

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Model-based estimates of age-structured SARS-CoV-2 epidemiology in households

Damon J.A. Toth^{1,2,3*}, Theresa R. Sheets³, Alexander B. Beams^{3,#a}, Sharia M. Ahmed¹, Nathan Seegert⁴, Jay Love¹,
Lindsay T. Keegan^{1,2}, Matthew H. Samore^{1,2}

¹ Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, United States of America

² Department of Veterans Affairs Salt Lake City Healthcare System, Salt Lake City, Utah, United States of America

³ Department of Mathematics, University of Utah, Salt Lake City, Utah, United States of America

⁴ Department of Finance, University of Utah David Eccles School of Business, Salt Lake City, Utah, United States of America

^{#a} Current address: Department of Mathematics, Simon Fraser University, Burnaby, British Columbia, Canada

* Corresponding author

E-mail: Damon.Toth@hsc.utah.edu (DT)

18 **Abstract**

19 Understanding how infectious disease transmission varies from person to person, including associations with age
20 and contact behavior, can help design effective control strategies. Within households, transmission may be highly
21 variable because of differing transmission risks by age, household size, and individual contagiousness. Our aim was
22 to disentangle those factors by fitting mathematical models to SARS-CoV-2 household survey and serologic data.
23 We surveyed members of 3,381 Utah households from January-April 2021 and performed SARS-CoV-2 antibody
24 testing on all available members. We paired these data with a probabilistic model of household importation and
25 transmission composed of a novel combination of transmission variability and age- and size-structured
26 heterogeneity. We calculated maximum likelihood estimates of mean and variability of household transmission
27 probability between household members in different age groups and different household sizes, simultaneously with
28 importation probability and probabilities of false negative and false positive test results. 12.8% of the individual
29 participants showed serologic evidence of prior infection or reported a prior positive test on the survey, and 17.4%
30 of the participating households showed evidence of at least one SARS-CoV-2 importation. Serologically positive
31 individuals in younger age groups were less likely than older adults to have tested positive during their infection
32 according to our survey results. Our model results suggested that adolescents and young adults (ages 13-24)
33 acquired SARS-CoV-2 infection outside the household at a rate substantially higher than younger children and older
34 adults. Our estimate of the household secondary attack rate (HSAR) among adults aged 45 and older exceeded
35 HSARs to and/or from younger age groups. We found lower HSAR in households with more members, independent
36 of age differences. Our findings from age-structured transmission analysis suggest that age groups contact each
37 other at different rates within households, a key insight for understanding community outbreak patterns and
38 mechanisms of differential infection risk.

39 **Author Summary**

40 Infectious diseases can spread through human communities in irregular patterns, partly because different
41 demographic groups, such as age groups, experience different transmission risks due to contact or other behavioral
42 or physiological differences. Understanding the factors driving age differences in transmission can help predict
43 patterns of disease spread and suggest efficient public health strategies to mitigate outbreaks. Households are inter-
44 age mixing locations where age differences in transmission can be studied. In early 2021, we collected blood
45 samples from all members of thousands of households in Utah and tested them for SARS-CoV-2 antibodies, from
46 which prior COVID-19 infection can be inferred. We paired these data with mathematical models that quantify
47 probabilities that different combinations of household members end up infected for different assumptions about non-
48 household infection and within-household transmission. Our estimates suggest that adolescents and young adults
49 acquired infection outside the household more frequently than did other age groups. After a household importation
50 occurred, middle-aged and older adults living together transmitted to each other more readily than all other age
51 pairings for a given household size. The age patterns of household transmission we found suggest that within-
52 household contact rate differences play a significant role in driving household transmission epidemiology.

53 **1 Introduction**

54 The spread of an infectious disease through a community can occur in complicated patterns, especially when the
55 person-to-person transmission potential of infected individuals is highly variable [1]. Such variability can occur
56 physiologically, through varying duration, severity, and contagiousness of human infections, as well as socially,
57 through varying social settings and contact behaviors that provide transmission opportunities [2]. These variabilities
58 may be partially explained by correlating transmission risk with observable demographic variables such as age and
59 number of contacts in primary social settings such as households [3]. Understanding the mechanisms and
60 associations along which transmission variabilities arise helps improve the efficiency of infection control via
61 targeted pharmaceutical and social interventions [4].

62 Transmission mechanisms and variabilities could be studied by directly observing the transmissions and
63 transmission opportunities for each infected person, but such data are difficult and costly to collect. Cross-sectional
64 prevalence data, e.g. from testing serological samples for the presence of antibodies among individuals who have
65 had contact with each other, can be used to indirectly estimate transmission rates and patterns. When such data are
66 paired with mechanistically derived mathematical models, insights can be derived about the source of transmission
67 variabilities using appropriate analytic and computational techniques [5, 6]. In this work, our goal was to use such a
68 model-based approach to study transmission variabilities of SARS-CoV-2, with a focus on age-based associations
69 effects, using household seroprevalence data.

70 SARS-CoV-2 poses high risk of transmission from person to person in indoor public spaces and households [7],
71 where spread of SARS-CoV-2 has been well documented [8]. Studying transmission at the household level is
72 convenient relative to other indoor locations with more transient populations, and household transmission research
73 can provide illuminating insights into risk factors and mechanisms of transmission. Furthermore, households provide
74 an intergenerational mixing ground with contact across age groups, and both household and age structure can be
75 crucially important for accurate characterization of community-wide transmission [3].

76 Studying differences in susceptibility, transmissibility, and contact by age has been an important area of research
77 with implications for public health policy for many infectious diseases. During the COVID-19 pandemic, younger
78 individuals have been less likely to experience severe disease but are a potential source of SARS-CoV-2 contagion
79 [9]. Infected children with mild or no symptoms may nevertheless transmit to other more vulnerable contacts,

80 including to highly vulnerable aged adults in multi-generational households [10]. Some SARS-CoV-2 and influenza
81 studies have found that households with more members have lower transmission rates [11-13], and it is unclear the
82 extent to which that observation might be attributed to different age configurations of larger households or other
83 factors. Understanding differences in transmission among and between different age groups, and the mechanisms by
84 which those differences might arise, can shed light on the benefits of age-group targeting of transmission-
85 interruption measures, including vaccination [14].

86 Serological surveys, in which blood samples from individuals in a population are tested for the presence of
87 antibodies, can be used to infer which individuals had experienced prior infection, including those who had not been
88 diagnosed [15]. When performed over entire households, serological surveys offer one way to assess household
89 transmission: the number of household members testing positive provides information to infer the distribution of the
90 total number of household members infected during household outbreaks that ran their course prior to data
91 collection. Quantitative model-based techniques [16] can be used to extract information from these datasets on the
92 transmissibility of the pathogen under conditions of sustained close contact. When such datasets include metadata on
93 household member demographics, they can help answer questions about the relative susceptibility to or
94 transmissibility of infection across demographics such as age [17].

95 In this work, we developed a novel mathematical model of household transmission that incorporates transmission
96 variability [18] and heterogeneity over age and household size. A central goal achieved by our model-based findings
97 was to estimate quantitative age differences in household COVID-19 epidemiology using detailed serological data
98 from Utah households and elucidate the extent to which these estimates reveal potential mechanisms of within-
99 household transmission differences by age.

100 **2 Methods**

101 *2.1 Data collection from Utah households*

102 Our data was collected via the Utah Health & Economic Recovery Outreach (HERO) Project [19], which recruited
103 participants from households in the state of Utah during several distinct phases in 2020 [20] and 2021 [21].

104 Members of participating households completed surveys and were subsequently invited to a mobile testing site to
105 contribute samples for testing. Participants contributing the specific subset of the Utah HERO Project data analyzed
106 and modeled for this manuscript were recruited during two different phases of the Utah HERO Project in 2021:

107 “school district testing” and “hotspot testing” (see Supporting Information for recruitment details). In contacted
108 households, all household members aged 12 years or older were asked to complete an individual survey, adults were
109 asked to complete the individual survey on behalf of children <12 years of age, and the household was asked to
110 complete an overall household survey. Each household member >6 months of age was asked to provide a blood
111 sample at a mobile testing site.

112 The University of Utah Institutional Review Board reviewed the surveillance project that produced the data analyzed
113 in this manuscript and determined it as non-research public health surveillance, waived the requirement for
114 documented consent, and determined that use of these data for analysis to understand the dynamics of SARS-CoV-2
115 transmission was exempt from further review (IRB_00132598). Individuals were informed of the project procedures
116 and that participation was voluntary. Participants provided their agreement to participate and were given the chance
117 to opt out of having their data used for future research. The data were analyzed anonymously for this manuscript.

118 The following data from the surveys were retained for our analysis. From the household survey, we used the
119 household size (number of people living in the household). From the individual surveys, we used the date of birth,
120 whether the individual had previously tested positive for COVID-19, the date of a previous positive test, the number
121 of prior COVID-19 vaccine doses received, and the date of any prior vaccine doses. We calculated the age of
122 surveyed participants on the day of serological testing, using the date of birth from the survey. Prior test result data
123 from the survey were classified as prior positive reported (P) or no prior positive reported (N).

124 We excluded households from our analysis if the number of people living in the house (household size) or the age of
125 any household member were missing. We also excluded households for which the reported size was inconsistent
126 with the individual-level data (i.e., the reported household size was less than the number of individuals contributing
127 data for that household).

128 Whole blood was collected from participants at mobile testing facilities, via finger prick for participants aged
129 between 6 months and 12 years and via venous puncture for participants aged 12 years or older. Serum specimens
130 from blood collected via venous puncture were analyzed using all three of the following tests: Abbott SARS-CoV-2
131 IgG assay performed on an Abbott Architect i2000 instrument (Abbott Laboratories), Euroimmun Anti-SARS-CoV-
132 2 ELISA (IgG) assay, and Siemens SARS-CoV-2 IgG (Cov2G) assay. Only the Euroimmun test was used to analyze
133 specimens from finger prick samples.

134 Methodology and criteria for a positive antibody result were defined according to the manufacturers' instructions
135 [22-24] and as in prior studies [25, 26]. Antibody test results for each test were classified according to negative (N),
136 positive (P), or not performed / indeterminate (X). Because the Euroimmun and Siemens assays detect anti-SARS-
137 CoV-2 IgG for the S1 domain of viral spike protein (anti-S1), which can be produced by vaccination, participants
138 who reported having received at least one dose of COVID-19 vaccine prior to data collection and tested positive on
139 either assay were automatically labeled as indeterminate (X) for purposes of informing prior infection. The Abbott
140 assay detects anti-SARS-CoV-2 IgG for the nucleocapsid protein (anti-N), which we assumed was not impacted by
141 prior vaccination and that vaccinated patients testing Abbott-positive had tested positive for prior infection [27]. We
142 defined an antibody test result as being "informative" for potential prior infection status when the result was
143 negative or positive for the Abbott test, negative or positive for Euroimmun or Siemens for participants reporting no
144 prior vaccination, or negative for Euroimmun or Siemens for participants reporting any prior vaccination.

145 Each member of a household retained for analysis was labeled according to their combination of five data points:
146 reported prior positive test (N or P); Abbott test result (N, P, or X); Euroimmun test result (N, P, or X); Siemens test
147 result (N, P, or X). For example, a participant labeled "NPXN" reported no prior positive test, tested positive by
148 Abbott, had an indeterminate Euroimmun test, and tested negative by Siemens. Further labels were added indicating
149 age group and household size, as described in later sections. Then we created a table tallying the number of
150 participants with each label in each household: a column for each unique label that at least one participant received
151 and a row for each household. Each table element is the number of participants in each household with each label.
152 We then tallied the number of unique rows, C , and labeled the sets of numbers in each unique row as vectors
153 $\mathbf{y}_1, \dots, \mathbf{y}_C$. Finally, the entire dataset is represented by the vector $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_C, f_1, \dots, f_C)$, where frequency f_i is the
154 number of households described by the corresponding \mathbf{y}_i .

155 The following sections describe the mathematical model designed to produce a likelihood function $\mathcal{L}(\mathbf{y})$ to quantify
156 the probability of observing the dataset \mathbf{y} under a given set of epidemiological assumptions. The de-identified data
157 and codes, written in R version 4.0.3, to produce \mathbf{y} , calculate and optimize $\mathcal{L}(\mathbf{y})$, and produce the results in this
158 manuscript, are publicly available at <https://github.com/damontoth/householdTransmission2021>.

159 *2.2 Household transmission model*

160 First, we define the probability that a single infected person transmits infection to different numbers of people living
 161 in the same household. We separate a household's population into age groups that can differ in both susceptibility to
 162 acquisition from infected housemates and in transmissibility to susceptible housemates from each group. For
 163 simplicity, we limited the number of transmission age groups to three, named U, V, and W. We use the notation $F_{\mathbf{y};\mathbf{z}}$
 164 to denote the probability that a single infected person with $\mathbf{z} = (z_u, z_v, z_w)$ susceptible housemates in each group
 165 directly transmits infection to $\mathbf{y} = (y_u, y_v, y_w)$ of them, respectively.

166 To derive our formula for F , we assume that the probability of transmission to a u-group person is a random variable
 167 taking the form $1 - e^{-x}$, where x is a gamma-distributed random variable with shape k and rate r . The gamma
 168 distribution models the mean and variability of infected individuals' transmissibility to housemates, where the k
 169 parameter is equivalent to the dispersion parameter k often used to model transmission variability and
 170 superspreading. The value of x can be considered the product of the total dose of exposure experienced by an
 171 infected person's susceptible housemate and the per-dose probability of infection, as in the exponential dose-
 172 response model. Then, the probabilities of transmission to a person in group V and W is $1 - e^{-\sigma_v x}$ and $1 - e^{-\sigma_w x}$,
 173 where the coefficients σ_v and σ_w alter the probability of infection relative to a person in group U, from a given
 174 infected housemate.

