

21 **SUMMARY**

22 SARS-CoV-2 superspreading occurs when transmission is highly efficient and/or an individual
23 infects many others, contributing to rapid spread. To better quantify heterogeneity in SARS-
24 CoV-2 transmission, particularly superspreading, we performed a systematic review of
25 transmission events with data on secondary attack rates or contact tracing of individual index
26 cases published before September 2021, prior to emergence of variants of concern and
27 widespread vaccination. We reviewed 592 distinct events and 9,883 index cases from 491
28 papers. Meta-analysis of secondary attack rates identified substantial heterogeneity across 12
29 event types/settings, with the highest transmission (25–35%) in co-living situations including
30 households, nursing homes, and other congregate housing. Among index cases, 67% produced
31 zero secondary cases and only 3% (287) infected >5 secondary cases (“superspreaders”). The
32 highest percentage of superspreaders was among symptomatic individuals, individuals aged 49–
33 64 years, and individuals with over 100 total contacts. However, only 55% of index cases
34 reported age, sex, symptoms, real-time PCR cycle threshold values, or total contacts. Despite the
35 limitations, our review highlighted that SARS-CoV-2 superspreading is more likely in settings
36 with prolonged close contact and among symptomatic adults with many contacts. Enhanced
37 reporting on transmission events and contact tracing could help explain heterogeneity and
38 facilitate control efforts.

39

40 **KEYWORDS**

41 coronavirus; COVID-19; transmission; heterogeneity; infectious disease epidemiology

42

43 INTRODUCTION

44 Following the emergence of SARS-CoV-2 in late 2019, the virus spread worldwide,
45 resulting in the coronavirus disease (COVID-19) pandemic [1]. Understanding drivers of SARS-
46 CoV-2 transmission was crucial for formulating control measures, especially prior to the
47 development of vaccines. Early in the pandemic, heterogeneity in transmission, particularly
48 superspreading, was investigated because of its ability to cause large outbreaks [2–4].
49 Superspreading involves two distinct but non-mutually exclusive phenomena: a setting where
50 many people become infected due to an environment conducive to transmission (e.g., crowded
51 indoor settings), and individuals who are outliers in the number of secondary cases they infect,
52 due to biological heterogeneity in infectiousness and/or high-risk behaviors [5,6]. Both
53 phenomena have garnered considerable attention in the literature. For example, over 140
54 individuals were infected during a Christmas event in Belgium in December 2020, causing over
55 26 deaths [7]. Likewise, one individual infected dozens of people during a choir practice in
56 Washington, USA, in March 2020 [8].

57 Because superspreading events contributed substantially to local and global SARS-CoV-
58 2 transmission [9], public health interventions were enacted to reduce their risk of occurrence.
59 These interventions included school closures, limitations on indoor gatherings, and restrictions
60 on visiting hospitalized patients or long-term care facilities. Many of these policies were based
61 on limited data from early in the pandemic. Moreover, published systematic reviews and
62 modeling of SARS-CoV-2 superspreading from this period were limited in scope and did not
63 adequately address the dual nature of superspreading. For example, studies of setting-specific
64 transmission rates have focused on household and healthcare transmission or geographic and
65 temporal trends [2,10–13], but did not address transmission heterogeneity across other social

66 settings. Previous meta-analyses of individual-level superspreading included only a small
67 number of papers (<26) that calculated overdispersion in transmission, missing the majority of
68 published transmission trees and capturing data primarily from Asia [14,15]. Early investigations
69 of individual-level characteristics related to superspreading were also limited by incomplete
70 contact tracing [16,17] and a focus on clinical over demographic characteristics [16]. A more
71 complete summary of superspreading is needed to understand the scale of transmission
72 heterogeneity across settings and identify causes of individual heterogeneity.

73 The objective of this review was to summarize global heterogeneity in SARS-CoV-2
74 transmission events prior to widespread vaccination and the role of environmental and individual
75 factors in superspreading. Specifically, this review aimed to identify: 1) which settings had the
76 highest attack rates; 2) the individual offspring distribution for SARS-CoV-2; and 3) the
77 characteristics of superspreading individuals.

78

79 **METHODS**

80 **Literature search and data extraction**

81 We conducted this systematic review and meta-analysis according to the Preferred
82 Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [18]; see
83 Appendix 1 for the PRISMA checklist. We included all studies of SARS-CoV-2 that contained
84 data on: 1) transmission chains; 2) numbers of index cases, contacts, and infected contacts; 3)
85 numbers of index cases and infected contacts; or 4) secondary attack rates, i.e., number of
86 infected contacts divided by number of contacts. We excluded studies that were not about
87 humans. A clinical informationist searched PubMed, the WHO COVID database, the I Love
88 Evidence COVID database, and Embase on 9 September 2021. No restrictions on language or

89 start date were applied. Results were imported into EndNote X9 (Clarivate, London, UK) where
90 duplicates with exact matches in the author, year, and title fields were removed. Team members
91 screened titles and abstracts and performed full text review in Covidence (Veritas Health
92 Innovation, Melbourne, Australia).

