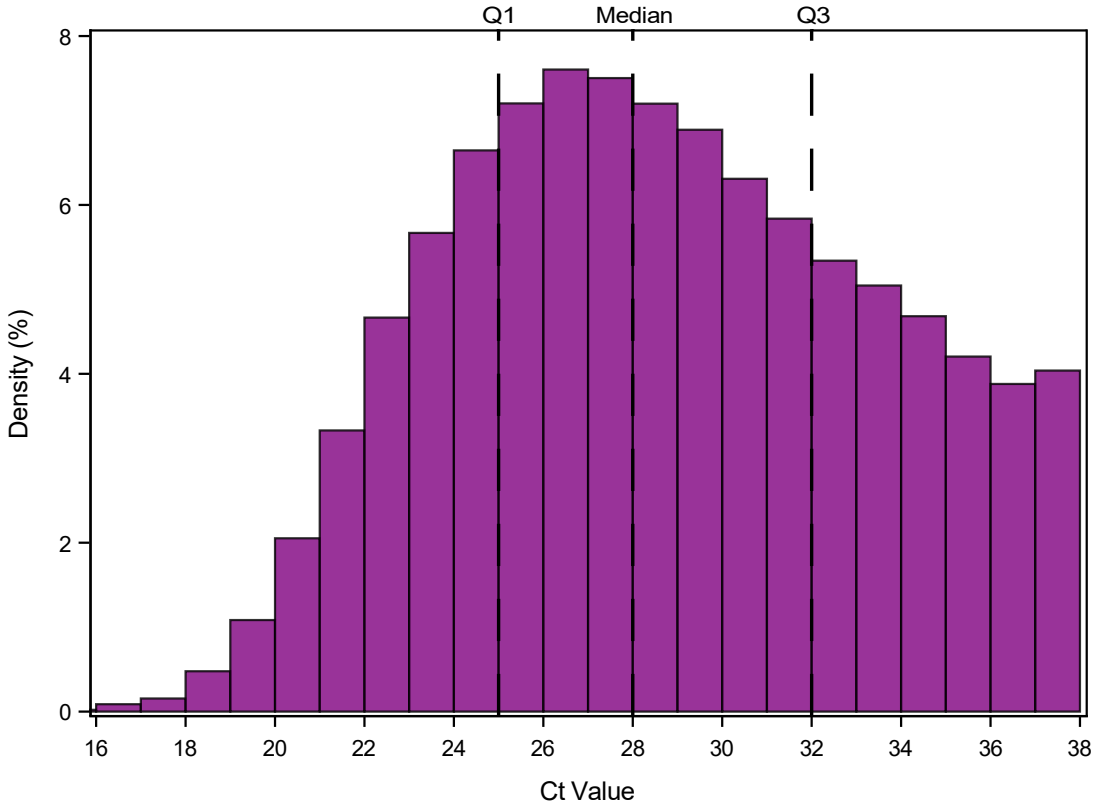
medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table S2: Summary Statistics, Cases

	Primary Cases	Potential Secondary Cases	Positive Secondary Cases
Total	66,311	213,576	103,389
Sex			
Male	33,545	107,994	51,766
Female	32,766	105,550	51,619
Age			
0 - 5	1,005	14,306	3,313
5 - 10	2,775	15,263	5,960
10 - 15	5,163	20,596	8,908
15 - 20	9,251	24,197	12,440
20 - 25	7,745	19,639	10,091
25 - 30	6,135	15,332	8,446
30 - 35	4,805	12,531	7,079
35 - 40	4,179	12,562	6,582
40 - 45	4,642	16,582	7,617
45 - 50	4,941	19,030	8,252
50 - 55	5,004	16,710	8,043
55 - 60	4,323	11,495	6,515
60 - 65	2,709	6,572	4,144
65 - 70	1,514	3,699	2,469
70 - 75	1,201	2,830	1,982
75 - 80	663	1,512	1,076
80 - 85	218	539	393
85 - 90	38	123	70
90 - 95	-	21	<5
>95	-	5	<5
Household Members			
2	24,242	48,489	32,999
3	15,364	46,100	22,674
4	16,860	67,451	28,614
5	7,543	37,720	14,291
6	2,302	13,816	4,811