175 From the above assumptions, the following formula for F applies (see Supporting Information for derivation):

176
$$F_{\mathbf{y};\mathbf{z}}(r, k, \sigma_v, \sigma_w) =$$

177
$$\binom{z_u}{y_u} \binom{z_v}{y_v} \binom{z_w}{y_w} \sum_{j_u=0}^{y_u} \sum_{j_v=0}^{y_v} \sum_{j_w=0}^{y_w} (-1)^{j_u+j_v+j_w} \binom{y_u}{j_u} \binom{y_v}{j_v} \binom{y_w}{j_w} \left(1 + \frac{z_u - y_u + j_u + \sigma_v(z_v - y_v + j_v) + \sigma_w(z_w - y_w + j_w)}{r}\right)^{-k}$$

178 Because we also model differential transmissibility from each group, we use three sets of $(r, k, \sigma_v, \sigma_w)$ parameters
 179 governing transmission: $(r_u, k_u, \sigma_{uv}, \sigma_{uw})$, $(r_v, k_v, \sigma_{vv}, \sigma_{vw})$, and $(r_w, k_w, \sigma_{vw}, \sigma_{ww})$, which are the parameters used
 180 when a person from group U, V, and W is the potential transmitter, respectively. Using those parameters, we can
 181 also define the directional household secondary attack rates (HSARs) p_{ij} , the mean transmission probability from an
 182 infected person from group i to a susceptible person from group j in the same household:

183
$$p_{uu} = 1 - (1 + 1/r_u)^{-k_u}, \quad p_{uv} = 1 - (1 + \sigma_{uv}/r_u)^{-k_u}, \quad p_{uw} = 1 - (1 + \sigma_{uw}/r_u)^{-k_u}$$

184
$$p_{vu} = 1 - (1 + 1/r_v)^{-k_v}, \quad p_{vv} = 1 - (1 + \sigma_{vv}/r_v)^{-k_v}, \quad p_{vw} = 1 - (1 + \sigma_{vw}/r_v)^{-k_v}$$

185
$$p_{wu} = 1 - (1 + 1/r_w)^{-k_w}, \quad p_{wv} = 1 - (1 + \sigma_{wv}/r_w)^{-k_w}, \quad p_{ww} = 1 - (1 + \sigma_{ww}/r_w)^{-k_w}$$

186 For clarity, we report our estimates for the set of transmission parameters using the above nine directional attack
 187 rates, plus the dispersion parameters k_i . The values of r_i and σ_{ij} can then be calculated by inverting the above
 188 equations.

189 With the above function F defined for transmissions from one infectious individual in a household, we now derive
 190 formulae for multiple potential transmitters. First, let $H_{\mathbf{x};\mathbf{y};\mathbf{z}}$ be the probability that, in a household with $\mathbf{x} =$
 191 (x_u, x_v, x_w) infected members from each group, those infected persons transmit infection directly to a total of $\mathbf{y} =$
 192 (y_u, y_v, y_w) of $\mathbf{z} = (z_u, z_v, z_w)$ susceptible housemates from each group. First, we define the trivial case where $\mathbf{x} =$
 193 $(0,0,0)$: no infected household members to transmit:

194
$$H_{000;\mathbf{y};\mathbf{z}} = \begin{cases} 1, & y_u + y_v + y_w = 0 \\ 0, & y_u + y_v + y_w > 0 \end{cases}$$

195 Next, we have already defined the cases where $x_u + x_v + x_w = 1$ (exactly one infected household member):

196
$$H_{100;\mathbf{y};\mathbf{z}} = F_{\mathbf{y};\mathbf{z}}(r_u, k_u, \sigma_{uv}, \sigma_{uw})$$

197
$$H_{010;\mathbf{y};\mathbf{z}} = F_{\mathbf{y};\mathbf{z}}(r_v, k_v, \sigma_{vv}, \sigma_{vw})$$

198
$$H_{001;\mathbf{y};\mathbf{z}} = F_{\mathbf{y};\mathbf{z}}(r_w, k_w, \sigma_{wv}, \sigma_{ww})$$

199 Then we iteratively calculate $H_{\mathbf{x};\mathbf{y};\mathbf{z}}$ for each (x_u, x_v, x_w) in increasing sequence of $x_u + x_v + x_w = 2, 3, \dots$:

200
$$H_{\mathbf{x};\mathbf{y};\mathbf{z}} =$$

201
$$\sum_{i_u=0}^{y_u} \sum_{i_v=0}^{y_v} \sum_{i_w=0}^{y_w} \frac{x_u H_{\mathbf{x}-(1,0,0);i;\mathbf{z}} H_{100;\mathbf{y}-i;\mathbf{z}-i} + x_v H_{\mathbf{x}-(0,1,0);i;\mathbf{z}} H_{010;\mathbf{y}-i;\mathbf{z}-i} + x_w H_{\mathbf{x}-(0,0,1);i;\mathbf{z}} H_{001;\mathbf{y}-i;\mathbf{z}-i}}{x_u + x_v + x_w}$$

202 With H in hand, we then calculate $T_{\mathbf{x};\mathbf{y};\mathbf{z}}$: the probability that \mathbf{x} already infected (u-group, v-group, w-group)
 203 household members lead to a *total* of \mathbf{y} transmissions to \mathbf{z} susceptible household members from each respective
 204 group. In other words, $T_{\mathbf{x};\mathbf{y};\mathbf{z}}$ is the probability that the vector of final household outbreak sizes of each group is $\mathbf{x} + \mathbf{y}$

205 given that \mathbf{x} were initially infected and \mathbf{z} were initially susceptible. The probability accounts for transmissions
 206 directly from initially infected members and in any subsequent transmission generations.

207 We calculate values of T iteratively, starting with trivial case of no initial infections in the household:

$$208 \quad T_{000;y;z} = \begin{cases} 1, & y_u + y_v + y_w = 0 \\ 0, & y_u + y_v + y_w > 0 \end{cases}$$

209 Then, starting with $y_u + y_v + y_w = 0$ and then for increasing $y_u + y_v + y_w$ in sequence, we iteratively calculate:

$$210 \quad T_{x;y;z} = \sum_{i_u=0}^{y_u} \sum_{i_v=0}^{y_v} \sum_{i_w=0}^{y_w} H_{x;i;z} T_{i;y-i;z-i}$$

211 *2.3 Total household infection size model*

212 The probability of infections within a household also depends on the likelihood of members acquiring infection from
 213 non-household members (called “community” acquisitions). We calculated those probabilities for each household
 214 member based on characteristics that could extend beyond membership in the U, V, and W transmission groups
 215 described in the last section.

216 Let $\mathbf{n} = (n_u, n_v, n_w)$ be the number of household members in each of the three transmission groups. Let l_1, \dots, l_n be
 217 the infection indicators for each household member, where the l_i values equal 1 if infected and 0 if uninfected. Let
 218 u_1, \dots, u_n be indicators for whether each household member belongs to transmission group U (1 if yes, 0 if no), and
 219 similarly v_1, \dots, v_n and w_1, \dots, w_n are indicators for membership in groups V and W. Then we define $\mathbf{l} =$
 220 $(l_u, l_v, l_w) = (\sum_{j=1}^n l_j u_j, \sum_{j=1}^n l_j v_j, \sum_{j=1}^n l_j w_j)$.

221 Let p_1, \dots, p_n be the community acquisition probability of each household member. Now we define the likelihood
 222 M_{l_1, \dots, l_n} for household members indicated by the l values ending up infected by either community acquisition or
 223 household transmission.

$$224 \quad M_{l_1, \dots, l_n} = \sum_{i_1=0}^{l_1} \dots \sum_{i_n=0}^{l_n} (p_1)^{i_1} (1 - p_1)^{1-i_1} \dots (p_n)^{i_n} (1 - p_n)^{1-i_n} T_{\mathbf{i}; \mathbf{l}-\mathbf{i}, \mathbf{n}-\mathbf{i}}$$

225 The values $\mathbf{i} = (i_u, i_v, i_w) = (\sum_{j=1}^n i_j u_j, \sum_{j=1}^n i_j v_j, \sum_{j=1}^n i_j w_j)$. The M values depend on both the household
 226 transmission parameters defined in the prior section and the community acquisition probabilities.

227 **2.4 Data accuracy model**

228 Applying the M formula from the prior section directly to our data would be problematic because the true infection
229 history of each household member is not known with certainty. Our data include up to 4 sources of COVID-19 test
230 information by which prior infection status of individual household members can be probabilistically inferred:
231 results from up to 3 different antibody tests (Abbot, Siemens, and Euroimmun) and results from surveys in which
232 participants could report results of a prior test.

233 Antibody test results are subject to imperfect sensitivity and specificity due to false negatives and false positives. To
234 account for these, we added 6 additional parameters to our likelihood model: ϕ_{AA} , ϕ_{AS} , and ϕ_{AE} are the antibody test
235 sensitivities for Abbot, Siemens, and Euroimmun, respectively, and π_{AA} , π_{AS} , and π_{AE} are the specificities. Prior test
236 results for SARS-CoV-2 reported on the survey also do not perfectly identify those with prior infections. Among
237 people who had a prior infection, a portion of them never tested positive because they did not receive a test while
238 their infection was identifiable or because they falsely tested negative. Because the degree of this under-
239 ascertainment has been known to vary widely by age, we modeled it using a separate probability for each household
240 member: $(\phi_{S1}, \dots, \phi_{Sn})$ are the probabilities that each of the n surveyed persons in a household reported receiving a
241 positive test for the virus had they been infected. The parameter π_S is the probability that any surveyed person with
242 no prior infection did not report receiving a positive test.

243 For each person $i = 1, \dots, n$ in a household of size n , we considered the overall combination of that person's test
244 results to calculate ϕ_i and π_i , the probability of that test result combination occurring if the person had been infected
245 and never infected, respectively. The following provide illustrated examples.

246 Example 1: result combination “NNNN”: no prior positive reported on survey, and negative by all 3 antibody tests:

$$247 \quad \phi_i = (1 - \phi_{Si})(1 - \phi_{AA})(1 - \phi_{AE})(1 - \phi_{AS}), \quad \pi_i = \pi_S \pi_{AA} \pi_{AE} \pi_{AS}$$

248 Example 2: result combination “NNPN”: no prior positive reported on survey, negative by Abbot, positive by
249 Euroimmun, and negative by Siemens:

$$250 \quad \phi_i = (1 - \phi_{Si})(1 - \phi_{AA})\phi_{AE}(1 - \phi_{AS}), \quad \pi_i = \pi_S \pi_{AA} (1 - \pi_{AE})\pi_{AS}$$

251 Example 3: result combination “PNXN”: prior positive reported on survey, negative by Abbot, no or indeterminate
 252 Euroimmun result, and negative by Siemens:

$$253 \quad \phi_i = \phi_{Si}(1 - \phi_{AA})(1 - \phi_{AS}), \quad \pi_i = (1 - \pi_S)\pi_{AA} \pi_{AS}$$

254 With the ϕ_i and π_i values for each household member in hand, the following formula specifies the overall
 255 likelihood of observing the entire household’s set of test results together:

$$256 \quad \mathcal{L} = \sum_{l_1=0}^1 \dots \sum_{l_n=0}^1 (\phi_1)^{l_1} (\pi_1)^{1-l_1} \dots (\phi_n)^{l_n} (\pi_n)^{1-l_n} M_{l_1, \dots, l_n}$$

257 *2.5 Age- and size-based parameter values*

258 The formulas in the previous sections included three different components that we assumed could depend on the
 259 characteristics of household member i . First, the “transmission group” specified by (u_i, v_i, w_i) ; second, the
 260 community acquisition probability p_i ; and third, the prior infection ascertainment probability ϕ_{Si} . We modeled each
 261 of these dependencies based on the individual’s age at the time of data collection. Specifically, we categorized ages
 262 using combinations of age ranges for terms defined by Medical Subject Headings (MeSH) [28] (Table 1). Because
 263 we had limited data from the youngest individual age groups, we combined the newborn, infant, and preschool child
 264 groups into one category (aged 0 to 5 years). Other MeSH-defined age categories and sub-categories were retained
 265 separately, for a total of 8 categories.

266 **Table 1. Age categories and age-specific model parameters**

MeSH age category names	Age range (years)	Community acquisition parameter	Prior infection ascertainment parameter
Newborn/Infant/Preschool Child	0 – 5	p_a	ϕ_{Sa}
Child	6 – 12	p_b	ϕ_{Sb}
Adolescent	13 – 18	p_c	ϕ_{Sc}
Young adult	19 – 24	p_d	ϕ_{Sd}
Adult (excluding young adult)	25 – 44	p_e	ϕ_{Se}
Middle aged	45 – 64	p_f	ϕ_{Sf}
Aged (excluding 80 and over)	65 – 79	p_g	ϕ_{Sg}
Aged, 80 and over	80+	p_h	ϕ_{Sh}

267 MeSH = Medical Subject Headings

268 The community acquisition parameter for each individual p_i in a household was set to one of the 8 values p_a through
 269 p_h according to age group, and similarly each individual’s prior infection ascertainment parameter ϕ_{Si} was set to a
 270 value among ϕ_{Sa} through ϕ_{Sh} . For the transmission groupings, because we limited our model to three groups, we
 271 tested different ways to combine our 8 MeSH age groups from Table 1 into three groups of adjacent MeSH age

272 groups. We designated transmission group “U” as the oldest age group, “V” as the middle age group, and “W” as the
 273 youngest.

274 We also allowed within-household transmission to vary by total household size (number of household members)
 275 using two additional parameters: σ_{34} and σ_{5+} scale the probability of acquiring infection from an infected housemate
 276 of a person living in a size 3-4 or size 5-or-more household, respectively, relative to a person living in a size-2
 277 household. These parameters multiply each of the σ_{xy} parameters that model age differences, so that each age-to-age
 278 transmission probability is assumed to scale by household size in the same way. Thus, we calculate three sets of T
 279 values quantifying household transmission chain probabilities, for a given set of epidemiological parameters:

$$280 \quad T = \begin{cases} T(r_u, k_u, \sigma_{uv}, \sigma_{uw}, r_v, k_v, \sigma_{vv}, \sigma_{vw}, r_w, k_w, \sigma_{ww}, \sigma_{ww}), & \text{size-2 house} \\ T(r_u/\sigma_{34}, k_u, \sigma_{34}\sigma_{uv}, \sigma_{34}\sigma_{uw}, r_v/\sigma_{34}, k_v, \sigma_{34}\sigma_{vv}, \sigma_{34}\sigma_{vw}, r_w/\sigma_{34}, k_w, \sigma_{34}\sigma_{ww}, \sigma_{34}\sigma_{ww}), & \text{size 3/4 house} \\ T(r_u/\sigma_{5+}, k_u, \sigma_{5+}\sigma_{uv}, \sigma_{5+}\sigma_{uw}, r_v/\sigma_{5+}, k_v, \sigma_{5+}\sigma_{vv}, \sigma_{5+}\sigma_{vw}, r_w/\sigma_{5+}, k_w, \sigma_{5+}\sigma_{ww}, \sigma_{5+}\sigma_{ww}), & \text{size 5+ house} \end{cases}$$

281 In our full model we sought to use our data to simultaneously estimate the 22 epidemiological parameters

$$282 \quad \theta_1 = (p_a, \dots, p_h, r_u, k_u, \sigma_{uv}, \sigma_{uw}, r_v, k_v, \sigma_{vv}, \sigma_{vw}, r_w, k_w, \sigma_{ww}, \sigma_{ww}, \sigma_{34}, \sigma_{5+})$$

283 and the 15 data accuracy parameters

$$284 \quad \theta_2 = (\phi_{Sa}, \dots, \phi_{Sh}, \phi_{AA}, \phi_{AE}, \phi_{AS}, \pi_S, \pi_{AA}, \pi_{AE}, \pi_{AS})$$

285 using maximum likelihood estimation (MLE).