93 We extracted data using a pre-designed, study-specific spreadsheet, collecting
94 information on paper metadata and target variables for transmission events and individual index
95 cases (Table 1). Events were defined as discrete transmission events where secondary attack
96 rates for defined groups of people could be calculated as the number of infected cases divided by
97 the total number of exposed individuals. This definition of secondary attack rates includes both
98 clinical and subclinical infections in some studies. Due to the limited details published in the
99 literature, we did not attempt to distinguish events associated with individual transmission chains
100 from a single source (potentially with confirmatory sequencing data) from events that aggregated
101 multiple transmission chains together. In lieu of this distinction, we separated events into
102 different settings and by the duration of the event (i.e., exposure window, in days) reported in
103 each paper. Twelve event types were chosen to classify each event/setting described in a paper
104 (Table 2). To describe individual contributions to transmission, we extracted data on index cases
105 for whom contacts were followed to identify secondary transmission. We only entered data from
106 papers where it was clear from the methods that contact tracing was done for at least one week to
107 capture secondary transmission from individual index cases. For studies that did not report
108 SARS-CoV-2 variants, we imputed the dominant variant from CoVariants data for the country
109 and time period of interest [19]. See the Supplementary Material for additional details on the
110 identification of papers, data extraction (Supplementary Tables S2–S3), and assessment of study
111 bias.

112

113 **Statistical analyses**

114 Descriptive analysis of event data included the number of each event type, starting year
115 of the data, focal countries, diagnostic methods, event duration, and level of missingness for all
116 variables. Because not all individuals potentially exposed during an event were tested in each
117 study, secondary attack rates for individual events were calculated separately using the total
118 number of exposed individuals or the total number tested. If either of these quantities were
119 missing, the value was imputed based on the value present (i.e., assuming the number tested was
120 equal to the number exposed or vice versa). Sensitivity of results to this choice of denominator
121 was assessed in the meta-analysis of events (see Supplementary Material).

122 Meta-analysis was performed on secondary attack rates across event types using the
123 *metafor* package in R v4.2.2 [20]. We converted secondary attack rates for each event to
124 Freeman-Tukey double arcsine transformed proportions [21] and calculated the sampling
125 variance. We fit a hierarchical model with a nested random effect for event within study and no
126 fixed effects to assess the heterogeneity in secondary attack rates attributable to these factors
127 using restricted maximum likelihood. We calculated I^2 , the percentage of variance attributable to
128 true heterogeneity, for each random effect [22] and used Cochran's Q test to test if estimated
129 heterogeneity in secondary attack rates was greater than expected from the sampling error alone.
130 We then fit additional mixed-effects models that included the same random effects but also event
131 type and event duration as fixed effects. Cochran's Q was performed on these model to assess
132 whether residual heterogeneity in secondary attack rates was greater than expected after
133 accounting for sampling error and fixed effects. Fitted coefficients and 95% confidence intervals
134 (CI) from meta-analysis were back-transformed to proportions using the geometric mean of the

135 tested individuals across all studies in each event type [21].

136 The overall distribution of secondary cases generated by index cases was fit to a negative
137 binomial distribution, following Lloyd-Smith et al. [23]. We estimated the percentile of index
138 cases producing 80% of all secondary infections using a formula and code from Endo et al. [24].
139 Based on the availability of demographic characteristics and other features of index cases in the
140 literature, we examined differences in distributions of secondary cases produced by index cases
141 according to sex, presence/absence of symptoms, age, real-time PCR cycle threshold (Ct) value,
142 and the total number of contacts each index case had. Additional statistical tests compared these
143 listed factors between “superspreaders” (index cases with >5 secondary cases, following Adam
144 et al. [3]) and “non-superspreaders” (index cases with ≤5 secondary cases): Chi-square tests of
145 proportions to compare the proportion of women, the proportion of symptomatic cases, and
146 proportion of adults or across age bins; Student’s t-tests to compare mean age and Ct value; and
147 a Kruskal-Wallis test to compare the highly skewed distributions of total contacts among index
148 cases. All statistical tests used $\alpha = 0.05$ as the statistical significance threshold.