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure S1: Distribution of Ct Values for Primary Cases



Notes: Only Ct values of the first positive test of primary cases are shown. Q1 = 1st quartile (25th percentile, P25), etc. An RT-PCR test is positive if the Ct value is ≤38.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

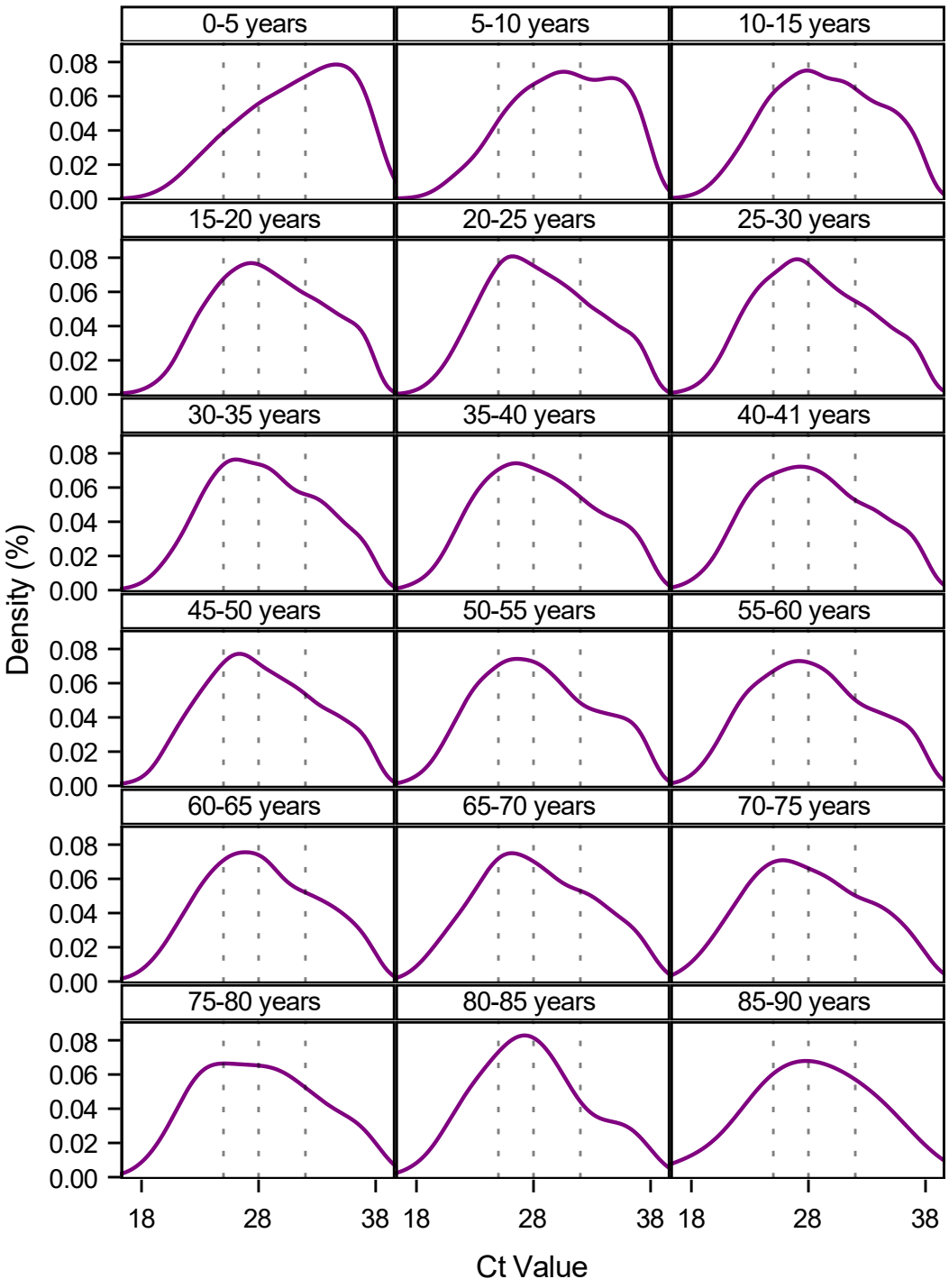
Table S3: Summary Statistics: Cases and Ct Values