286 The log likelihood of the dataset \mathbf{y} described in Section 2.1 with variable set $\theta = (\theta_1, \theta_2)$ is

$$287 \quad \ln \mathcal{L}(\mathbf{y}|\theta) = f_1 \ln \mathcal{L}(\mathbf{y}_1|\theta) + \dots + f_c \ln \mathcal{L}(\mathbf{y}_c|\theta)$$

288 **2.6 Likelihood optimization and uncertainty**

289 We maximized the log likelihood over the unknown parameters θ using the observations for each household, to
 290 produce the MLE $\hat{\theta}$. The log likelihood maximization was performed using an iterative scheme in R, in which R’s
 291 native “optim” function was applied alternately to the set of epidemiological parameters θ_1 and then to the set of
 292 data accuracy parameters θ_2 , holding the other set constant at the previous solution and continuing back and forth
 293 until the estimates converged.

294 We first tested models with simpler assumptions for the dispersion parameters k_u , k_v , and k_w . The value $k = 1$ is a
295 special case representing exponentially-distributed (special case of gamma-distributed) transmissibility by an
296 infected person. The special case $k \rightarrow \infty$ represents no variance in transmissibility. We tested models assuming
297 $k_u = k_v = k_w = 1$ and $k_u = k_v = k_w = \infty$ against one that optimized those values along with other parameters; if
298 one of the simpler models is favored, we could move forward with other model comparisons described next using
299 less computationally costly formulae.

300 We also tested different age-group choices for the household transmission groups U, V, and W described above, and
301 we selected the grouping that produced the MLE with the highest likelihood. We refer to this result as the MLE of
302 the “full model” to which alternate, simpler models were compared, to test the extent to which the full model MLE
303 might suffer from overfitting.

304 For the full model, we derived approximate confidence interval boundaries for an individual parameter θ_i using the
305 likelihood ratio test, using the statistic $2 \log(\mathcal{L}(\hat{\boldsymbol{\theta}})/\mathcal{L}(\boldsymbol{\theta}))$, where $\boldsymbol{\theta}$ consists of θ_i freely varying and the other
306 elements of $\boldsymbol{\theta}$ held at their optimal value. We defined a 95% confidence interval boundary where θ_i produces a
307 value for this statistic equal to the 95th percentile of the chi-squared distribution with 1 degree of freedom.

308 *2.7 Alternate model comparisons*

309 We compared the MLE of the full model against simpler models with fewer free parameters. To calculate P-values
310 at which a simpler model could be rejected in favor of one with more free parameters and of which the simpler
311 model is a special case, we used the chi-squared distribution with degrees of freedom equal to difference in the
312 number of free parameters between the two models. We quantified and compared the parsimony of different models
313 using the Akaike information criterion, which aims to balance goodness of fit with model simplicity.

314 For the set of average HSARs, which consists of 9 free age-group-to-age-group values in the full model, we tested
315 alternate models that use fewer parameters via simplifying assumptions. We tested an “equal contact” model, which
316 assumes that the extent to which household members contact each other is independent of age, but that susceptibility
317 and/or transmissibility per contact can differ by age. The equal contact model corresponds to the assumption that
318 $\sigma_{uv} = \sigma_{vv} = \sigma_{vw}$ and $\sigma_{uw} = \sigma_{vw} = \sigma_{ww}$. Thus, in this model the 9 HSARs are governed by 5 free parameters: 2
319 relative susceptibility (σ) parameters and 3 average transmissibility parameters.

320 Next, we tested a “reciprocal transmission” model, which assumes that the average household transmission
321 probability between an infected and susceptible person from two different age groups is the same regardless of
322 which person is infected, i.e., $p_{vu} = p_{uv}$, $p_{wu} = p_{uw}$, and $p_{vw} = p_{vw}$. With those 3 constraints, the reciprocal
323 transmission model is governed by 6 free parameters to determine the 9 HSARs.

324 To seek a maximally parsimonious model, we tested an additional series of models applying a variety of simplifying
325 assumptions, as defined by optimal AIC. These included assumptions of age-independence of transmission to and/or
326 from each of the individual transmission age groups. For example, assuming $p_{uu} = p_{uv} = p_{uw}$ enforces
327 transmission from age group U to be equally likely to a housemate of any age. Likewise, assuming $p_{uu} = p_{vu} = p_{wu}$
328 enforces acquisition of infection by age group U to be equally likely from an infected housemate of any age, and
329 assuming $p_{uu} = p_{uv} = p_{uw} = p_{vu} = p_{wu}$ enforces an equal HSAR between age group U and all ages in both
330 directions. We repeated the MLE procedure for each of these models and each age group, as well as combinations of
331 them for multiple age groups in cases where multiple sets of assumptions produced favorable fits by the AIC. We
332 also tested models that collapse the different age-group-based community acquisition parameters p_a through p_h and
333 prior infection ascertainment parameters ϕ_{sa} through ϕ_{sh} into fewer age groupings, as well as models that assume
334 no effect of household size on transmission rates. We combined any of the above simplifications that were favored
335 by the AIC criteria to arrive at our maximally parsimonious model.

336 Finally, we analyzed implications of our results for the most parsimonious model. We explored potential
337 mechanisms of household transmission difference by age that are consistent with estimated age differences in
338 HSAR, including age-specific differences in household contact rates and per-contact transmissibility or
339 susceptibility. We also calculated the overall age-specific infection risk for each group implied by the model results
340 and the age configurations of households in the data to provide a summary of the estimated rate of infection from
341 household and non-household sources for each age group.

342 **3. Results**

343 ***3.1 Data summary***

344 Of the 96,547 total households contacted across both the school district and hotspot testing phases, we compiled
345 serological test data and associated survey data from 3,381 (3.5%) of the households and 7,707 individuals (Table A
346 in Supporting Information).

347 Of the 3,381 total households, we retained 2,514 (74.4%) for our analysis. The 867 excluded households were
348 removed because of unknown household size (n=55), a reported household size inconsistent with individual data
349 (n=57), or incomplete age data obtained for the household members (n=755). In the 2,514 retained households, there
350 were 5,855 (2.33 per household) household members.

351 Among the 5,855 retained participants, 1,337 (22.8%) reported having received at least one dose of a COVID-19
352 vaccine prior to serological data collection, and 727 (12.4%) reported having received two vaccine doses. For those
353 reporting a plausible date for their first vaccine dose (n = 1274), the reported date occurred a median of 32.5 days
354 prior to serological sample collection (range 0–112 days). For those reporting plausible dates for two vaccine doses
355 (n = 654), the reported date of the second dose occurred a median of 29 days prior to data collection (range 0–91
356 days). Participants reporting prior vaccination tested antibody positive by Euroimmun and Siemens tests at highly
357 elevated rates compared to those not reporting prior vaccination (Table B in Supporting Information), consistent
358 with the presence of vaccine-induced antibodies; these positive results were labeled non-informative ('X') when
359 calculating prior SARS-CoV-2 infection likelihood in our model.

360 5,626 (96.1%) of the retained participants provided a blood sample that produced at least one informative antibody
361 test result, and 2,848 (48.6%) of the retained participants received informative results for all three antibody tests. Of
362 those, 2,412 (84.7%) were all three negative, 230 (8.1%) were all three positive, and 206 (7.2%) were a mixture of
363 positive and negative results. The most common conflicting result combinations among the latter group were
364 Abbott-negative conflicting with Euroimmun-positive and Siemens-positive results (n=99) and Euroimmun-positive
365 conflicting with Abbott-negative and Siemens-negative results (n=63). We tallied other result combinations,
366 including indeterminate (non-informative) results marked by 'X' as well as prior-positive test reporting for
367 participants in each antibody test result group (Table C in Supporting Information).

368 In all, 748 (12.8%) of the 5,855 participants were positive on at least one informative antibody test or reported prior
369 test. These 748 participants lived in 438 of the 2514 households, thus 17.4% of the participating households had at
370 least one informative positive result. Among those 438 households with at least 1 informative positive participant,

371 246 had exactly 1 and 192 had more than 1, ranging from 2 to 8. A crude HSAR estimate, i.e., assuming exactly 1
372 importation in each of the 438 households and calculating the fraction of housemates of an importer who were also
373 informative positive, was 33.9%. Our likelihood model relaxed those assumptions in several ways, including
374 accounting for the probability of multiple household importations, false negative and false positive tests, and other
375 features, as detailed in the following sections.

376 *3.2 Maximum likelihood estimates*

377 *3.2.1 Preliminary model comparisons: transmission dispersion and age groupings*

378 We compared models that produced estimates of each individual transmission dispersion parameter (k_u, k_v, k_w) to
379 simpler models of transmission variability. First, the simpler model with exponentially distributed transmissibility
380 ($k_u = k_v = k_w = 1$) produced similar maximal likelihood to the full model and a superior AIC, while the model
381 with no variability in transmissibility ($k_u = k_v = k_w = \infty$) produced inferior maximal likelihood (Table D in
382 Supporting Information). For simplicity we fixed the most parsimonious $k_u = k_v = k_w = 1$ assumption for all
383 following models.

384 We also ran our MLE procedure for each possible configuration of the (U, V, W) transmission groups (Table E in
385 Supporting Information). The best configuration producing the maximal log-likelihood occurred when the oldest
386 group, reference group U, included middle aged and aged individuals (age 45+), the middle group V included adults
387 excluding young adults (age 25-44), and the youngest group W included children, adolescents, and young adults
388 (age 0-24). We used those three age ranges for the three transmission age groups in all following models.

389 *3.2.2 Full model results*

390 The model with the assumptions defined in the prior sub-section and no other simplifying assumptions is called our
391 “full model” and includes 34 free parameters. We derived the MLE and confidence intervals for all 34 parameters
392 (Table 2).

393 **Table 2: Full model MLE results with confidence intervals**

Parameter	MLE (95% CI)	394
Mean community acquisition probability		
– infant/pre-school 0-5 (p_a)	3.2% (0.6% – 6.8%)	
– child 6-12 (p_b)	2.0% (0.6% – 4.1%)	
– adolescent 13-18 (p_c)	15.1% (11.4% – 19.4%)	
– young adult 19-24 (p_d)	12.6% (8.8% – 17.2%)	
– adult 25-44 (p_e)	6.7% (5.5% – 8.0%)	
– middle aged 45-64 (p_f)	6.7% (5.3% – 8.4%)	
– aged 65-79 (p_g)	5.2% (3.7% – 7.1%)	
– aged, 80 and over (p_h)	0.0% (0.0% – 3.4%)	
HSAR to age 45+ adults in 2-person households		
– From middle aged and older adults aged 45+ (p_{uu2})	68% (56% – 77%)	
– From adults aged 25-44 (p_{vu2})	45% (11% – 66%)	
– From children, adolescents, and young adults aged 0-24 (p_{wu2})	42% (23% – 57%)	
HSAR to age 25-44 adults in 2-person households		
– From middle aged and older adults aged 45+ (p_{uv2})	52% (22% – 73%)	
– From adults aged 25-44 (p_{vv2})	44% (29% – 58%)	
– From children, adolescents, and young adults aged 0-24 (p_{wv2})	63% (50% – 74%)	
HSAR to age 0-24 individuals in 2-person households		
– From middle aged and older adults aged 45+ (p_{uw2})	34% (11% – 53%)	
– From adults aged 25-44 (p_{vw2})	47% (31% – 60%)	
– From children, adolescents, and young adults aged 0-24 (p_{ww2})	45% (33% – 56%)	
Altered transmission coefficient to members of larger households		
– with 3 or 4 members ($\sigma_{3,4}$)	0.43 (0.33 – 0.57)	
– with 5 or more members (σ_{5+})	0.22 (0.16 – 0.29)	
Probability that prior infected person reported receiving a prior positive test		
– infant / pre-school, 0-5 (ϕ_{sa})	16% (6% – 32%)	
– child, 6-12 (ϕ_{sb})	36% (23% – 49%)	
– adolescent, 13-18 (ϕ_{sc})	53% (42% – 63%)	
– young adult 19-24 (ϕ_{sd})	62% (47% – 75%)	
– adult 25-44 (ϕ_{se})	59% (52% – 65%)	
– middle aged 45-64 (ϕ_{sf})	73% (65% – 80%)	
– aged 65-79 (ϕ_{sg})	73% (61% – 83%)	
– aged, 80 and over (ϕ_{sh})	100% (38% – 100%)	
Probability that prior infected person (any age) tested positive for antibodies:		
– Abbot (ϕ_{AA})	72% (68% – 76%)	
– Euroimmun (ϕ_{AE})	99.2% (97.5% – 99.98%)	
– Siemens (ϕ_{AS})	87% (83% – 90%)	
Probability that uninfected person did not report receiving a prior positive test		
– any age (π_s)	99.2% (98.9% – 99.4%)	
Probability that uninfected person (any age) tested negative for antibodies		
– Abbot (π_{AA})	99.7% (99.4% – 99.8%)	
– Euroimmun (π_{AE})	98.1% (97.6% – 98.6%)	
– Siemens (π_{AS})	99.85% (99.65% – 99.95%)	

395 MLE = maximum likelihood estimate; 95% CI = confidence interval, derived from comparing the likelihood ratio test statistic to the chi-squared
 396 distribution with one degree of freedom.

397 For community acquisition probability, the highest estimates occurred for adolescents: $p_c = 15.1\%$ (11.4% – 19.4%)
 398 and young adults: $p_d = 12.6\%$ (8.8% – 17.2%). The confidence intervals for those estimates did not overlap with
 399 those of any other age group, younger or older. The community acquisition probability for children aged 6–12: $p_b =$
 400 2.0% (0.1% – 4.1%) was substantially lower than the adjacent age group of adolescents, and p_b was also lower than
 401 the result for adults aged 25-44: $p_e = 6.7\%$ (5.5% – 8.0%) and middle-aged adults: $p_f = 6.7\%$ (5.3% – 8.4%). The

402 result for the oldest age group, 80 and over: $p_h = 0.0\%$ (0.0% – 3.4%), was substantially lower than any other adult
403 or adolescent age group.