149

150 **RESULTS**

151 **Study selection**

152 We identified 13,632 articles from the four databases searched, representing 8,339 unique
153 references (Figure 1). Of these, we excluded 7,358 records during the abstract review. For the
154 981 records that underwent full text review, we excluded 384 records that were reviews or letters
155 to the editor without data, contained no data on our variables of interest, or were duplicate
156 records (preprints, true duplicates, or duplicated datasets). A total of 598 papers were assessed
157 for eligibility for data extraction and a further 107 papers were excluded that did not contain data

158 on our outcome variables of interest or were duplicates (Figure 1). We extracted data from 491
159 studies: 232 studies provided event data only, 195 studies provided individual index case data
160 only, and 64 studies provided both data types, yielding evidence from 592 distinct events and
161 9,883 index cases. The 491 analyzed studies were from 67 countries, with most from China
162 (26%), the USA (17%), and South Korea (5%) (Supplementary Figure S1A). Although our
163 search included two-thirds of 2021, nearly all studies covered data from 2020 (94% of events,
164 99% of index case symptom onset or positive test dates).

165

166 **Heterogeneity in event secondary attack rates**

167 Event data were most commonly from the USA (27%), China (15%), the UK (8%), and
168 South Korea (6%) (Supplementary Figure S1B). Published papers were missing information on
169 many variables that we aimed to extract about events (Supplementary Figure S2A). Of the 46
170 target data fields from articles about events, 17 had high data completeness (>80%), including
171 those for the study and event metadata, event description, time period of the event (describing
172 the start and end dates of exposure), location of the event (country and state/province or city),
173 and number of exposed individuals and secondary cases (Supplementary Table S3). Event
174 durations were highly skewed, with a median duration of 34 days and an interquartile range of
175 13–60 days (Supplementary Figure S3). Studies used a variety of diagnostic methods to identify
176 SARS-CoV-2 cases, though PCR was the dominant method (Supplementary Figure S4A). Other
177 approaches included antigen tests, retrospective case identification by serology, diagnosis via
178 symptoms or chest tomography in early papers, or a mixture of approaches. Because most studies
179 covered events prior to emergence of variants, most events (N = 532, 90%) likely involved only
180 wild-type/ancestral SARS-CoV-2, while 14 events involved Alpha, six Beta, eight Delta, and 31

181 likely included a mixture of variants (e.g., during periods of variant emergence and replacement
182 of the dominant variant).

183 Secondary attack rates varied substantially within and among event types (Figure 2).
184 Interquartile ranges of attack rates were lower for transport (0–11%), hospital/healthcare (1–
185 20%), and mixed events (3–12%), whereas congregate housing (9–63%), households (15–60%),
186 social venues (8–53%), and cruise ships (9–41%) had higher heterogeneity, with some events
187 reporting attack rates of 100% (Table 2). Meta-analysis of secondary attack rates including a
188 nested random effect for event within study detected significant heterogeneity in secondary
189 attack rates ($I^2 = 99%$, Cochran's $Q_{E,591} = 141,765$, $P < 0.0001$). The random effect for study
190 accounted for most of the heterogeneity ($I^2_{study} = 58%$), followed by event nested within study
191 ($I^2_{event} = 41%$). Addition of a fixed effect for event type to the model indicated that secondary
192 attack rates varied significantly across event types (Cochran's $Q_{M,11} = 122$, $P < 0.0001$). Mean
193 secondary attack rates from meta-analysis were lowest for shopping (0%), hospitals and
194 healthcare (6%), transportation other than cruise ships (9%), and schools (11%) (Figure 2).
195 Comparatively, estimated mean attack rates were two to three times higher (25–35%) in nursing
196 homes, cruise ships, households, and other congregate housing settings (e.g., homeless shelters,
197 prisons). Models including event duration and an interaction term between event type and event
198 duration as additional fixed effects found similar levels of heterogeneity (Cochran's $Q_{M,23} = 135$,
199 $P < 0.0001$) and identified a common trend of decreasing attack rates with longer event durations
200 across different event types, with the exception of cruise ships and shopping (Supplementary
201 Figure S5).

202

203 **Heterogeneity in transmission across individual index cases**

204 Individual index case data with offspring distributions overwhelmingly came from China
205 (36%) and India (35%) (Supplementary Figure S1C). Index case data exhibited higher
206 missingness compared to events (Supplementary Figure S2B): of the 74 data fields that we
207 extracted for individual index cases, the highest completeness (>60%) was seen for study and
208 index case numbers, location of the index case (country and state/province or city), total number
209 of contacts infected, method of testing for the index case and contacts, and SARS-CoV-2 variant
210 (Supplementary Table S4). We identified five key characteristics of index cases that could be
211 related to superspreading, though most of these were also missing from the published literature:
212 46% of cases included data on age, 48% on sex, 10% on presence/absence of symptoms, 6% on
213 total number of contacts, and only 2% had Ct values reported. A total of 5,437 index cases (55%)
214 contained data on at least one of these five variables. Diagnostic methods for identification of
215 individual index cases and their associated secondary cases were only reported in 61% of cases,
216 with PCR as the primary approach (Supplementary Figure S4B,C). The majority of index cases
217 (N = 8,565, 87%) were assumed to be infected with wild-type SARS-CoV-2 based on location
218 and timing of the study or test confirmation date. A mixture of variants was likely in 1,282 cases
219 (13%), while one index case was reported with Alpha, two Beta, 11 Delta, and 22 Epsilon.