Ct value	Primary Cases	Potential Secondary Cases	Positive Secondary Cases
<18	159	507	239
18	214	660	422
19	515	1,572	928
20	1,060	3,347	1,878
21	1,809	5,723	3,204
22	2,771	8,832	4,804
23	3,538	11,196	6,004
24	4,114	13,142	6,843
25	4,612	14,826	7,510
26	4,970	15,889	7,982
27	5,082	16,171	8,051
28	4,868	15,717	7,643
29	4,662	14,810	7,105
30	4,336	14,016	6,694
31	4,026	13,156	6,085
32	3,651	11,753	5,444
33	3,395	11,141	5,037
34	3,221	10,641	4,787
35	2,900	9,549	4,153
36	2,654	8,717	3,693
37	2,565	8,427	3,407
38	1,189	3,784	1,476
Total	66,311	213,576	103,389

Notes: An RT-PCR test is positive if the Ct value is $\leq$38.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure S2: Distribution of Ct Values for primary cases stratified by age



Notes: Only Ct values of the first positive test of primary cases are shown. The vertical dotted reference lines are the 1st quartile, the median, and 3rd quartile of the total population. An RT-PCR test is positive if the Ct value is $\leq$38.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table S4: Distribution of Ct-values for Primary Cases

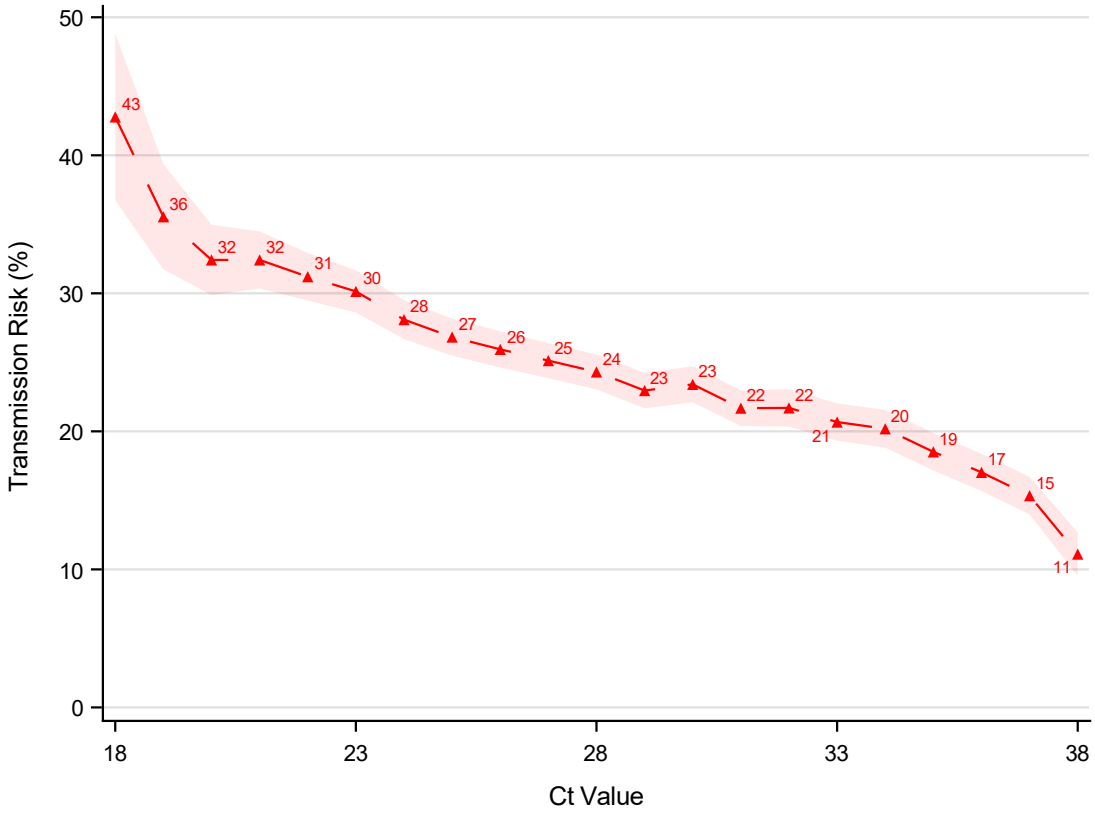
	P5	**P10**	**P25**	**P50**	**P75**	**P90**	**P95**	**Mean**	**Observations**
Total	21	23	25	28	32	35	37	29	66,311
Sex									
Male	21	22	25	28	32	35	37	28	33,545
Female	22	23	25	29	33	36	37	29	32,766
Age									
0 - 5	23	25	28	32	35	37	37	31	1,005
5 - 10	23	25	27	31	34	36	37	31	2,775
10 - 15	22	23	26	29	33	36	37	29	5,163
15 - 20	22	23	25	29	33	35	37	29	9,251
20 - 25	22	23	25	28	32	35	37	29	7,745
25 - 30	21	22	25	28	32	35	37	28	6,135
30 - 35	21	23	25	28	32	35	36	29	4,805
35 - 40	21	22	25	28	32	35	37	28	4,179
40 - 45	21	22	25	28	32	35	37	28	4,642
45 - 50	21	22	25	28	32	35	37	28	4,941
50 - 55	21	22	25	28	32	35	37	28	5,004
55 - 60	21	22	24	28	32	35	37	28	4,323
60 - 65	21	22	25	28	32	35	36	28	2,709
65 - 70	21	22	24	28	32	35	37	28	1,514
70 - 75	20	22	24	28	32	35	36	28	1,201
75 - 80	21	22	24	28	32	35	37	28	663
80 - 85	21	22	24	28	31	35	36	28	218
85 - 90	19	22	25	29	32	35	36	28	38