404 Our results also exhibit age differences in directional household transmission probabilities (HSARs) at given
405 household sizes. The highest estimated average HSAR among all the 9 possible directional age group pairings in this
406 model was from middle-aged and older adults (age 45+) to other middle-aged and older adults: $p_{uu2} = 68\%$ (56% –
407 77%) in size-2 households. The lowest estimate for size-2 households was from middle-aged and older adults to
408 children, adolescents, and young adults (age 0-24): $p_{uw2} = 34\%$ (11% – 53%). We also found significantly smaller
409 acquisition probability for those living in larger households versus smaller ones: $\sigma_{34} = 0.43$ (0.33 – 0.57), and
410 $\sigma_{5+} = 0.22$ (0.16 – 0.29). This means that our model estimates that each age-to-age HSAR estimate for size-2
411 households (Table 2) is lower for size 3-4 households and lower still for size 5+ households. For example, our MLE
412 HSAR estimates among age 45+ adults living together in the three different household size groups are $p_{uu2} = 68\%$,
413 $p_{uu34} = 43\%$, and $p_{uu5+} = 31\%$.

414 We found age-based differences in the probability that individuals with prior infections tested positive during their
415 infection (as reported on our survey). Infants and preschoolers (age 0–5) had the lowest probability estimate: $\phi_{sa} =$
416 16% (6% – 33%), with a confidence interval that overlapped only with that of children aged 6–12: $\phi_{sb} = 36\%$ (23%
417 – 49%), reflective of lower infection ascertainment for young children. The MLE for each other age group was
418 above 50%, with middle aged and older individuals having the highest estimates at 73% or higher. The false-positive
419 rate estimate for survey-reported prior test results among individuals with no prior infection was 0.8%: $\pi_s = 99.2\%$
420 (98.9% – 99.4%).

421 We also estimated sensitivity and specificity parameters for each of the antibody tests. Our sensitivity estimates
422 were substantially different for the three tests, with Euroimmun highest at $\phi_{AE} = 99.2\%$ (97.5% – 99.98%) followed
423 by Siemens, $\phi_{AS} = 87\%$ (83% – 90%), and Abbott, $\phi_{AA} = 72\%$ (68% – 76%). Our sensitivity estimate for
424 Euroimmun was higher than what the test manufacturer found from samples of individuals who had tested PCR
425 positive 14+ days earlier. By contrast, our sensitivity estimates for Siemens and Abbott were lower than the
426 respective test manufacturer estimates.

427 For antibody test specificity, our results produced the lowest estimate for Euroimmun, $\pi_{AE} = 98.1\%$ (97.5% –
428 98.6%) and substantially higher estimates for Abbott, $\pi_{AA} = 99.7\%$ (99.4% – 99.8%) and Siemens, $\pi_{AS} = 99.85\%$

429 (99.65% – 99.95%). Those specificity intervals overlap those obtained by each test manufacturer from testing stored,
 430 presumed negative samples collected prior to the COVID-19 pandemic: 99.0% (98.4% – 99.4%) for Euroimmun
 431 [22], 99.6% (99.1% – 99.9%) for Abbott [23], and 99.89% (99.66% – 99.98%) for Siemens [24].

432 3.2.3 Equal contact and reciprocal transmission models

433 The “equal contact” model, in which the assumptions $\sigma_{uv} = \sigma_{vv} = \sigma_{wv}$ and $\sigma_{uw} = \sigma_{vw} = \sigma_{ww}$ are enforced,
 434 produces a MLE result (Table 3) in which the oldest transmission age group has the highest transmissibility, the
 435 middle transmission age group has the lowest transmissibility and the highest susceptibility, and the youngest age
 436 group has the lowest susceptibility. However, the maximal likelihood of this estimate is much poorer than that of the
 437 full model, such that the equal contact model can be rejected by the likelihood ratio test ($P = 0.01$) and has a poorer
 438 AIC score (difference of 5.0) despite using 3 fewer parameters than the full model. By contrast, the “reciprocal
 439 transmission” model, in which the assumptions $p_{uv} = p_{vu}$, $p_{uw} = p_{wu}$, and $p_{vw} = p_{wv}$ are enforced, cannot be
 440 rejected by the likelihood ratio test and produces a superior AIC score (difference of 4.2) to the full model (Table 3)
 441 due to using 3 fewer parameters while achieving a similar likelihood at the MLE.

442 **Table 3. Comparison of equal contact and reciprocal transmission models to full model**

	\hat{p}_a	\hat{p}_{ww2}	\hat{p}_{wv2}	\hat{p}_{wu2}	σ_{34}	$\hat{\phi}_{Sa}$	$\hat{\pi}_S$	Log likelihood	P value	Par	ΔAIC
	\vdots	\hat{p}_{vw2}	\hat{p}_{vv2}	\hat{p}_{vu2}	σ_{5+}	\vdots	$\hat{\pi}_{AA}$				
	\hat{p}_h	\hat{p}_{uw2}	\hat{p}_{uv2}	\hat{p}_{uu2}		$\hat{\phi}_{Sh}$	$\hat{\phi}_{AA}$				
							$\hat{\phi}_{AE}$				
							$\hat{\pi}_{AS}$				
Full model	3.2%	45%	63%	42%	0.43	16%	99.2%	-3021.0	Ref	34	Ref
	2.0%	47%	44%	45%	0.22	36%	99.7%				
	15.1%	34%	52%	68%		53%	99.2%				
	12.6%					62%	87%				
	6.7%					59%					
	6.7%					73%					
	5.2%					73%					
	0.0%					100%					
Equal contact model	3.6%	44%	57%	53%	0.42	16%	99.2%	-3026.5	0.03	30	3.0
	2.1%	37%	50%	46%	0.22	36%	99.7%				
	15.0%	50%	62%	58%		52%	99.2%				
	12.6%					62%	87%				
	6.6%					58%					
	6.8%					73%					
	5.3%					73%					
	0.0%					100%					
Reciprocal transmission model	2.7%	39%	54%	37%	0.45	16%	99.1%	-3021.8	0.62	31	-4.2
	1.7%	54%	45%	49%	0.24	35%	99.7%				
	14.5%	37%	49%	67%		53%	99.2%				
	12.4%					62%	87%				
	7.0%					59%					
	6.8%					73%					
	5.2%					73%					
	0.0%					100%					

443 The equal contact model estimates $\sigma_{uv} = \sigma_{vv} = \sigma_{wv} = 1.17$ and $\sigma_{uw} = \sigma_{vw} = \sigma_{ww} = 0.73$ are used when calculating the 9 HSAR values shown.
444 The reciprocal transmission model enforces $p_{vu} = p_{uv}$, $p_{wu} = p_{uw}$, and $p_{wv} = p_{vw}$. P values are for rejection of the model in favor of the full
445 model, by the likelihood ratio test with degrees of freedom equal to the difference in the number of parameters estimated (Par). $\Delta AIC =$
446 difference in Akaike information criterion value versus the reference (Ref) model; a negative ΔAIC means the model is favored over the
447 reference model.

448 **3.2.4 Most parsimonious model results**

449 We tested a series of models using simplifying assumptions determining the age-specific HSAR values, using fewer
450 free parameters to determine the 9 values. The models enforcing equal HSAR to and/or from age group V (age 25-
451 44) produced better AIC scores than those that did not (Table F in Supporting Information). When combining this
452 with other favored assumptions, the best AIC score was produced by a 5-parameter model that used 3 HSAR values
453 to parameterize the 3-by-3 matrix of directional HSAR by age group for a particular household size group, plus the 2
454 size-effect parameters σ_{34} and σ_{5+} (Table F in Supporting Information). This model retains the assumption of equal
455 AR between age group V (age 25-44) and all age groups in either direction, and also equal HSAR between age
456 group W (age 0-24) and age groups W and U (age 45+) in either direction, which leaves the HSAR among age
457 group U as the third independent value for a particular household size group.

458 We also tested simpler models that used two rather than three transmission age groups by consolidating adjacent
459 groups of our three-group model (Table F in Supporting Information). The most parsimonious among those two-
460 group models used ages 0-44 and 45+ as the two groups, with reciprocal transmission – it produced a likelihood
461 inferior to the best three-age-group model just described, with the same number of estimated parameters. Our testing
462 also included even simpler models that used fewer than 5 transmission parameters (Table G in Supporting
463 Information), which could each be rejected in favor of the most parsimonious model ($P \leq 0.03$). Those rejected
464 models include models with no age-difference assumptions on transmission (household size being the only correlate
465 with HSAR) or no household size effect ($\sigma_{34} = \sigma_{5+} = 1$). The simplest model with no age- or size-structure
466 produced a single HSAR estimate of 27% applying to all ages and household sizes, but that result could be strongly
467 rejected ($P < 0.0001$).

468 We tested whether adjacent age groups could be consolidated to use fewer free parameters for community
469 acquisition and case ascertainment while achieving comparable model fit. Guided by comparing confidence
470 intervals from the full model MLE results, we tested a model that combined the age groups for community

471 acquisition to reduce the number of groups from 8 to 4, where the 4 groups are ages 0–12 ($p_a = p_b$), 13–24 ($p_c =$
472 p_d), 25–79 ($p_e = p_f = p_g$), and 80+ (p_h). For case ascertainment, we reduced the number of groups from 8 to 3:
473 ages 0–12 ($\phi_{sa} = \phi_{sb}$), 13–44 ($\phi_{sc} = \phi_{sd} = \phi_{se}$), and 45+ ($\phi_{sf} = \phi_{sg} = \phi_{sh}$). Compared to the full model, the
474 MLE for the model with those new groupings for both community acquisition and case ascertainment could not be
475 rejected by the likelihood ratio test ($P = 0.46$) and produced superior AIC score with a difference of 9.2 (Table H in
476 Supporting Information). Models using alternate groupings using the same or fewer numbers of age groups
477 produced lower maximal likelihood and inferior AIC scores.

478 The model combining all the simplifying assumptions producing the best AIC scores in the above comparisons is the
479 most parsimonious overall model we found, producing the best overall AIC score. The most parsimonious model
480 includes 3 household transmission parameters for age-specific HSAR, 2 for household size-adjustment coefficients
481 for HSAR, 4 age-specific importation parameters, 3 age-specific prior case ascertainment parameters, and 7 age-
482 independent sensitivity and specificity parameters (Table 4). The HSAR values shown apply only between members
483 of 2-person households: 66% HSAR between age 45+ individuals residing together, 47% between age 25-44
484 individuals and anyone else, and 35% between remaining pairings involving age 0-24 individuals with each other or
485 with age 45+ individuals. For households with 3 or 4 residents, the model adjusts those HSAR estimates down to
486 49%, 31%, and 21%, respectively, and for households with 5 or more residents, the respective HSAR estimates are
487 35%, 21%, and 13%.

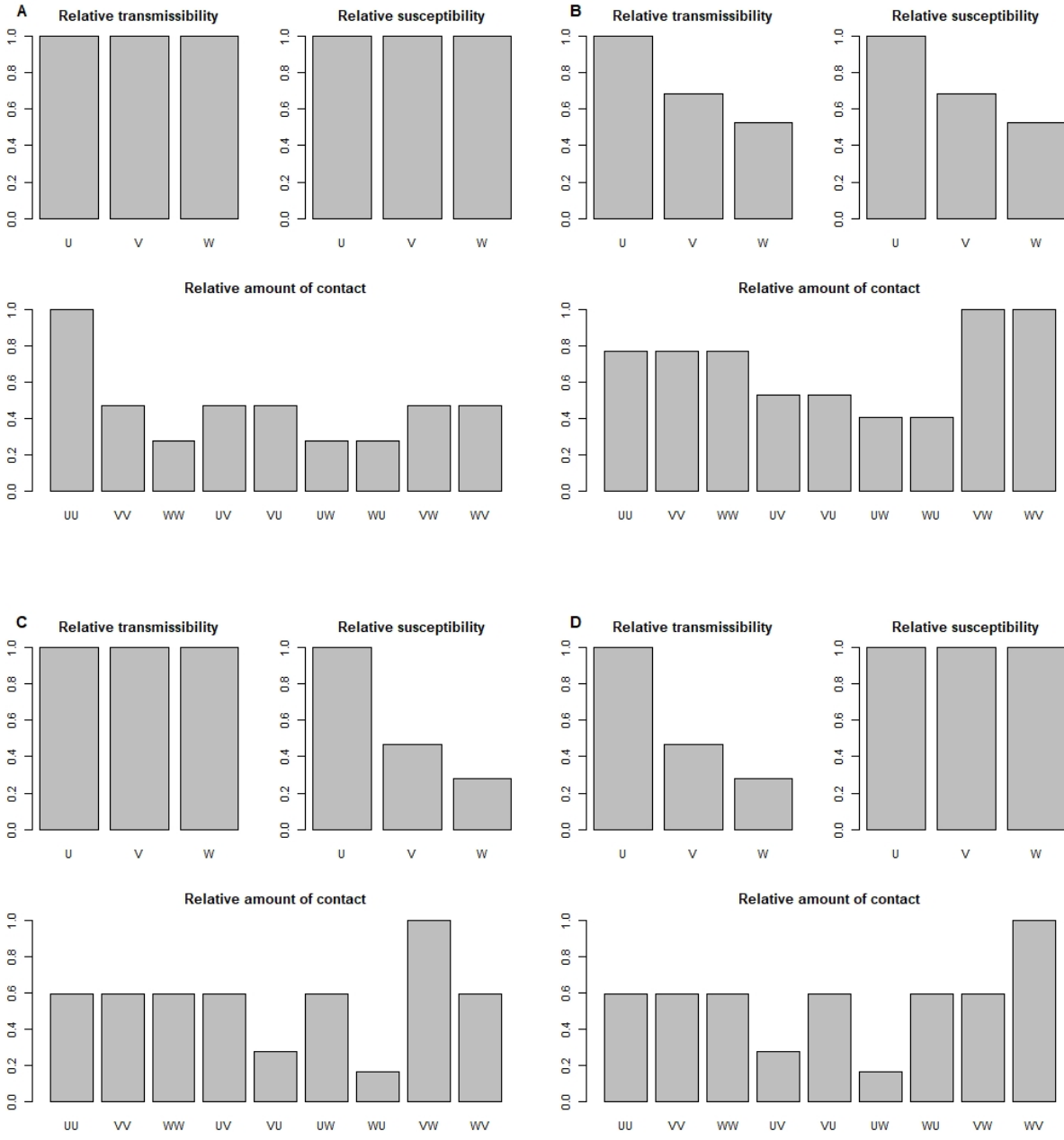
488 **Table 4. Most parsimonious model with confidence intervals**

Parameter	MLE (95% CI)
Mean community acquisition probability	
– infant / pre-school / child 0-12 ($p_{a,b}$)	2.1% (0.8% – 3.8%)
– adolescent / young adult 13-24 ($p_{c,d}$)	13.7% (10.9% – 16.7%)
– adult / middle aged / aged 25-79 ($p_{e,f,g}$)	6.6% (5.8% – 7.4%)
– aged, 80 and over (p_h)	0.0% (0.0% – 3.4%)
Mean HSAR in 2-person households	
– Among age 45+ (p_{uu2})	66% (54% – 75%)
– From or to age 25-44 ($p_{vu2,vv2,vw2,uv2,wv2}$)	47% (41% – 54%)
– Among age 0-24 and between age 0-24 and age 45+ ($p_{ww2,uw2,wu2}$)	35% (27% – 42%)
Altered exposure / susceptibility of members of larger households	
– with 3 or 4 members (σ_{34})	0.51 (0.39 – 0.67)
– with 5 or more members (σ_{5+})	0.29 (0.21 – 0.38)
Probability that prior infected person reported receiving a prior positive test	
– infant / pre-school / child, 0-12 ($\phi_{sa,b}$)	28% (19% – 39%)
– adolescent / young adult / adult, 13-44 ($\phi_{sc,d,e}$)	57% (52% – 63%)
– middle aged and older, 45+ ($\phi_{sf,g,h}$)	73% (67% – 79%)
Probability that prior infected person (any age) tested positive for antibodies:	
– Abbot (ϕ_{AA})	72% (68% – 76%)
– Euroimmun (ϕ_{AE})	99.3% (97.6% – 100%)
– Siemens (ϕ_{AS})	87% (83% – 90%)
Probability that uninfected person did not report receiving a prior positive test	
– any age (π_s)	99.2% (98.9% – 99.4%)
Probability that uninfected person (any age) tested negative for antibodies	
– Abbot (π_{AA})	99.7% (99.4% – 99.8%)
– Euroimmun (π_{AE})	98.1% (97.6% – 98.6%)
– Siemens (π_{AS})	99.85% (99.65% – 99.95%)

489 MLE = maximum likelihood estimate; 95% CI = confidence interval, derived from comparing the likelihood ratio test statistic to the chi-squared
 490 distribution with one degree of freedom.