220 Most index cases (67%) did not transmit SARS-CoV-2 to another person and 17%
221 transmitted to only one other individual (Figure 3). There were 287 “superspreaders” with >5
222 contacts infected, representing 3% of index cases from the included studies. The distribution of
223 secondary infections fit a negative binomial distribution with a mean of 0.88 (CI: 0.84–0.92) and
224 a dispersion parameter k of 0.27 (CI: 0.25–0.28). Using the formula from Endo et al. [24] and the
225 estimated mean and k for the negative binomial distribution, the top 17% most infectious index

226 cases would be expected to generate 80% of all secondary cases.

227 We observed substantial differences between superspreaders and non-superspreaders
228 (Table 3). The proportion of index cases with reported symptoms was higher in superspreaders
229 (89%) than non-superspreaders (76%; $\chi^2_1 = 5.4$, $P = 0.02$). Superspreaders had more than two
230 times the mean number of contacts (79) compared to non-superspreaders (36; $\chi^2_1 = 56.6$, $P <$
231 0.0001). Adults also made up a greater proportion of superspreaders (99%) than non-
232 superspreaders (84%; $\chi^2_1 = 14.1$, $P < 0.0001$). Index cases over 25 years of age were
233 overrepresented among superspreaders and no superspreaders 12 years of age and under were
234 reported (Figure 4). When age was analyzed as a continuous variable, the number of contacts
235 infected and the frequency of superspreaders increased with age, up to around 60 years of age
236 (Supplementary Figure S6). No significant differences by sex or Ct values were observed (Table
237 3). However, two adult male index cases produced the highest number of secondary infections,
238 infecting 81 of their 104 contacts and 101 of their 300 contacts, respectively. The former was a
239 lecturer in Tonghua, China [25] and the latter a fitness instructor in Hong Kong, China [26].

240 Symptomatic cases had a higher mean number of infected contacts (2.1) compared to
241 asymptomatic cases (0.7) (Table 4). The dispersion parameter k was higher for symptomatic
242 cases than asymptomatic cases (0.43 vs. 0.11), indicating lower variance in the number of
243 secondary cases produced by a symptomatic case. This variance is exemplified by the lower
244 percentage of non-transmitters (44%) and higher percentage of superspreaders (9%) among
245 symptomatic cases compared to asymptomatic cases (79% and 4%, respectively). Compared to
246 other age groups, individuals aged 49–64 years had the highest mean number of infected contacts
247 (1.2), lower variance (higher k , 0.43), and a higher percentage of superspreaders (3%). Data on
248 total reported contacts showed a different pattern, with a higher mean number of infected

249 contacts (8) as well as higher variance (lower k , 0.28) among index cases with >100 total
250 contacts compared to individuals with fewer contacts. This was accompanied by a substantially
251 higher percentage of superspreaders (28%) among individuals with >100 total contacts compared
252 to individuals with 11–100 contacts (19%) or those with 0–10 contacts (2%). Considering only
253 symptomatic adults with a known number of total contacts ($N = 129$), the percentage of
254 superspreaders was consistently smaller as the number of contacts decreased: 26% (5/19) for
255 individuals with over 100 contacts, 24% (8/34) for those with 21–100 contacts, 8% (2/24) for
256 those with 11–20 contacts, and 0% for those with 10 or fewer contacts (0/52).

257

258 **DISCUSSION**

259 Superspreading is a key form of heterogeneity in transmission of numerous viruses,
260 including SARS-CoV, MERS-CoV, SARS-CoV-2, Nipah, Ebola, and measles [14,15,23,27–29].
261 The potential for superspreading depends on the environment and social context of contact, as
262 well as individual biological or behavioral factors that influence infectiousness. The existing
263 literature on SARS-CoV-2 superspreading has done little to disaggregate this phenomenon into
264 the distinct contributions of environment and individual characteristics.

265 Our ranking of event types by attack rate reinforces our existing understanding of SARS-
266 CoV-2 and other pathogens that transmit through the air, that transmission is more likely in
267 dense indoor gatherings or close and frequent contact among co-living individuals, especially in
268 households [9]. Published meta-analyses covering the early pandemic (pre-2021) estimated
269 pooled household secondary attack rates of 17–21% [10,12,13,30,31], with household attack
270 rates consistently higher than those in healthcare, work, or travel settings [10,13]. Our pooled
271 household secondary attack rate over 115 events was 29%, higher than these earlier studies but

272 similar to the 31% estimate from Madewell et al. [12] for studies covering July 2020 to March
273 2021. The higher value may be explained by the emergence of the Alpha and Delta variants and
274 the larger second and third waves of the pandemic occurring in some countries during 2021.