Notes: P5 = 5th percentile, P25 = 25th percentile (1st quartile), etc.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

7 Appendix B: Sensitivity Analyses

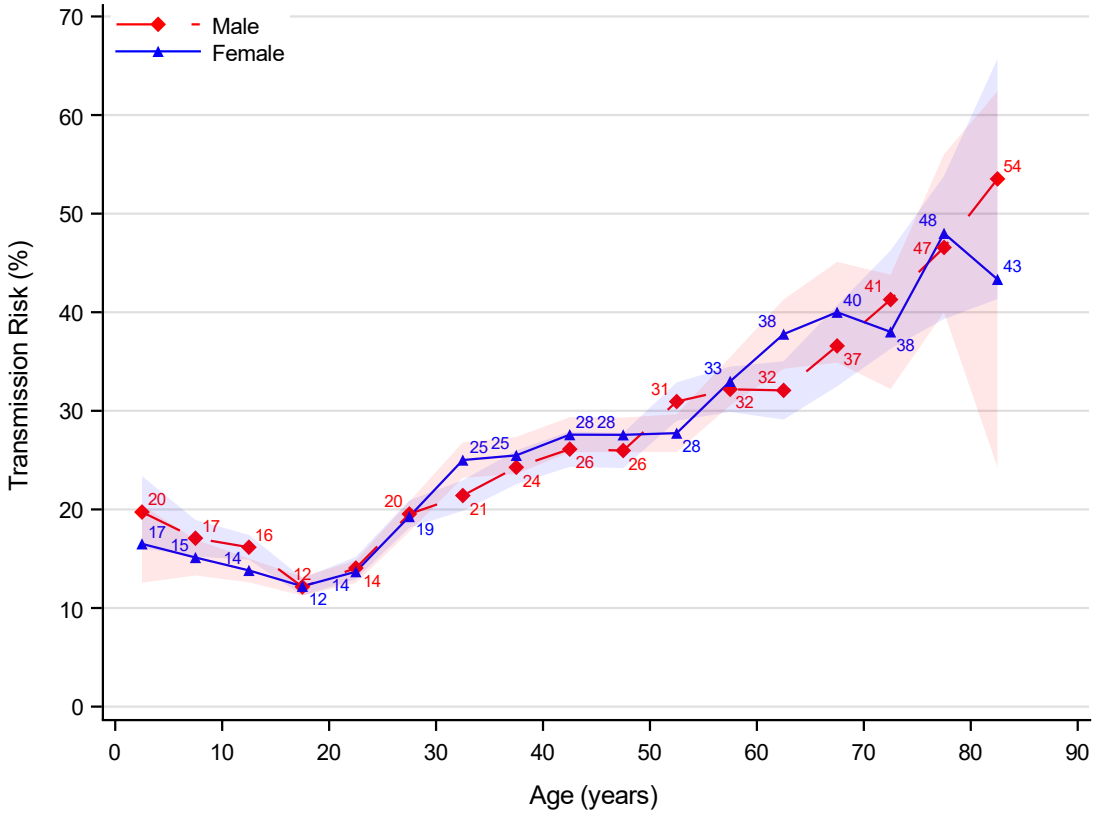
Figure S3: Association between Ct value (quartiles) and Age Structured Transmission Risk