491 **3.3 Mechanisms of household transmission differences**

492 The three different age-specific HSARs at a given household size estimated by the most parsimonious model could
 493 be explained by a combination of age-based differences in transmissibility, susceptibility, and the amount of
 494 household contact. We cannot uniquely estimate the age differences for each of those three mechanisms
 495 simultaneously, but we can use our HSAR estimates to constrain their values and explore possible combinations that
 496 are consistent with our findings (Fig 1). We showed in section 3.2.3 that assuming an equal amount of household
 497 contact between all age groups is inconsistent with our results, as age differences in transmissibility and
 498 susceptibility alone cannot reproduce the attack rate matrix. Thus, all depicted possibilities exhibit age-based contact
 499 differences (Fig 1). However, uncertainty remains about the potential role of age-based transmissibility and
 500 susceptibility differences in combination with different magnitudes of contact differences.



501

502

503 **Fig 1. Possible mechanisms of transmission differences consistent with model estimates.** W: age group 0-24; V: age group 25-44; U: age
 504 group 45+; labels 'XY' in relative-amount-of-contact figures refer to the amount of contact between age groups X and Y when X is infected and
 505 Y is susceptible. (A) Relative amount of contact when assuming no age difference in susceptibility nor transmissibility; (B) results when
 506 assuming equal relative age differences in both susceptibility and transmissibility and no age differences in within-age-group contact; (C) results
 507 when assuming no age difference in transmissibility nor in within-age-group contact; (D) results when assuming no age differences in
 508 susceptibility nor in within-age-group contact.

509 One possibility consistent with our results is that transmissibility and susceptibility per contact are independent of
510 age, and that attack rate differences are entirely explained by contact differences (Fig 1A). With this scenario, we
511 estimate the amount of contact between age 25-44 adults and their housemates of any age would have to be about
512 47% the amount that age 45+ adults contact each other. The amount of contact between age 0-24 and age 0-24 or
513 45+ would be about 28% the amount of contact within the over-45 age group.

514 Another possibility is that within-age-group contacts are the same, and the increase of within-group attack rates is
515 explained by both susceptibility and transmissibility increasing with age by equal proportions (Fig 1B). With this
516 scenario, we estimate that susceptibility and transmissibility of those aged 25-44 and 0-24 are about 68% and 53%
517 less than those aged 45+, respectively. This scenario also requires that the highest amount of contact occurred
518 between age 25-44 and age 0-24 individuals.

519 It is also possible that age groups have equal transmissibility but different susceptibility (Fig 1C) or vice-versa (Fig
520 1D). However, those scenarios require that contact behavior depends on the age group of the infected person in an
521 infected-susceptible pairing. For example, if aged 25-44 adults are more susceptible than children but equally
522 transmissible, the amount of household contact in a potential transmission scenario between adults and children
523 must be higher when adults were infected and lower when children were infected.

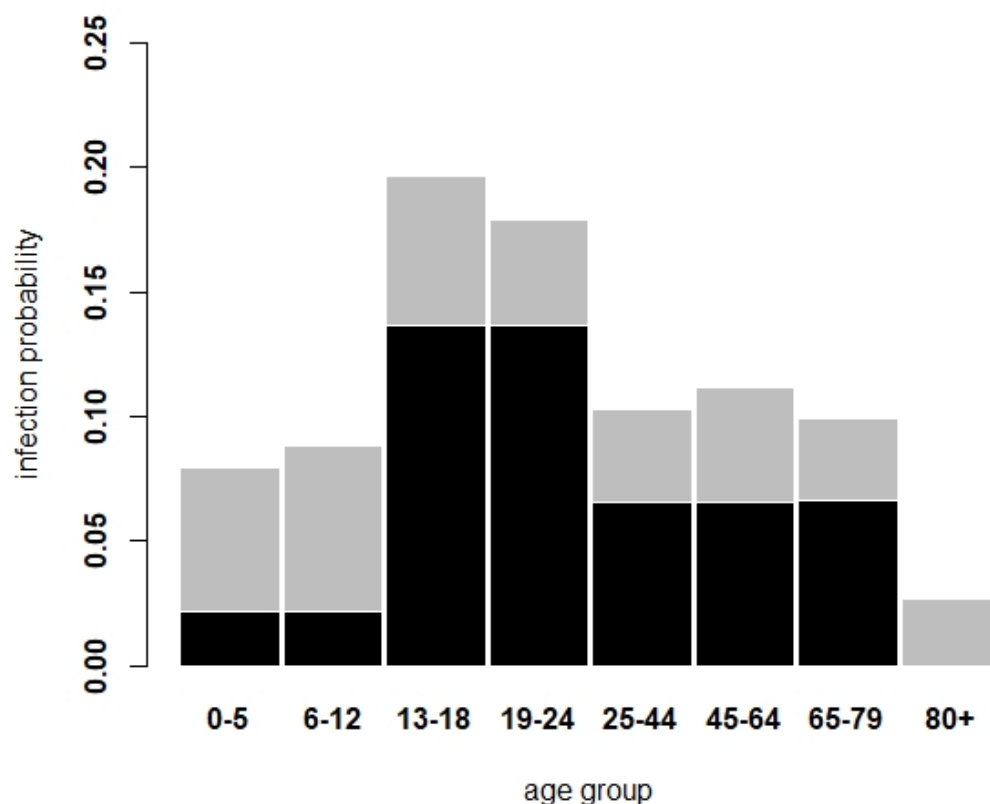
524 The depicted scenarios illustrate certain boundary cases and are not exhaustive of all possibilities. We derived a full
525 mathematical description of the possible mechanistic parameter combinations consistent with our MLE results
526 (Supporting Information).

527 ***3.4 Infection risk by age group***

528 Estimates from the most parsimonious model exhibit notable age differences in the overall risk of infection prior to
529 data collection, as well as in relative contribution of non-household and household sources (Fig 2). Members of the
530 youngest (0-12 years) and oldest (80+ years) age groups experienced higher probability of within-household
531 acquisition compared to their risk of acquiring infection from non-household sources. Those in the intermediate age
532 groups experienced most of their acquisition from non-household sources, but within-household sources still
533 contributed between 23% and 41% of their overall infection probability.

534 Children and adolescents were estimated to have had the highest absolute probability of acquiring infection within
535 the household (6–7%), despite being members of the transmission age group with the lowest HSAR estimates. This
536 was because they lived in larger households than an average adult and were more likely to be exposed to an
537 introduction. Relative to other age groups, adolescents experienced high acquisition risk from both non-household
538 and within-household sources.

539 The age 25-44 and 45-79 groups experienced similar overall within-household acquisition risk despite the
540 differences in their transmission probabilities; the higher HSAR of age 45+ individuals among themselves was
541 balanced out by their lower rate of acquisition from the youngest age groups. Finally, the age 80+ age group were
542 estimated to have negligible non-household acquisition risk and the lowest within-household risk, the latter due to
543 their living with individuals in the younger groups less frequently.



544

545 **Fig 2. Estimated infection probability of individuals from five age groups from non-household (black) and within-household (grey)**
546 **sources.** Estimates were derived using parameters from the most parsimonious model (Table 4).

547 **4 Discussion**

548 Our novel mathematical model of household infection, transmission, and test data accuracy can simultaneously
549 incorporate multiple highly important features of SARS-CoV-2: variability in transmissibility of infected people and
550 heterogeneity in non-household acquisition, case ascertainment, and household transmission across age groups and
551 household sizes. Our model could produce nuanced estimates for household transmission dynamics from data and
552 provide novel insights about potential mechanisms of disparate infection rates across households of different sizes
553 and age combinations.

554 Our findings show strong evidence that adolescents and young adults aged 13-24 were more likely than younger
555 children and older adults to acquire SARS-CoV-2 infection outside the household prior to data collection in early

556 2021 in Utah. Children aged 12 and younger were less likely to acquire outside the household than any adult group
557 other than those aged 80 and over. We found little evidence of any acquisitions by the age 80+ group from outside
558 the household, suggesting that the primary risk to these elderly individuals living in households may have been
559 transmission from their housemates. Our data did not include elderly residing in congregate living facilities such as
560 nursing homes, where residents likely experienced a much different risk profile compared to aged individuals
561 covered by our sampling of households [29].

562 The results also suggest that there were differences in household transmission rates across age group pairings. The
563 highest age-specific HSAR for a given household size was from middle-aged and older adults (age 45+) to other
564 middle-aged and older adults. The next-highest HSAR estimate was for adults aged 25-44, from and to all age
565 groups. The lowest HSAR estimate was for those younger than 25 among each other and to and from those aged 45
566 and older. That age pattern of attack rates was inconsistent with a model that assumes an age-independent rate of
567 contact among infected and susceptible pairs of housemates. With that equal-contact model, the age group most
568 susceptible to acquiring infection for a given amount of infectious contact would experience the highest attack rate
569 from all three age groups. However, our findings showed that, while the 45-and-older group experienced the highest
570 attack rate from age 45+ housemates, they experienced a lower attack rate from the under-25 group compared to that
571 experienced by the 25-44 group. Similarly, while infected 45-and-older individuals exhibited the highest attack rate
572 to other 45-and-older housemates, they transmitted at a lower rate to under-25 housemates compared to the
573 corresponding attack rate from the 25-44 group. Therefore, our findings suggest that contact within households was
574 not uniform across age groups.

575 Our finding that the attack rate estimates were similar in both directions for each given age-group pairing is
576 consistent with a hypothesis that there are not substantial age differences in susceptibility or transmissibility for a
577 given rate of contact. For example, if 45-and-older individuals were substantially more susceptible to acquiring
578 infection from a given contact compared to those under 25, one might expect the under-25 to over-45 attack rate to
579 be higher than the over-45 to under-25 attack rate. However, it is possible that an equal attack rate in both directions
580 could be explained by multiple mechanistic differences that cancel each other out, such as age-dependent
581 susceptibility, transmissibility, or contact behavior changes during infection.

582 Additional data on age-structured, transmission-relevant contact behavior in households would help further constrain
583 the possible mechanisms driving the age differences in transmission we found. For example, if data showed that
584 contact between younger parents and younger children (i.e. between “V” and “W” groups in our model) was more
585 substantial than contact between other age pairings within households, and especially if that contact persisted during
586 illness because of the necessities of childcare [30], then our findings would suggest stronger evidence that younger
587 age groups had lower susceptibility and/or transmissibility per contact compared to older age groups (Fig 1 panels
588 B-D).

589 We also found lower HSARs for larger household sizes, even when controlling for our age difference findings. This
590 finding is consistent with other household transmission studies [11-13] that favored models with transmission rates
591 scaled down with increasing household size, e.g. via frequency-dependent-like transmission assumptions. The
592 precise mechanisms by which transmission probabilities for a given age-to-age pairing might decrease with
593 increasing household size are unclear. There could be contact rate differences among members of larger vs. smaller
594 families [12], or risk differences due to domicile size and layout or demographic factors other than age that correlate
595 with household size. Differences in precaution behavior to prevent transmission might also play a role, particularly
596 in later generations of infection after a given household introduction, which would be a larger factor for more
597 populous households where multiple rounds of transmission are possible.

598 In combination, our findings for age-dependency and household size-dependency of household transmission result
599 in dramatically HSAR depending on the scenario. Our model’s HSAR estimates range from 66% for middle-aged or
600 older adults living alone together, down to 13% among age 0-24 individuals or between age 0-24 and 45+
601 individuals living together in households of size 5 or more. This finding suggests that standard estimates of a single
602 averaged HSAR can be highly misleading if used to judge the risk of transmission in all households.

603 Our estimates quantifying transmission variability are consistent with our prior finding that the low-variability
604 binomial transmission assumption traditionally used in household models can be rejected in favor of models
605 incorporating higher variability (lower value of a dispersion parameter) for individual transmissibility [18]. Notably,
606 our model factors out potential sources of variability that could correlate with age and/or household size. Yet we still
607 found that incorporating additional variability for individual transmissibility within members of the same age/size

608 groups was warranted, suggesting that other sources of variability beyond those determined by age and household
609 size are important to represent [2].

610 The dispersion parameter, k , that we used to quantify transmission variability is analogous to the dispersion
611 parameter k commonly used in negative binomial transmission models, in that it derives from the shape parameter of
612 gamma-distributed individual transmissibility [1]. However, in the negative binomial model, k represents all
613 potential contributions to variability, including variability in number of people contacted [31] and variation in non-
614 contact-based characteristics that might correlate with age. Our model explicitly represents those features and leaves
615 k to represent any remaining contributions to variability. I.e., we use k to represent variability of transmission
616 probability to *each* housemate of a given age group, from a given age group, in a household of a given size.
617 Therefore, if contact degree and and/or age/size-based differences contribute significantly to overall transmission
618 variation, we would expect our estimate of k to be higher (lower variability) than estimated k from a negative
619 binomial model. Indeed, a systematic review [32] found that 93% of 28 studies estimated $k < 1$ for the negative
620 binomial model fit to COVID-19 transmission data, whereas we did not find strong support for $k < 1$ in our model (k
621 = 1 could not be rejected and produced the most parsimonious model fit). However, our rejection of the lower
622 variance binomial assumption ($k \rightarrow \infty$) suggests the possibility that a large portion of overall transmission
623 variability could be attributed to factors other than number of contacts or age/size-associated differences, such as the
624 duration of infectiousness, organism shedding rate [33], or individual contact timing and intensity.