275 The literature on SARS-CoV-2 transmission events rarely reported on the
276 epidemiological context and characteristics of different populations exposed, which could help
277 explain variation in attack rates. While the timing and location of events may help to explain
278 some of the variation within event types, the remaining variation could depend on event duration
279 (as shown by Supplementary Figure S5) and time spent indoors, types of activities occurring
280 (e.g., exercise, singing) [32,33], varying interventions in place (e.g. masking requirements,
281 physical distancing), and the age groups present at the event. For example, the age of individuals
282 interacting in these contexts appears to also influence propensity for transmission, as evidenced
283 by the large difference in attack rates within schools versus nursing homes. Children and
284 adolescents are frequently found to have lower household infection risk than working age adults
285 [12,13,17,31] and older adults have higher risk of infection and severe disease than younger ages
286 [12,31]. In studies that assessed transmission among school-aged children, teachers, and their
287 household contacts, attack rates among children at school were lower than among teachers and
288 the household contacts of children and teachers [34,35]. Environmental factors such as humidity,
289 room size, ventilation, and air flow [5] could also augment transmission across settings but these
290 were very rarely reported in the literature.

291 Our results also indicate substantial heterogeneity in the infectiousness of individuals,
292 which is reflected elsewhere [36–38]. Analysis of index case demographics also highlights age as
293 an important factor in SARS-CoV-2 transmission and superspreading. While age was only
294 reported in 46% of index cases, nearly all superspreading individuals were adults and there were

295 no reported superspreaders 12 years of age and under. Presence of symptoms was reported less
296 frequently in the literature (10% of papers), but among the cases with data, symptoms were more
297 frequent in superspreaders than in non-superspreaders. Individual and age-related heterogeneity
298 in the amount and assortative patterns of social contacts likely influence superspreading as well.
299 Evidence supports lower transmission from children compared to adults, but effect sizes have
300 been small in some studies [10,17,30,37]. Remaining heterogeneity in individual infectiousness
301 may derive from differences in genetic susceptibility [39,40], body size (accounting for age)
302 [41], baseline lung volume and function [38], immunocompromising disease or co-infection
303 [42,43], or the loudness and wetness of speech [32]. The relative importance of these
304 characteristics to SARS-CoV-2 transmission at a population level are unknown and may be
305 challenging to measure and report at scale. Future work on COVID-19 and other respiratory
306 diseases should address these hypotheses.

307 It is clear from our analysis that SARS-CoV-2 exhibits pronounced individual
308 heterogeneity of infectiousness, as evidenced by the degree distribution for index cases and the
309 estimate of the dispersion parameter k . Our estimate of k (0.27, CI: 0.25–0.28) is within the range
310 of previous estimates for a similar period of the pandemic, with values frequently in the range of
311 0.1–0.7 [3,14,15,24]. Caution should be taken when interpreting k values, which are sensitive to
312 changes in the tails of a distribution, such as superspreaders or individuals that cause no
313 secondary infections. Without robust isolated case finding and follow-up, contact tracing efforts
314 may undercount the number of zeroes, biasing k upwards [44,45]. Alternatively, backwards
315 contact tracing may be susceptible to attachment bias, where infections are preferentially
316 attributed to a known superspreader rather than a separate (known or unknown) transmitter [45].
317 Additionally, there may be publication bias or more complete contact tracing for large outbreaks

318 with an individual superspreader [9,45]. These effects would bias k downwards. Without
319 knowledge of the relative impact of these biases, it is challenging to interpret whether k is a true
320 representation of SARS-CoV-2 transmission heterogeneity. We recommend that contact tracing
321 efforts use both backward and forward contact tracing [9,17,46], with sufficient follow-up time
322 to identify non-infecting individuals, and complete reporting of contact tracing efforts (e.g.,
323 anonymized line lists with infector-infectee and other demographic information).

324 A principal limitation of our analysis was the incomplete data available in the published
325 literature. Beyond information provided about the timing and location of events, very few studies
326 reported any demographics of the exposed individuals, their COVID-19 vaccination status (once
327 introduced) or history of prior SARS-CoV-2 infection, or the density and amount of time
328 indoors. For individual index cases, some studies reported demographic information and the
329 presence/absence of symptoms, but this atypical. We also experienced difficulty with deducing
330 whether contact tracing was performed for all reported cases in transmission chains, especially
331 for terminal nodes. It was not always clear whether cases did not transmit or whether data were
332 missing due to lack of contact tracing, so these cases had to be omitted from the analysis. To
333 improve the field and our understanding of the drivers of heterogeneity in transmission, we
334 propose standard and consistent reporting on transmission for all outbreaks, as feasible, including
335 details on the epidemiological context of transmission and complete line lists of cases following
336 contact tracing, with information on case demographics (age, sex, occupation), diagnosis
337 (presence/absence of symptoms, symptom description, test date and results), the duration of
338 contact tracing, and the total number of contacts and the demographic information for contacts
339 (see Appendix 2). Details on the duration of contact tracing should include the entire time period
340 of case finding and how long cases were followed to detect any secondary cases. We recognize