Notes: This figure shows the transmission risk by Ct value controlling for age of the primary case. The shaded areas show the 95% confidence bands clustered on the household level.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

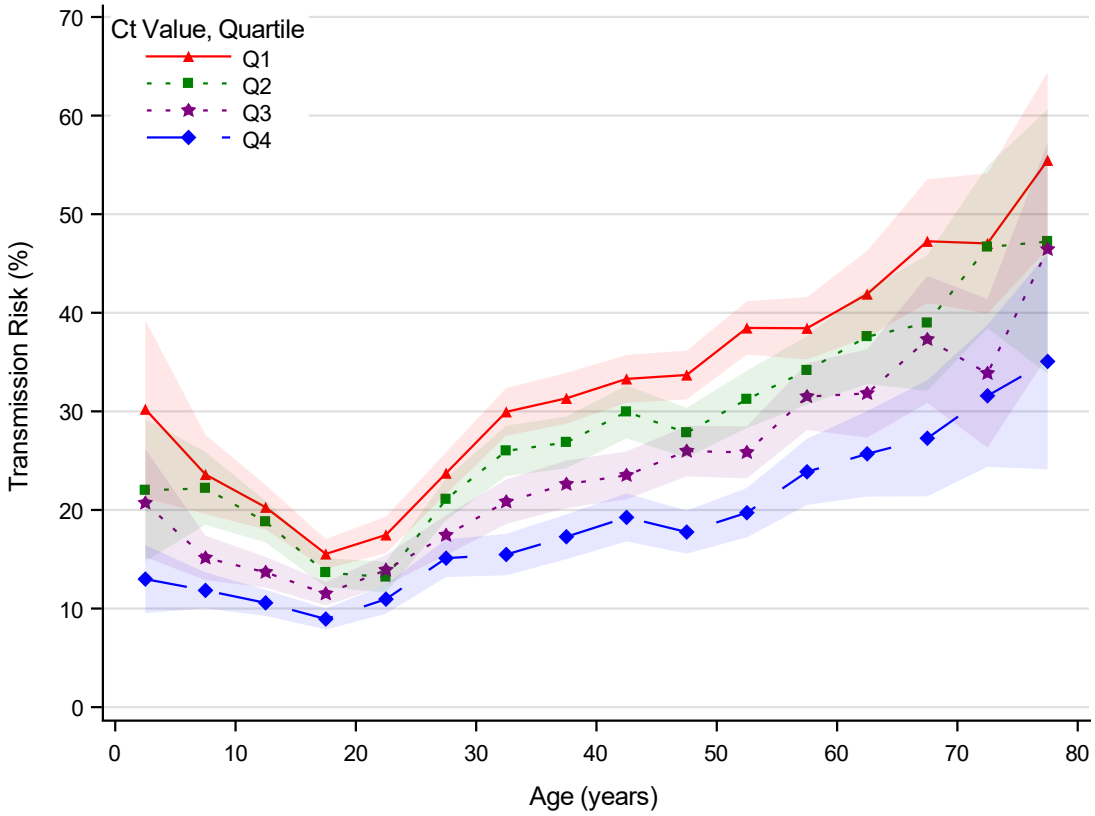
Figure S4: Association between Age Structured Transmission Risk and Sex



Notes: This figure shows the age structured transmission risk estimates stratified by age. It is estimated with regression equation 2. The shaded areas show the 95% confidence bands clustered on the household level.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