625 Our detailed data including results from up to three different antibody tests also presented a unique opportunity for
626 comparing test accuracies against each other and against data from the test manufacturers or other sources. Our
627 independent estimates of the antibody test specificities were remarkably consistent with results from independent
628 data. However, our results for antibody test sensitivities included some disagreements with other sources,
629 particularly for the Euroimmun test. Further research could illuminate why our Euroimmun sensitivity estimates
630 were so high, perhaps by building a model that utilizes the quantitative Euroimmun results to investigate whether the
631 manufacturer-suggested positive/indeterminate/negative thresholds were less appropriate for our implementation.
632 Our extensive antibody test data also allowed us to estimate case ascertainment across age groups. By comparing
633 survey data about whether antibody-positive individuals had ever tested positive prior to serological testing, we
634 found a high level of age dependence: younger individuals with a prior infection were substantially less likely to

635 have tested positive during their infection compared to older age groups. This finding provides evidence that case
636 ascertainment in Utah had been substantially lower for children aged 0-12 compared to adolescents and adults, and
637 somewhat lower for those aged 13-44 compared to those middle-aged and older, perhaps suggesting that those
638 younger groups were less likely to experience symptoms that warranted seeking testing [34].

639 This study has several limitations. We did not test factors other than age composition and household size that might
640 affect the way community acquisition risk or within-household transmission varies by household. Some households
641 could have experienced higher collective community acquisition risk than others of similar age composition due to
642 living in higher-risk locations, common work or school exposure, or common social exposure. Household
643 transmission could also be driven by non-age-based properties of households such as age-independent contact
644 behavior, health composition of household members, physical properties of the domicile such as size and
645 ventilation, or other properties that could increase infection or transmission risk of all household members together.
646 It is possible that further expansion of our model to include correlates with non-age data and/or household-level
647 variability distributions for community acquisition or household transmission could provide further insights, but we
648 suspect that these additions would push the limits of parameter identifiability against our data.

649 Some members of participating households might have been actively or very recently infected with SARS-CoV-2 at
650 the time of sampling for serological testing. When ongoing or recent household outbreaks involved multiple
651 infections, some of the infected members could have tested seropositive while others had yet to become infected or
652 had yet to seroconvert on the day of sampling. For that scenario, our assumption that household outbreaks had
653 reached their final size would be violated, contributing to underestimates of HSAR. We suspect that the proportion
654 of those households was low, given the short time window in which those conditions would apply in a multi-person
655 outbreak, and households might have been unlikely to choose to travel to a testing site while experiencing or
656 recovering from symptomatic infection.

657 Our findings may contain biases related to non-participation rates of households that were selected and approached
658 for inclusion in the study. Our data collection in the school district testing phase included a complicated sampling
659 design with weights introduced partly to account for different rates of nonresponse across demographic strata. For
660 simplicity we ignored these details and sampling weights for the analysis presented here. In the hotspot testing phase
661 households were selected for invitation by simple random sample of addresses, and those choosing to participate

662 may not be representative of those areas. Also, while the included sampling phases and locations cover a broad area
663 of the most populous region in Utah, there may be important differences in households both within those areas and
664 from the areas of Utah not covered. Thus, households with higher COVID-19 risk may be overrepresented or
665 underrepresented in our data relative to their frequency in the broader population of households in Utah.

666 We used participants' prior vaccination information only to indicate whether positive test results for the two spike-
667 protein-based antibody tests might have been triggered by the vaccine rather than prior SARS-CoV-2 infection; we
668 did not attempt to assess whether the vaccine produced a protective affect against community and/or within-
669 household acquisition in the period after receiving one or two vaccine doses. Our raw data show that vaccinated
670 individuals tested positive by the Abbott (nucleocapsid) test at a lower rate than non-vaccinated individuals,
671 particularly in the 65-79 age group, which had received higher vaccination coverage than younger age groups. Thus,
672 it is possible that including a vaccination effect parameter on community acquisition or household transmission
673 probability would produce a result suggesting a reduced infection probability. However, such a result would not be
674 conclusive of a causal effect of the vaccine on acquisition risk, for the following reasons.

675 As vaccines had only recently become available in early 2021, only 12% of participants had received the full two-
676 dose series prior to serological sampling, and about half of those had received their second dose less than one month
677 prior. Thus, even the minority of those participants who were considered fully vaccinated at the time of testing had
678 spent only a small portion of their cumulative time at risk since the onset of the pandemic in a potentially protected
679 state. The subset of participants who had experienced both prior infection and vaccination could include both those
680 who were infected prior to vaccination as well as break-through infections post vaccination, and inferring the
681 relative timing of these events would be difficult with our data. Also, inferring a potential causal effect of
682 vaccination on infection risk would be hampered by confounders, including risk-based vaccination eligibility, the
683 potential reverse causal effect of known prior infection on the decision to vaccinate, and potential correlations
684 between exposure risk and vaccine access and/or refusal among those eligible. If the vaccinated participants did
685 experience substantial COVID-19 exposures between their vaccination and serological testing, and the vaccine
686 substantially reduced their likelihood of infection by those exposures, then our results for the relative acquisition
687 probabilities of older age groups (who had higher vaccine coverage) might be biased low.

688 Our dataset's collection period ended on April 10, 2021, just prior to the widespread emergence of the Delta variant
689 of SARS-CoV-2, which was first detected in Utah in mid-April 2021 and spread rapidly in the following weeks [35].
690 Thus, our estimates may not accurately reflect infection and transmission characteristics of dominant SARS-CoV-2
691 strains circulating in the U.S. and elsewhere from late Spring 2021 to the present, particularly because different
692 strains have been found to exhibit different household transmission rates [36]. Also, as the state of Utah has a
693 younger population, larger average household size, and other demographic differences compared to U.S. national
694 averages [18], our household-based epidemiological estimates might be atypical. However, as we expressed our
695 household transmission estimates as per-capita transmission probabilities from and to particular age groups and
696 household sizes, our findings could serve as tool to study how overall household transmission patterns would differ
697 in communities with different household size and age distributions.

698 In conclusion, we found evidence of important age differences in COVID-19 epidemiology in Utah households
699 during 2020 and early 2021, including differences in case ascertainment, non-household acquisition, and within-
700 household transmission. Household transmission differences by age occurred in patterns that suggest substantial
701 differences in household contact rates between different age groups. A better understanding of those differences
702 might be crucial to deciphering community-wide transmission patterns and mitigation efforts.

703 **5 Acknowledgments**

704 The authors acknowledge David Swerdlow, Fred Angulo, Jessica Atwell, and Farid Khan for inspiration for this data
705 modeling work and comments on this manuscript. We acknowledge Stephen Alder, Adam Looney, Tom Greene,
706 Brian Orleans, Yue Zhang, and other members of the Utah HERO Project team and the University of Utah Study
707 Design & Biostatistics Center who helped design and/or implement the household participant recruitment and
708 survey effort. We acknowledge Julio Delgado for directing the validation and implementation of the antibody assays
709 at ARUP Laboratories and overseeing the validation of finger prick sample collection, as well as the staff at ARUP
710 Laboratories who processed and tested sample specimens to produce the antibody data used in our analysis.

711 **6 References**

- 712 1. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual
713 variation on disease emergence. *Nature* **2005**; 438(7066): 355-9.
- 714 2. VanderWaal KL, Ezenwa VO. Heterogeneity in pathogen transmission: mechanisms and
715 methodology. *Functional Ecology* **2016**; 30(10): 1606-22.

- 716 3. Pellis L, Cauchemez S, Ferguson NM, Fraser C. Systematic selection between age and household
717 structure for models aimed at emerging epidemic predictions. *Nat Commun* **2020**; 11(1): 906.
- 718 4. Buckee C, Noor A, Sattenspiel L. Thinking clearly about social aspects of infectious disease
719 transmission. *Nature* **2021**; 595(7866): 205-13.
- 720 5. Ma KC, Menkir TF, Kissler S, Grad YH, Lipsitch M. Modeling the impact of racial and ethnic
721 disparities on COVID-19 epidemic dynamics. *Elife* **2021**; 10.
- 722 6. Ferguson NM, Donnelly CA, Anderson RM. Transmission dynamics and epidemiology of dengue:
723 insights from age-stratified sero-prevalence surveys. *Philos Trans R Soc Lond B Biol Sci* **1999**;
724 354(1384): 757-68.
- 725 7. Meyerowitz EA, Richterman A, Gandhi RT, Sax PE. Transmission of SARS-CoV-2: A Review of
726 Viral, Host, and Environmental Factors. *Ann Intern Med* **2021**; 174(1): 69-79.
- 727 8. Qian H, Miao T, Liu L, Zheng X, Luo D, Li Y. Indoor transmission of SARS-CoV-2. *Indoor Air* **2020**;
728 ina.12766.
- 729 9. Yonker LM, Neilan AM, Bartsch Y, et al. Pediatric Severe Acute Respiratory Syndrome
730 Coronavirus 2 (SARS-CoV-2): Clinical Presentation, Infectivity, and Immune Responses. *J Pediatr*
731 **2020**; 227: 45-52 e5.
- 732 10. Ghosh AK, Venkatraman S, Soroka O, et al. Association between overcrowded households,
733 multigenerational households, and COVID-19: a cohort study. *Public Health* **2021**; 198: 273-9.
- 734 11. Hart WS, Abbott S, Endo A, et al. Inference of the SARS-CoV-2 generation time using UK
735 household data. *Elife* **2022**; 11.
- 736 12. Endo A, Uchida M, Kucharski AJ, Funk S. Fine-scale family structure shapes influenza
737 transmission risk in households: Insights from primary schools in Matsumoto city, 2014/15. *PLoS*
738 *Comput Biol* **2019**; 15(12): e1007589.
- 739 13. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boelle PY. A Bayesian MCMC approach to study
740 transmission of influenza: application to household longitudinal data. *Stat Med* **2004**; 23(22):
741 3469-87.
- 742 14. Bubar KM, Reinholt K, Kissler SM, et al. Model-informed COVID-19 vaccine prioritization
743 strategies by age and serostatus. *Science* **2021**; 371(6532): 916-21.
- 744 15. Metcalf CJ, Farrar J, Cutts FT, et al. Use of serological surveys to generate key insights into the
745 changing global landscape of infectious disease. *Lancet* **2016**; 388(10045): 728-30.
- 746 16. Longini I, Koopman JS. Household and community transmission parameters from final
747 distributions of infections in households. *Biometrics* **1982**; 38(1): 115-26.
- 748 17. Bi Q, Lessler J, Eckerle I, et al. Insights into household transmission of SARS-CoV-2 from a
749 population-based serological survey. *Nat Commun* **2021**; 12(1): 3643.
- 750 18. Toth DJA, Beams AB, Keegan LT, et al. High variability in transmission of SARS-CoV-2 within
751 households and implications for control. *PLoS One* **2021**; 16(11): e0259097.
- 752 19. Utah Health and Economic Recovery Outreach Home - Utah Hero. Available at:
753 <https://eccles.utah.edu/utah-hero/>. Accessed 11/10/2023.
- 754 20. Samore MH, Looney A, Orleans B, et al. Probability-Based Estimates of Severe Acute Respiratory
755 Syndrome Coronavirus 2 Seroprevalence and Detection Fraction, Utah, USA. *Emerg Infect Dis*
756 **2021**; 27(11): 2786-94.
- 757 21. Hershov RB, Wu K, Lewis NM, et al. Low SARS-CoV-2 Transmission in Elementary Schools - Salt
758 Lake County, Utah, December 3, 2020-January 31, 2021. *MMWR Morb Mortal Wkly Rep* **2021**;
759 70(12): 442-8.
- 760 22. Anti-SARS-CoV-2 ELISA (IgG) - Instructions for Use. Available at:
761 <https://www.fda.gov/media/137609/download>. Accessed 10/20/2023.
- 762 23. SARS-CoV-2 IgG Architect - Instructions for Use. Available at:
763 <https://www.fda.gov/media/137383/download>. Accessed 1/25/2021.

- 764 24. Dimension Vista SARS-CoV-2 IgG (COV2G) - Instructions for Use. Available at:
765 <https://www.fda.gov/media/145020/download>. Accessed 10/20/2023.
- 766 25. Tang MS, Hock KG, Logsdon NM, et al. Clinical Performance of Two SARS-CoV-2 Serologic Assays.
767 *Clin Chem* **2020**; 66(8): 1055-62.
- 768 26. Rychert J, Couturier MR, Elgort M, et al. Evaluation of 3 SARS-CoV-2 IgG Antibody Assays and
769 Correlation with Neutralizing Antibodies. *J Appl Lab Med* **2021**; 6(3): 614-24.
- 770 27. Demmer RT, Baumgartner B, Wiggen TD, et al. Identification of Natural SARS-CoV-2 Infection in
771 Seroprevalence Studies Among Vaccinated Populations. *Mayo Clin Proc* **2022**; 97(4): 754-60.
- 772 28. Age Groups - MeSH - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/mesh/68009273>.
- 773 29. Cronin CJ, Evans WN. Nursing home quality, COVID-19 deaths, and excess mortality. *J Health
774 Econ* **2022**; 82: 102592.
- 775 30. Paul LA, Daneman N, Schwartz KL, et al. Association of Age and Pediatric Household
776 Transmission of SARS-CoV-2 Infection. *JAMA Pediatr* **2021**; 175(11): 1151-8.
- 777 31. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and
778 network models in epidemiology. *J R Soc Interface* **2007**; 4(16): 879-91.
- 779 32. Wegehaupt O, Endo A, Vassall A. Superspreading, overdispersion and their implications in the
780 SARS-CoV-2 (COVID-19) pandemic: a systematic review and meta-analysis of the literature. *BMC
781 Public Health* **2023**; 23(1): 1003.
- 782 33. Ke R, Martinez PP, Smith RL, et al. Daily longitudinal sampling of SARS-CoV-2 infection reveals
783 substantial heterogeneity in infectiousness. *Nat Microbiol* **2022**; 7(5): 640-52.
- 784 34. Davies NG, Klepac P, Liu Y, et al. Age-dependent effects in the transmission and control of
785 COVID-19 epidemics. *Nat Med* **2020**; 26(8): 1205-11.
- 786 35. Keegan LT, Truelove S, Lessler J. Analysis of Vaccine Effectiveness Against COVID-19 and the
787 Emergence of Delta and Other Variants of Concern in Utah. *JAMA Netw Open* **2021**; 4(12):
788 e2140906.
- 789 36. Madewell ZJ, Yang Y, Longini IM, Jr., Halloran ME, Dean NE. Household Secondary Attack Rates
790 of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-
791 analysis. *JAMA Netw Open* **2022**; 5(4): e229317.