341 the challenge of collecting, storing, and sharing identifiable data from outbreak investigations
342 while continuing to assure confidentiality and improve trust in the health system. However,
343 developing such a reporting system should be a priority for public health as the information has
344 important implications for reducing the spread of infectious pathogens.

345 Another limitation of this review was the wide variation in case detection methods across
346 studies. Some studies reported only symptomatic cases or only performed diagnostic tests (e.g.,
347 PCR) on symptomatic individuals, thereby missing all reporting of asymptomatic or
348 pausisymptomatic individuals and any secondary cases produced. Therefore, some of the event
349 secondary attack rates, meta-analysis estimates, and individual case degree distributions are
350 likely underestimates. However, we have no reason to believe that these biases in case
351 ascertainment would vary across different event types/settings or demographic characteristics of
352 index cases. The rankings of event types by secondary attack rate from our meta-analysis and the
353 differential characteristics of superspreaders we observed are likely robust to these issues of
354 ascertainment, but there is additional work needed to measure these effects. Furthermore, clinical
355 cases, which were less likely to be missed, are more important as public health outcomes and
356 more likely to be implicated in transmission.

357 Despite these limitations, this review found substantial heterogeneity in the transmission
358 of SARS-CoV-2, highlighting the settings and individual characteristics that might be most
359 important to target for controlling superspreading. Secondary attack rates were highest in co-
360 living situations where prolonged contact between individuals facilitated transmission, though
361 there was substantial variation in attack rates within similar settings that remained unexplained.
362 Given the moderate attack rates among minors in school and the rarity of children among
363 superspreaders, interventions targeting these age groups may be less efficient at preventing

364 SARS-CoV-2 superspreading and could be deprioritized in favor of interventions focusing on
365 adults [17,47], especially those with symptoms and individuals with many daily close contacts.
366 We advocate for consistent reporting on infectious disease outbreaks, ideally with detailed line
367 lists, to facilitate knowledge synthesis about transmission patterns and superspreading in the
368 future. Our review only covered the first phase of the pandemic, so important questions remain
369 about whether patterns in attack rates and individual-level transmission still apply to later
370 pandemic phases with significant population-level immunity. Enhanced reporting of outbreak
371 data would expedite such future investigations.

372

373 **DATA AVAILABILITY**

374 All the data were from publicly available databases. The complete database of extracted
375 information from included studies is provided in Appendix 3.

376

377 **FINANCIAL SUPPORT**

378 The research was funded by the World Health Organization. The funding agency had no role in
379 study design, data collection and analysis, decision to publish, or preparation of the manuscript.

380

381 **COMPETING INTERESTS**

382 The authors declare none.

383

384 **REFERENCES**

385 1. **Davis JT, et al.** Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. *Nature*
386 2021; **600**: 127–132.