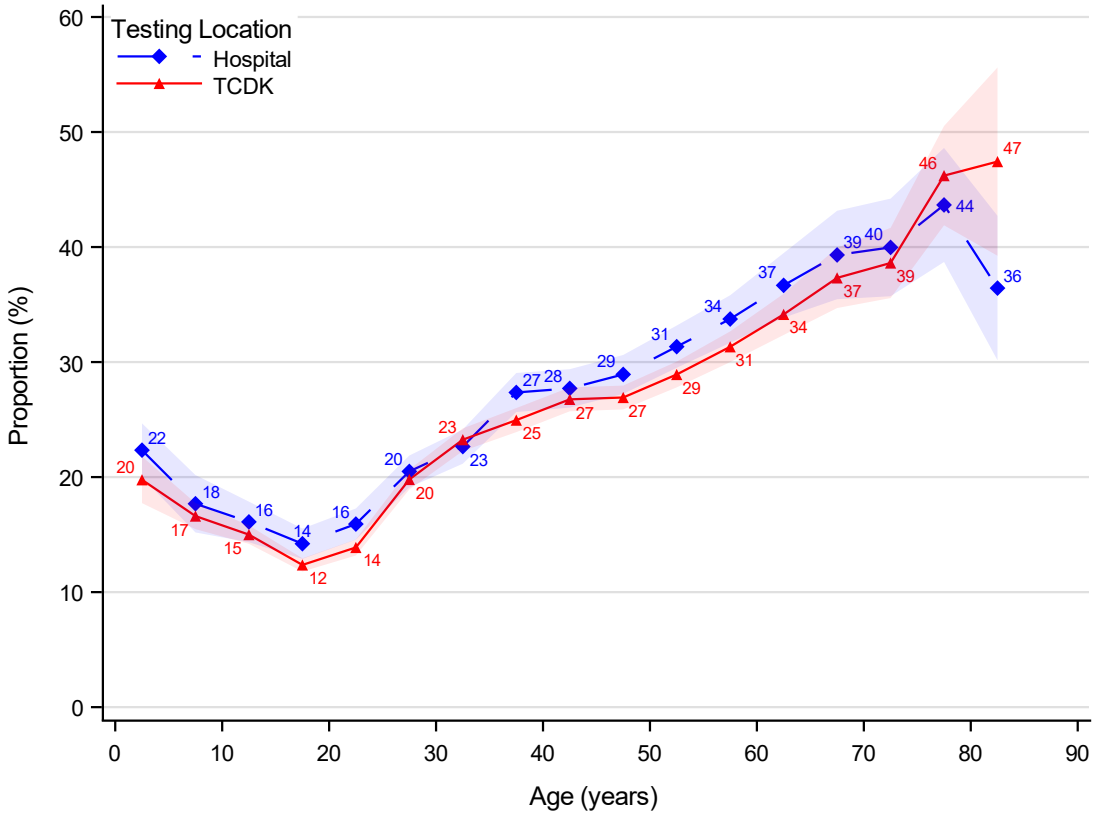
Figure S5: Association between Ct value (quartiles) and Age Structured Transmission Risk



Notes: This figure shows the age structured transmission risk estimates stratified by Ct value quartiles. It is estimated with regression equation 2. The shaded areas show the 95% confidence bands clustered on the household level.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