792

793 **Supporting Information**

794 *Participant recruitment:*

795 Participants contributing to our dataset were recruited during two different phases of the Utah HERO Project. The
796 first phase targeted households within the geographic boundaries of two school districts (“school district testing”):
797 the Salt Lake City School District, identical to the boundaries of Salt Lake City, and the adjacent Granite School
798 District, which includes other urban areas in Salt Lake County. These geographies were chosen to enable analysis of
799 relationships between school policies [21] and community transmission of SARS-CoV-2 and to understand the
800 extent of infection in this urban population.

801 For the school district testing phase, participants were recruited and enrolled starting from January 7, 2021, and
802 samples were collected from January 14 to April 10, 2021. The source population consisted of all households in the
803 two school district boundaries as defined in 2020-2021, for a total of 83,353 and 138,397 possible households in Salt
804 Lake City and Granite School Districts, respectively. Households were chosen using a random sampling design as
805 previously described [20]. A total of 75,067 households were invited by mail to participate, of which 35,936 were in
806 the Salt Lake City School District and 39,131 were in the Granite School District.

807 The second phase of recruitment contributing to our dataset occurred in response to a spike in COVID-19 cases in
808 specific areas (“hotspot testing”). The Utah Department of Health had notified the Utah HERO Project team of an
809 especially high incidence of COVID-19 cases in early 2021 among communities in the southern portion of Salt Lake
810 County and the northern portion of Utah County. In response, the HERO Project conducted a simple random sample
811 of households in the following cities/towns in Salt Lake County (Draper, Highland, South Jordan, West Jordan) and
812 Utah County (Alpine, American Fork, Bluffdale, Cedar Hills, Eagle Mountain, Herriman, Lehi, Lindon, Pleasant
813 Grove, Riverton, Saratoga Springs) in the State of Utah. Households that had been tested during previous phases of
814 the Utah HERO Project were not eligible for participation in the hotspot testing phase. A total of 21,480 households
815 were invited by mail to participate: 10,601 in Salt Lake County and 10,879 in Utah County. Recruitment for hotspot
816 testing began in late February 2021, and samples were collected from March 2 to March 13, 2021.

817 For both the school district and hotspot testing phases of recruitment, the Utah HERO Project team sent a postcard
818 and letter to each household in the sampling frame encouraging household members to participate. Study

819 participants received testing free of charge, and each household member who completed the survey and provided a
 820 blood sample received a \$30 gift card.

821 ***Derivation of household transmission distribution F_{yz} :***

822 When an infected household member has susceptible housemates, we require the probabilities that the infected
 823 person transmits to given numbers of them. The susceptible housemates can be members of up to three different age
 824 groups: U, V, and W. We assume the probability of transmission to a susceptible group-u housemate is $1 - e^{-x}$,
 825 where x is a gamma-distributed random variable with shape k and rate r . Then the transmission probability to a
 826 susceptible group-v housemate is $1 - e^{-\sigma_v x}$, and to a susceptible group-w housemate is $1 - e^{-\sigma_w x}$, where σ_v
 827 and σ_w are constants governing the altered transmission probability to age groups V and W relative to age group U.

828 F_{yz} is the joint probability of the infected person directly transmits to $\mathbf{y} = (y_u, y_v, y_w)$ of $\mathbf{z} = (z_u, z_v, z_w)$ susceptible
 829 housemates. We show here that these assumptions lead to the formula for this probability stated in the main text.

830 We use the binomial distribution with individual success probability distributed as described above:

831
$$F_{yz} = \int_0^{\infty} B(x, y_u, z_u) B(\sigma_v x, y_v, z_v) B(\sigma_w x, y_w, z_w) \frac{r^k}{\Gamma(k)} x^{k-1} e^{-rx} dx$$

832 Where B is the probability mass function of the binomial distribution taking the following form:

833
$$B(x, y, z) = \binom{z}{y} (1 - e^{-x})^y (e^{-x})^{z-y}$$

834 The expression for B can be rewritten by converting $(1 - e^{-x})^y$ into a sum:

835
$$(1 - e^{-x})^y = \sum_{j=0}^y \binom{y}{j} (-e^{-x})^j = \sum_{j=0}^y (-1)^j \binom{y}{j} e^{-jx}$$

836 So

837
$$B(x, y, z) = \binom{z}{y} \sum_{j=0}^y (-1)^j \binom{y}{j} e^{-jx} (e^{-x})^{z-y} = \binom{z}{y} \sum_{j=0}^y (-1)^j \binom{y}{j} e^{-(z-y+j)x}$$

838 Substituting within the integral for F we get:

$$839 \quad F_{yz} = \binom{z_u}{y_u} \binom{z_v}{y_v} \binom{z_w}{y_w} \sum_{j_u=0}^{y_u} \sum_{j_v=0}^{y_v} \sum_{j_w=0}^{y_w} (-1)^{j_u+j_v+j_w} \binom{y_u}{j_u} \binom{y_v}{j_v} \binom{y_w}{j_w} \int_0^\infty \frac{r^k}{\Gamma(k)} x^{k-1} e^{-(r+z_u-y_u+j_u+\sigma_v(z_v-y_v+j_v)+\sigma_w(z_w-y_w+j_w))x} dx$$

840 The integral can be solved:

$$841 \quad F_{yz} = \binom{z_u}{y_u} \binom{z_v}{y_v} \binom{z_w}{y_w} \sum_{j_u=0}^{y_u} \sum_{j_v=0}^{y_v} \sum_{j_w=0}^{y_w} (-1)^{j_u+j_v+j_w} \binom{y_u}{j_u} \binom{y_v}{j_v} \binom{y_w}{j_w} \left(\frac{r}{r+z_u-y_u+j_u+\sigma_v(z_v-y_v+j_v)+\sigma_w(z_w-y_w+j_w)} \right)^k$$

$$842 \quad = \binom{z_u}{y_u} \binom{z_v}{y_v} \binom{z_w}{y_w} \sum_{j_u=0}^{y_u} \sum_{j_v=0}^{y_v} \sum_{j_w=0}^{y_w} (-1)^{j_u+j_v+j_w} \binom{y_u}{j_u} \binom{y_v}{j_v} \binom{y_w}{j_w} \left(1 + \frac{z_u-y_u+j_u+\sigma_v(z_v-y_v+j_v)+\sigma_w(z_w-y_w+j_w)}{r} \right)^{-k}$$

843 *Mathematical description of mechanisms of household transmission differences:*

844 In our model with exponentially distributed transmissibility, the average household secondary attack rate P_{XY} from
845 an infected person in age group X to a susceptible housemate in age group Y can be expressed:

$$846 \quad P_{XY} = 1 - (1 + \tau_X \sigma_Y \gamma_{XY})^{-1},$$

847 where τ_X is the transmissibility of a person in age group X, σ_Y is the susceptibility of a person in age group Y, and
848 γ_{XY} is the amount of contact between people in age groups X and Y when the age-group-X person is infected and the
849 age-group-Y person is susceptible to infection. More specifically, we can define τ as the average total number of
850 organisms shed over the period of infection, σ as the per-organism probability of infection given a dose of exposure
851 (as in the exponential dose-response model), and γ as the portion of total shed organisms to which the susceptible
852 contact is exposed.

853 Our MLE results provide P_{XY} for each of the nine directional attack rates among the three age-groups in our model.
854 Because there are 15 unknown values in the above formulation (3 τ_X values, 3 σ_Y values and 9 γ_{XY} values), there is
855 not a unique solution for all values given the 9 equations. We require 6 additional equation-based assumptions to
856 arrive at a unique solution. First, because we have no independent information about absolute transmissibility (count
857 of shed organisms), susceptibility (dose-response), or contact exposures in households, we first make an arbitrary
858 assumption for these values for a particular age pairing, constrained by its MLE attack rate, and write equations for
859 the ratios of each component for the other age group against the chosen index. Using U-U as the baseline, we reduce
860 to the following set of 8 equations:

861
$$\frac{\tau_V \sigma_V \gamma_{VV}}{\tau_U \sigma_U \gamma_{UU}} = \frac{Q_{VV}}{Q_{UU}}$$

862
$$\frac{\tau_W \sigma_W \gamma_{WW}}{\tau_U \sigma_U \gamma_{UU}} = \frac{Q_{WW}}{Q_{UU}}$$

863
$$\frac{\sigma_V \gamma_{UV}}{\sigma_U \gamma_{UU}} = \frac{Q_{UV}}{Q_{UU}}$$

864
$$\frac{\sigma_W \gamma_{UW}}{\sigma_U \gamma_{UU}} = \frac{Q_{UW}}{Q_{UU}}$$

865
$$\frac{\tau_V \gamma_{VU}}{\tau_U \gamma_{UU}} = \frac{Q_{VU}}{Q_{UU}}$$

866
$$\frac{\tau_V \sigma_W \gamma_{VW}}{\tau_U \sigma_U \gamma_{UU}} = \frac{Q_{VW}}{Q_{UU}}$$

867
$$\frac{\tau_W \gamma_{WU}}{\tau_U \gamma_{UU}} = \frac{Q_{WU}}{Q_{UU}}$$

868
$$\frac{\tau_W \sigma_V \gamma_{WV}}{\tau_U \sigma_U \gamma_{UU}} = \frac{Q_{WV}}{Q_{UU}}$$

869 where $Q_{XY} = (1 - P_{XY})^{-1} - 1$. In these 8 equations, there are 12 unknown ratios: two each for τ and σ values for
 870 groups V and W relative to U and 8 for γ values relative to the U-U contact assumption. Note that the unknown
 871 ratios are not dependent on the values of τ_U , σ_U , and γ_{UU} , so the assumptions for their values are arbitrary if our goal
 872 is to explore the ratios.

873 We need 4 additional assumptions to determine potential unique solutions for the 12 unknown ratios in the above
 874 equation. In the main text, we made the following sets of assumptions to illustrate a range of possibilities,
 875 corresponding to Figure 2A-D (main text):

876 A:

877
$$\frac{\tau_V}{\tau_U} = 1, \quad \frac{\tau_W}{\tau_U} = 1, \quad \frac{\sigma_V}{\sigma_U} = 1, \quad \frac{\sigma_W}{\sigma_U} = 1$$

878 B:

879
$$\frac{\tau_V}{\tau_U} = \frac{\sigma_V}{\sigma_U}, \quad \frac{\tau_W}{\tau_U} = \frac{\sigma_W}{\sigma_U}, \quad \frac{\gamma_{VV}}{\gamma_{UU}} = 1, \quad \frac{\gamma_{WW}}{\gamma_{UU}} = 1$$

880 C:

881
$$\frac{\tau_V}{\tau_U} = 1, \quad \frac{\tau_W}{\tau_U} = 1, \quad \frac{\gamma_{VV}}{\gamma_{UU}} = 1, \quad \frac{\gamma_{WW}}{\gamma_{UU}} = 1$$

882 D:

883
$$\frac{\sigma_V}{\sigma_U} = 1, \quad \frac{\sigma_W}{\sigma_U} = 1, \quad \frac{\gamma_{VV}}{\gamma_{UU}} = 1, \quad \frac{\gamma_{WW}}{\gamma_{UU}} = 1$$

884

885 **Table A. Household recruitment and number of individuals tested by phase and location.**

Phase / Location	Households contacted	Households tested (response %)	Individuals tested	Individuals per household
School district: Granite	39,131	1,123 (2.9%)	2,539	2.26
School district: Salt Lake City	35,936	1,677 (4.7%)	3,262	1.94
Hotspot: Salt Lake County	10,601	290 (2.7%)	907	3.12
Hotspot: Utah County	10,879	291 (2.7%)	999	3.43
Total	96,547	3,381 (3.5%)	7,707	2.28

886

887 **Table B. Positivity rates by age and vaccination status**

MeSH age category	Prior positive / total participants	Abbott positive / total determinate	Euroimmun positive / total determinate	Siemens positive / total determinate
All ages	Unvax: 349 / 4518 (7.7%) Vax1: 37 / 610 (6.1%) Vax2: 41 / 727 (5.6%)	Unvax: 328 / 3436 (9.5%) Vax1: 37 / 598 (6.2%) Vax2: 40 / 712 (5.6%)	Unvax: 505 / 3587 (14.2%) Vax1: 378 / 513 (73.7%) Vax2: 664 / 672 (98.8%)	Unvax: 392 / 3434 (11.4%) Vax1: 373 / 597 (62.5%) Vax2: 693 / 713 (97.2%)
Newborn/ Infant/ Preschool (0 – 5)	Unvax: 7 / 350 (2.0%)	Unvax: 0 / 2 (0%)	Unvax: 32 / 312 (10.3%)	Unvax: 0 / 1 (0%)
Child (6 – 12)	Unvax: 21 / 633 (3.3%)	Unvax: 0 / 8 (0%)	Unvax: 55 / 573 (9.6%)	Unvax: 0 / 8 (0%)
Adolescent (13 – 18)	Unvax: 47 / 394 (11.6%) Vax1: 0 / 8 (0%) Vax2: 0 / 3 (0%)	Unvax: 57 / 369 (15.4%) Vax1: 0 / 8 (0%) Vax2: 0 / 3 (0%)	Unvax: 76 / 322 (23.6%) Vax1: 4 / 7 (57.1%) Vax2: 3 / 3 (100%)	Unvax: 72 / 369 (19.5%) Vax1: 5 / 8 (62.5%) Vax2: 3 / 3 (100%)
Young adult (19 – 24)	Unvax: 26 / 247 (10.5%) Vax1: 3 / 14 (21.4%) Vax2: 2 / 27 (7.4%)	Unvax: 26 / 233 (11.1%) Vax1: 3 / 13 (23.1%) Vax2: 2 / 27 (7.4%)	Unvax: 39 / 189 (20.6%) Vax1: 9 / 12 (75.0%) Vax2: 25 / 25 (100%)	Unvax: 33 / 233 (14.2%) Vax1: 10 / 13 (76.9%) Vax2: 27 / 27 (100%)
Adult (25 – 44)	Unvax: 115 / 1552 (7.4%) Vax1: 12 / 178 (6.7%) Vax2: 17 / 242 (7.0%)	Unvax: 120 / 1510 (7.9%) Vax1: 11 / 176 (6.3%) Vax2: 15 / 236 (6.4%)	Unvax: 157 / 1191 (13.2%) Vax1: 116 / 153 (75.8%) Vax2: 214 / 215 (99.5%)	Unvax: 141 / 1510 (9.3%) Vax1: 121 / 176 (68.8%) Vax2: 235 / 236 (99.6%)
Middle aged (45 – 64)	Unvax: 98 / 931 (10.5%) Vax1: 10 / 236 (4.2%) Vax2: 13 / 164 (7.9%)	Unvax: 86 / 910 (9.5%) Vax1: 12 / 232 (5.2%) Vax2: 16 / 160 (10.0%)	Unvax: 107 / 710 (15.1%) Vax1: 143 / 209 (68.4%) Vax2: 140 / 145 (96.6%)	Unvax: 100 / 909 (11.0%) Vax1: 132 / 231 (57.1%) Vax2: 152 / 160 (95.0%)
Aged (65 – 79)	Unvax: 34 / 381 (8.9%) Vax1: 11 / 160 (6.9%) Vax2: 9 / 251 (3.6%)	Unvax: 39 / 375 (10.4%) Vax1: 10 / 155 (6.5%) Vax2: 7 / 247 (2.8%)	Unvax: 41 / 273 (15.0%) Vax1: 99 / 123 (80.5%) Vax2: 243 / 245 (99.2%)	Unvax: 45 / 375 (12.0%) Vax1: 99 / 155 (63.9%) Vax2: 240 / 248 (96.8%)
Aged (80 and over)	Unvax: 1 / 30 (3.3%) Vax1: 1 / 14 (7.1%) Vax2: 0 / 40 (0%)	Unvax: 0 / 29 (0%) Vax1: 1 / 14 (7.1%) Vax2: 0 / 39 (0%)	Unvax: 1 / 17 (5.9%) Vax1: 7 / 9 (77.8%) Vax2: 39 / 39 (100%)	Unvax: 1 / 29 (3.4%) Vax1: 6 / 14 (42.9%) Vax2: 36 / 39 (92.3%)