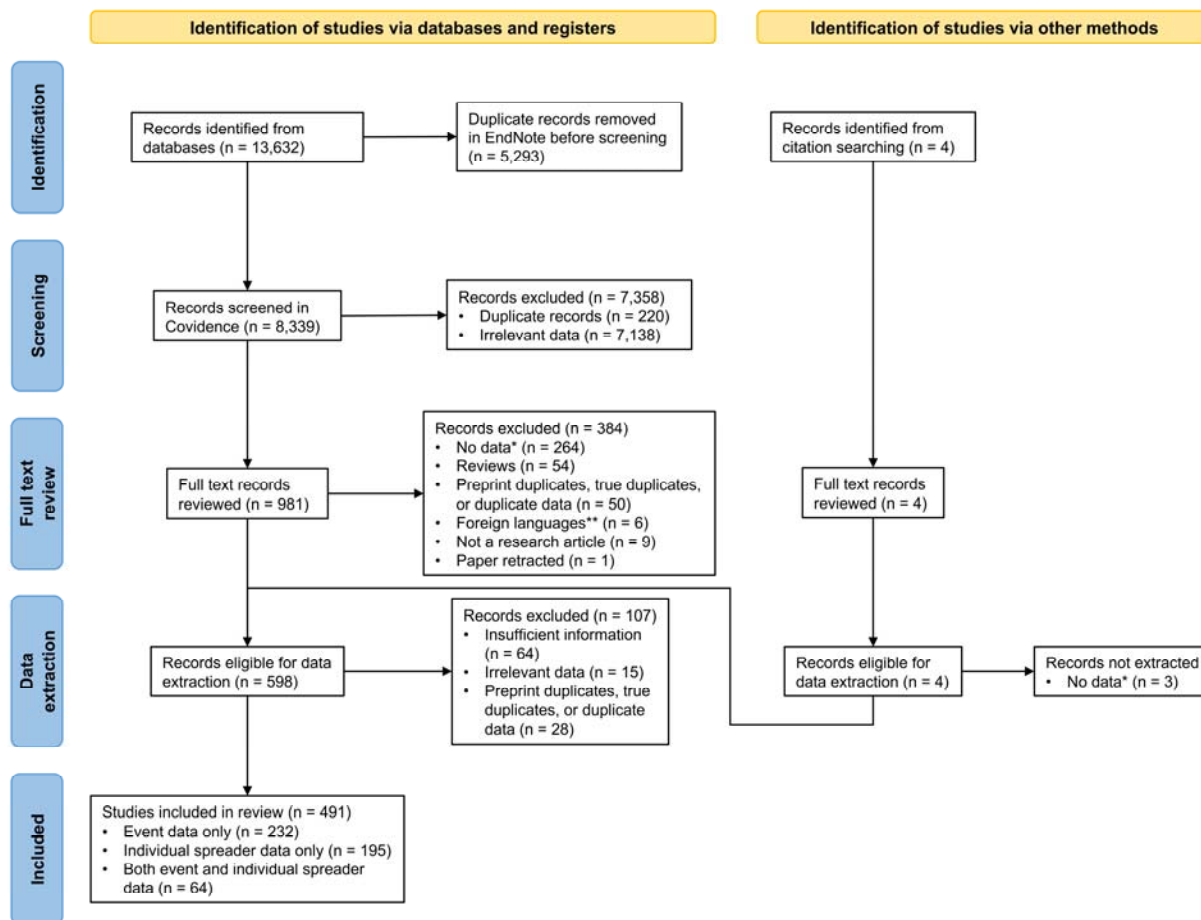
- 387 2. **Lau MSY, et al.** Characterizing superspreading events and age-specific infectiousness of
388 SARS-CoV-2 transmission in Georgia, USA. *Proceedings of the National Academy of*
389 *Sciences* 2020; **117**: 22430–22435.
- 390 3. **Adam DC, et al.** Clustering and superspreading potential of SARS-CoV-2 infections in Hong
391 Kong. *Nature Medicine* 2020; **26**: 1714–1719.
- 392 4. **Liu Y, Eggo RM, Kucharski AJ.** Secondary attack rate and superspreading events for
393 SARS-CoV-2. *The Lancet* 2020; **395**: e47.
- 394 5. **Lakdawala SS, Menachery VD.** Catch me if you can: superspreading of COVID-19. *Trends*
395 *in Microbiology* 2021; **29**: 919–929.
- 396 6. **Puhach O, Meyer B, Eckerle I.** SARS-CoV-2 viral load and shedding kinetics. *Nature*
397 *Reviews Microbiology* 2023; **21**: 147–161.
- 398 7. **Lewis D.** Superspreading drives the COVID pandemic — and could help to tame it. *Nature*
399 2021; **590**: 544–546.
- 400 8. **Hamner L, et al.** High SARS-CoV-2 attack rate following exposure at a choir practice —
401 Skagit County, Washington, March 2020. *MMWR. Morbidity and Mortality Weekly Report*
402 2020; **69**: 606–610.
- 403 9. **Lee EC, et al.** The engines of SARS-CoV-2 spread. *Science* 2020; **370**: 406–407.
- 404 10. **Thompson HA, et al.** Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
405 setting-specific transmission rates: a systematic review and meta-analysis. *Clinical Infectious*
406 *Diseases* 2021; **73**: e754–e764.
- 407 11. **Leclerc QJ, et al.** What settings have been linked to SARS-CoV-2 transmission clusters?
408 *Wellcome Open Research* 2020; **5**: 83.