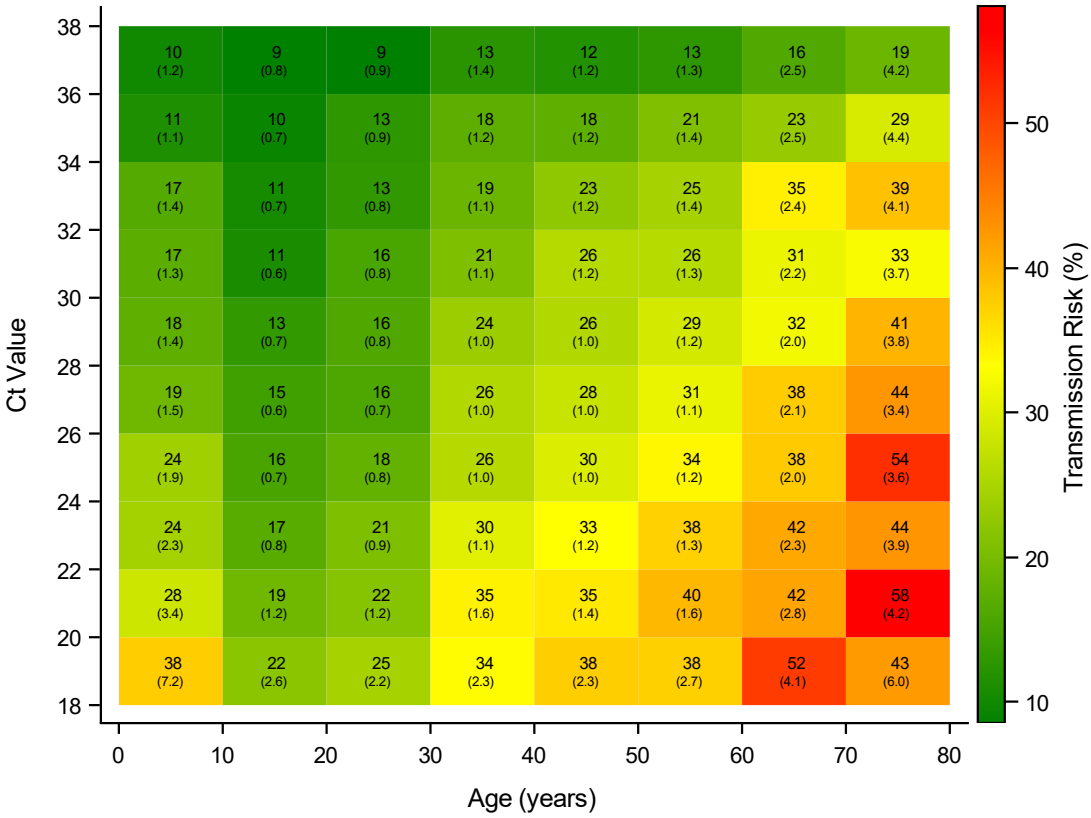
Figure S6: Association between Age Structured Transmission Risk stratified by Place of Testing



Notes: This figure shows the age structured transmission risk estimates stratified by TCDK and hospitals. It is estimated with regression equation 2. The shaded areas show the 95% confidence bands clustered on the household level.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure S7: Association between Age, Ct Value, and Transmission Risk

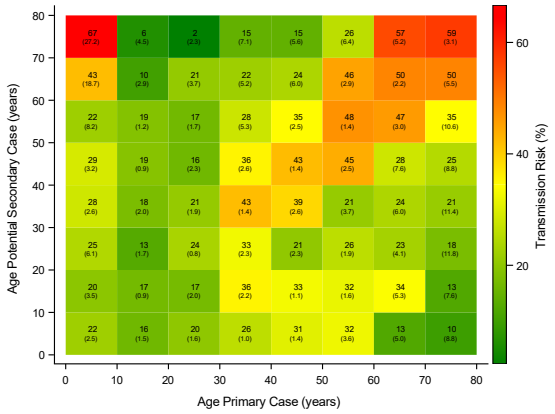


Notes: This figure shows the transmission risk from the age and Ct value interaction. It is estimated with regression equation 3. The estimates does not include controls for age of potential secondary cases. Standard errors clustered on the household level in parentheses.

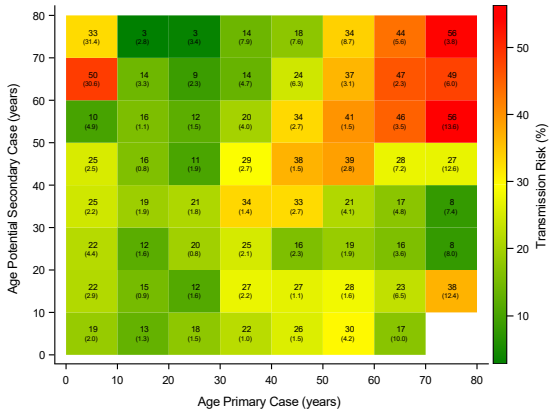
medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure S8: Association between age of primary case, potential secondary case, and transmission risk—stratified by Ct value quartile