888 MeSH = Medical Subject Headings [28]; Unvax = unvaccinated (participants who reported receiving no prior vaccine doses); Vax1 = participants

889 who reported receiving exactly 1 vaccine dose; Vax2 = participants who reported receiving 2 vaccine doses; total determinate = total number of

890 individuals receiving a test result of positive or negative; prior positive = individuals who reported on the survey that they had tested positive for

891 COVID-19

892 **Table C. Frequency of test and survey result combinations among individuals in the data**

Antibody Test results Abbot-Euroimmun-Siemens	Prior positive test reported? (No;Yes)
>0 negative, 0 positive	
NNN	2384; 28
NNX	1; 0
NXN	751; 11
NXX	972; 19
XNX	790; 3
>0 negative, >0 positive	
NPP	55; 44
NPN	53; 10
PPN	14; 14
PNN	10; 2
PXN	6; 3
NXP	3; 5
NNP	4; 0
NPX	1; 0
0 negative; >0 positive	
PPP	70; 160
XPX	66; 21
PXX	17; 58
PXP	22; 29
0 negative; 0 positive: XXX	
	209; 20

893 P=positive, N=negative, X=indeterminate. Indeterminate means the test result was indeterminate or, for Euroimmun and Siemens, the rest result

894 was positive for a participant reporting any prior vaccination.

895 **Table D. Transmission variability model comparison**

	\hat{p}_a	\hat{p}_{ww2}	\hat{p}_{wv2}	\hat{p}_{wu2}	σ_{34}	\hat{k}_w	$\hat{\phi}_{Sa}$	$\hat{\pi}_S$	Log likelihood	P value	Par	ΔAIC
	\vdots	\hat{p}_{vw2}	\hat{p}_{vv2}	\hat{p}_{vu2}	σ_{5+}	\hat{k}_v	\vdots	$\hat{\pi}_{AA}$				
	\hat{p}_h	\hat{p}_{uw2}	\hat{p}_{uv2}	\hat{p}_{uu2}		\hat{k}_u	$\hat{\phi}_{Sh}$	$\hat{\pi}_{AE}$				
							$\hat{\phi}_{AS}$	$\hat{\pi}_{AS}$				
Gamma-distributed transmissibility	3.2%	47%	68%	42%	0.42	0.75	16%	99.1%	-3020.8	Ref	37	Ref
	2.0%	48%	44%	44%	0.22	1.00	35%	72%				
	15.0%	34%	51%	68%		1.82	53%	99.2%				
	12.6%						62%	87%				
	6.6%						59%					
	6.7%						73%					
	5.2%						73%					
	0.0%						100%					
Exponentially distributed transmissibility	3.2%	45%	63%	42%	0.43	1	16%	99.2%	-3021.0	0.95	34	-5.7
	2.0%	47%	44%	45%	0.22	1	36%	72%				
	15.1%	34%	52%	68%		1	53%	99.2%				
	12.6%						62%	87%				
	6.7%						59%					
	6.7%						73%					
	5.2%						73%					
	0.0%						100%					
No-variance transmissibility	2.7%	46%	69%	37%	0.51	∞	17%	99.1%	-3022.7	0.28	34	-2.2
	1.9%	45%	42%	43%	0.27	∞	36%	72%				
	14.9%	35%	52%	69%		∞	53%	99.2%				
	12.5%						62%	87%				
	6.6%						58%					
	6.8%						73%					
	5.2%						73%					
	0.0%						100%					

896 P values are for rejection of the model in favor of the full model, by the likelihood ratio test with degrees of freedom equal to the difference in
 897 number of parameters (Par). Par = number of independent parameters simultaneously estimated. ΔAIC = difference in Akaike information
 898 criterion value versus the reference (Ref) model; a negative ΔAIC means the model is favored over the reference model, and the lowest ΔAIC
 899 model is favored over all the others.

900 **Table E. Maximum likelihood results for different transmission age groupings**

Rank	Group W age range	Group V age range	Group U age range	MLE likelihood
1	0 – 24	25 – 44	45+	-3021.0
2	0 – 12	13 – 44	45+	-3023.6
3	0 – 18	19 – 44	45+	-3022.2
4	0 – 44	45 – 64	65+	-3023.2
5	0 – 12	13 – 64	65+	-3025.3
6	0 – 18	19 – 24	25+	-3026.0
7	0 – 24	25 – 65	65+	-3026.1
8	0 – 12	13 – 24	25+	-3026.1
9	0 – 18	19 – 64	65+	-3026.4
10	0 – 12	13 – 18	19+	-3027.4

903 MLE = maximum likelihood estimate. The MLE for each row was calculated by optimizing over all test accuracy, importation, and transmission
904 parameters in the full model (fixing $k_u = k_v = k_w = 1$) after categorizing household members into U, V, and W transmission groups according
905 to the age ranges listed in that row. The higher-ranking age configurations produced a greater (log) likelihood at the MLE.

906

907 **Table F. Comparison of simpler age- and size-structured transmission models to full model**

	\hat{p}_{ww2}	\hat{p}_{vv2}	\hat{p}_{wu2}	σ_{34}	Log likelihood	P value	Par	Δ AIC
	\hat{p}_{vw2}	\hat{p}_{vp2}	\hat{p}_{vu2}	σ_{5+}				
	\hat{p}_{uw2}	\hat{p}_{uv2}	\hat{p}_{uu2}					
Full model	45%	63%	42%	0.43	-3020.98	Ref	11	Ref
	47%	44%	45%	0.22				
	34%	52%	68%					
Equal HSAR from U	50%	71%	49%	0.32	-3025.25	0.01	9	4.5
	56%	46%	49%	0.16				
	62%	62%	62%					
Equal HSAR from V	45%	62%	42%	0.44	-3021.03	0.95	9	-3.9
	45%	45%	45%	0.21				
	32%	51%	67%					
Equal HSAR from W	47%	47%	47%	0.48	-3023.33	0.10	9	0.7
	46%	46%	38%	0.24				
	28%	52%	66%					
Equal HSAR to U	53%	69%	59%	0.35	-3023.98	0.05	9	2.0
	51%	45%	59%	0.17				
	30%	52%	59%					
Equal HSAR to V	38%	48%	34%	0.56	-3021.83	0.43	9	-2.3
	42%	48%	36%	0.30				
	29%	48%	65%					
Equal HSAR to W	39%	57%	32%	0.53	-3021.36	0.68	9	-3.2
	39%	42%	36%	0.27				
	39%	51%	65%					
Equal HSAR to and from V	37%	47%	34%	0.51	-3022.17	0.88	7	-5.6
	47%	47%	47%	0.28				
	33%	47%	66%					
Consolidate W and V	47%	47%	39%	0.48	-3023.64	0.37	6	-4.7
	47%	47%	39%	0.24				
	37%	37%	67%					
Consolidate V and U	47%	54%	54%	0.41	-3027.18	0.03	6	2.4
	43%	56%	56%	0.21				
	43%	56%	56%					
Consolidate W and V; reciprocal transmission	48%	48%	40%	0.46	-3023.69	0.49	5	-6.6
	48%	48%	40%	0.22				
	40%	40%	67%					
Equal HSAR to and from V; equal HSAR W-to-W, U-to-W, & W-to-U	34%	47%	34%	0.51	-3022.26	0.97	5	-9.4
	47%	47%	47%	0.29				
	34%	47%	66%					

908 HSAR values (p estimates in a 3-by-3 grid) that share bold and/or italics font type were assumed to be equal for that model. P values are for
 909 rejection of the model in favor of the full model, by the likelihood ratio test with degrees of freedom equal to the difference in number of
 910 parameters (Par). Par = number of independent household transmission parameters simultaneously estimated. Δ AIC = difference in Akaike
 911 information criterion value versus the reference (Ref) model; a negative Δ AIC means the model is favored over the reference model, and the
 912 lowest Δ AIC model is favored over all the others.

913

914 **Table G. Comparison of simpler age- and size-structured transmission models to most parsimonious model**

	\hat{p}_{ww2}	\hat{p}_{wv2}	\hat{p}_{wu2}	σ_{34}	Log likelihood	P value	Par	Δ AIC
	\hat{p}_{vw2}	\hat{p}_{vp2}	\hat{p}_{vu2}	σ_{5+}				
	<i>\hat{p}_{uw2}</i>	<i>\hat{p}_{uv2}</i>	<i>\hat{p}_{uu2}</i>					
Most parsimonious model	<i>34%</i>	47%	<i>34%</i>	0.51	-3022.26	Ref	5	Ref
	47%	47%	47%	0.29				
	<i>34%</i>	47%	66%					
Equal HSAR except U-to-U	44%	44%	44%	0.47	-3024.70	0.03	4	2.9
	44%	44%	44%	0.24				
	44%	44%	66%					
Equal HSAR for all ages	54%	54%	54%	0.36	-3029.35	0.0008	3	10.2
	54%	54%	54%	0.17				
	54%	54%	54%					
No household size effect	<i>16%</i>	30%	<i>16%</i>	1	-3031.11	0.0001	3	13.7
	30%	30%	30%	1				
	<i>16%</i>	30%	57%					
Equal HSAR for all ages and sizes	27%	27%	27%	1	-3053.50	<0.0001	1	54.5
	27%	27%	27%	1				
	27%	27%	27%					

915 HSAR values (p estimates in a 3-by-3 grid) that share bold or italics font were assumed to be equal for that model. The “no household size effect”
 916 assumption fixes $\sigma_{34} = \sigma_{5+} = 1$. P values are for rejection of the model in favor of the most parsimonious model, by the likelihood ratio test with
 917 degrees of freedom equal to the difference in number of parameters (Par). Par = number of independent household transmission parameters
 918 simultaneously estimated. Δ AIC = difference in Akaike information criterion value versus the reference (Ref) model; positive Δ AIC means the
 919 model is not favored over the reference model.

920

921 **Table H. Consolidated age groupings for community acquisition and case ascertainment**

	\hat{p}_a	\hat{p}_{ww2}	\hat{p}_{wv2}	\hat{p}_{wu2}	σ_{34}	$\hat{\phi}_{Sa}$	$\hat{\pi}_S$	Log likelihood	P value	Par	Δ AIC	
	\hat{p}_h	\hat{p}_{vw2}	\hat{p}_{vv2}	\hat{p}_{vu2}	σ_{5+}	$\hat{\phi}_{Sh}$	$\hat{\phi}_{AA}$	$\hat{\pi}_{AA}$				
		\hat{p}_{uw2}	\hat{p}_{uv2}	\hat{p}_{uu2}			$\hat{\phi}_{AE}$	$\hat{\pi}_{AE}$				
							$\hat{\phi}_{AS}$	$\hat{\pi}_{AS}$				
Full model	3.2%	45%	63%	42%	0.43	16%		99.2%	-3020.98	Ref	34	Ref
	2.0%	47%	44%	45%	0.22	36%	72%	99.7%				
	15.1%	34%	52%	68%		53%	99.2%	98.1%				
	12.6%					62%	87%	99.85%				
	6.7%					59%						
	6.7%					73%						
	5.2%					73%						
	0.0%					100%						
Full transmission model;	2.5%	45%	63%	43%	0.44	29%		99.2%	-3025.37	0.46	25	-9.2
consolidated age groups for p_i and ϕ_{Si}	2.5%	46%	44%	44%	0.22	29%	72%	99.7%				
	14.0%	33%	54%	67%		57%	99.2%	98.1%				
	14.0%					57%	87%	99.85%				
	6.4%					57%						
	6.4%					73%						
	6.4%					73%						
	0.0%					73%						
Parsimonious transmission model;	2.1%	35%	47%	35%	0.51	28%		99.2%	-3026.80	0.71	19	-18.4
consolidated age groups for p_i and ϕ_{Si}	2.1%	47%	47%	47%	0.29	28%	72%	99.7%				
	13.7%	35%	47%	66%		57%	99.3%	98.1%				
	13.7%					57%	87%	99.85%				
	6.6%					57%						
	6.6%					73%						
	6.6%					73%						
	0.0%					73%						

922 The “consolidated age groups for p_i and ϕ_{Si} ” versions enforce the following assumptions: $p_a = p_b$, $p_c = p_d$, $p_e = p_f = p_g$, $\phi_{Sa} = \phi_{Sb}$, $\phi_{Sc} =$
923 $\phi_{Sd} = \phi_{Se}$, and $\phi_{Sf} = \phi_{Sg} = \phi_{Sh}$. The parsimonious transmission model enforces $p_{ww} = p_{wu} = p_{uw}$ and $p_{vu} = p_{vv} = p_{vw} = p_{uv} = p_{wv}$ within
924 each household size group. All models in this table assume exponentially distributed variability in transmissibility ($k_u = k_v = k_w = 1$). P values
925 are for rejection of the model in favor of the full model, by the likelihood ratio test with degrees of freedom equal to the difference in number of
926 parameters (Par). Par = number of independent parameters simultaneously estimated. Δ AIC = difference in Akaike information criterion value
927 versus the reference (Ref) model; a negative Δ AIC means the model is favored over the reference model, and the lowest Δ AIC model is favored
928 over all the others.