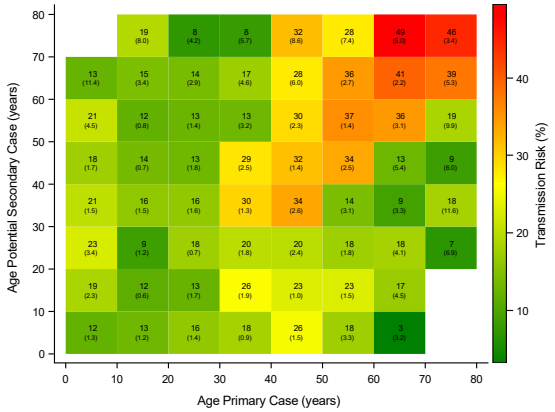
(a) Ct value, Q1



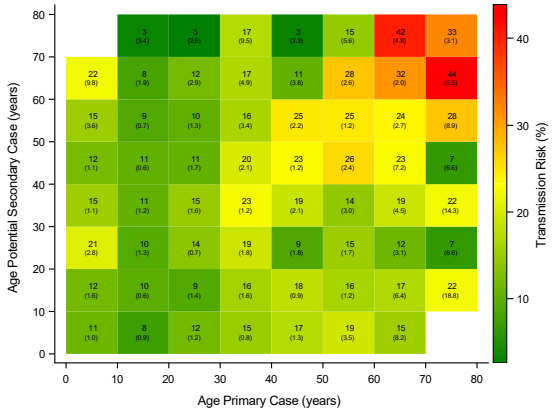
(b) Ct value, Q2



(c) Ct value, Q3



(d) Ct value, Q4



Notes: This figure shows the transmission risk from the interaction between age of the primary case and age of the potential secondary cases, stratified by Ct value quartiles. Panel (a) shows the transmission risk for primary cases with Ct values in the lowest quartile. Standard errors clustered on the household level in parentheses.

medRxiv preprint doi: https://doi.org/10.1101/2021.02.28.21252608; this version posted March 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table S5: Odds Ratio Estimates for Transmission Risk

	Univariable		Multivariable	
	Odds Ratio	**95% CI**	**Odds Ratio**	**95% CI**
Ct Value	0.97	(0.97-0.97)	0.97	(0.97-0.97)
Age				
0 - 5	0.87	(0.81-0.94)	1.11	(1.03-1.19)
5 - 10	0.75	(0.71-0.79)	0.95	(0.90-1.00)
10 - 15	0.67	(0.65-0.70)	0.82	(0.78-0.85)
15 - 20	0.64	(0.62-0.66)	0.7	(0.67-0.72)
20 - 25	0.90	(0.87-0.94)	0.73	(0.70-0.75)
25 - 30	1.09	(1.04-1.13)	0.87	(0.83-0.90)
30 - 35	1.00	(.)	1.00	(.)
35 - 40	0.96	(0.92-1.00)	1.07	(1.02-1.11)
40 - 45	1.01	(0.97-1.06)	1.13	(1.08-1.18)
45 - 50	1.09	(1.04-1.13)	1.11	(1.06-1.15)
50 - 55	1.36	(1.30-1.42)	1.15	(1.10-1.20)
55 - 60	1.74	(1.65-1.82)	1.19	(1.14-1.25)
60 - 65	2.10	(1.98-2.23)	1.27	(1.20-1.34)
65 - 70	2.69	(2.48-2.92)	1.49	(1.39-1.61)
70 - 75	3.08	(2.81-3.39)	1.64	(1.50-1.78)
75 - 80	3.88	(3.40-4.42)	1.98	(1.75-2.23)
80 - 85	4.67	(3.68-5.91)	2.45	(1.95-3.10)
85 - 90	3.53	(1.88-6.64)	1.98	(1.19-3.29)
Sex				
Male	1.00	(.)	1.00	(.)
Female	1.04	(1.02-1.06)	1.02	(1.01-1.04)
Household Size				
2	1.00	(.)	1.00	(.)
3	0.45	(0.45-0.46)	0.51	(0.50-0.52)
4	0.35	(0.34-0.35)	0.38	(0.38-0.39)
5	0.29	(0.28-0.29)	0.32	(0.31-0.33)
6	0.25	(0.24-0.26)	0.28	(0.27-0.29)
Number of observations		213,576		
Number of households		66,311		

Notes: 95% CI = 95% confidence intervals clustered on the household level.