- 409 12. **Madewell ZJ, et al.** Factors associated with household transmission of SARS-CoV-2: an
410 updated systematic review and meta-analysis. *JAMA Network Open* 2021; **4**: e2122240.
- 411 13. **Koh WC, et al.** What do we know about SARS-CoV-2 transmission? A systematic review
412 and meta-analysis of the secondary attack rate and associated risk factors. *PLOS ONE* 2020;
413 **15**: e0240205.
- 414 14. **Du Z, et al.** Systematic review and meta-analyses of superspreading of SARS-CoV-2
415 infections. *Transboundary and Emerging Diseases* 2022; **69**: e3007–e3014.
- 416 15. **Wang J, et al.** Superspreading and heterogeneity in transmission of SARS, MERS, and
417 COVID-19: a systematic review. *Computational and Structural Biotechnology Journal* 2021;
418 **19**: 5039–5046.
- 419 16. **Ramírez-del Real T, et al.** Individual factors associated with COVID-19 infection: a
420 machine learning study. *Frontiers in Public Health* 2022; **10**: 912099.
- 421 17. **Gupta M, et al.** Contact tracing of COVID-19 in Karnataka, India: superspreading and
422 determinants of infectiousness and symptomatic infection. *PLOS ONE* 2022; **17**: e0270789.
- 423 18. **Page MJ, et al.** The PRISMA 2020 statement: an updated guideline for reporting systematic
424 reviews. *Systematic Reviews* 2021; **10**: 89.
- 425 19. **Hodcroft EB.** *CoVariants: SARS-CoV-2 Mutations and Variants of Interest.*
426 2021(<https://covariants.org/>).
- 427 20. **Viechtbauer W.** Conducting meta-analyses in R with the metafor package. *Journal of*
428 *Statistical Software* 2010; **36**: 1–48.
- 429 21. **Lin L, Xu C.** Arcsine-based transformations for meta-analysis of proportions: pros, cons,
430 and alternatives. *Health Science Reports* 2020; **3**: e178.

- 431 22. **Senior AM, et al.** Heterogeneity in ecological and evolutionary meta-analyses: its magnitude
432 and implications. *Ecology* 2016; **97**: 3293–3299.
- 433 23. **Lloyd-Smith JO, et al.** Superspreading and the effect of individual variation on disease
434 emergence. *Nature* 2005; **438**: 355–359.
- 435 24. **Endo A, et al.** Estimating the overdispersion in COVID-19 transmission using outbreak sizes
436 outside China. *Wellcome Open Research* 2020; **5**: 67.
- 437 25. **Yao L, et al.** COVID-19 super spreading event amongst elderly individuals — Jilin
438 Province, China, January 2021. *China CDC Weekly* 2021; **3**: 211–213.
- 439 26. **Chu DKW, et al.** SARS-CoV-2 superspread in fitness center, Hong Kong, China, March
440 2021. *Emerging Infectious Diseases* 2021; **27**: 2230–2232.
- 441 27. **Althaus CL.** Ebola superspreading. *The Lancet Infectious Diseases* 2015; **15**: 507–508.
- 442 28. **Paunio M, et al.** Explosive school-based measles outbreak: intense exposure may have
443 resulted in high risk, even among revaccinees. *American Journal of Epidemiology* 1998; **148**:
444 1103–1110.
- 445 29. **Nikolay B, et al.** Transmission of Nipah virus — 14 years of investigations in Bangladesh.
446 *The New England Journal of Medicine* 2019; **380**: 1804–1814.
- 447 30. **Madewell ZJ, et al.** Household transmission of SARS-CoV-2: a systematic review and
448 meta-analysis. *JAMA Network Open* 2020; **3**: e2031756.
- 449 31. **Bi Q, et al.** Insights into household transmission of SARS-CoV-2 from a population-based
450 serological survey. *Nature Communications* 2021; **12**: 3643.
- 451 32. **Asadi S, et al.** Aerosol emission and superemission during human speech increase with voice
452 loudness. *Scientific Reports* 2019; **9**: 2348.

- 453 33. **Mürbe D, et al.** Aerosol emission in professional singing of classical music. *Scientific*
454 *Reports* 2021; **11**: 14861.
- 455 34. **Aizawa Y, et al.** Coronavirus disease 2019 cluster originating in a primary school teachers'
456 room in Japan. *The Pediatric Infectious Disease Journal* 2021; **40**: e418–e423.
- 457 35. **Gettings JR, et al.** SARS-CoV-2 transmission in a Georgia school district — United States,
458 December 2020–January 2021. *Clinical Infectious Diseases* 2022; **74**: 319–326.
- 459 36. **Chen PZ, et al.** Heterogeneity in transmissibility and shedding SARS-CoV-2 via droplets
460 and aerosols. *eLife* 2021; **10**: e65774.
- 461 37. **Ke R, et al.** Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial
462 heterogeneity in infectiousness. *Nature Microbiology* 2022; **7**: 640–652.
- 463 38. **Zhou J, et al.** Viral emissions into the air and environment after SARS-CoV-2 human
464 challenge: a phase 1, open label, first-in-human study. *The Lancet Microbe* 2023; **4**: e579–
465 e590.
- 466 39. **Severe Covid-19 GWAS Group, et al.** Genomewide association study of severe Covid-19
467 with respiratory failure. *The New England Journal of Medicine* 2020; **383**: 1522–1534.
- 468 40. **Zhao J, et al.** Relationship between the ABO blood group and the coronavirus disease 2019
469 (COVID-19) susceptibility. *Clinical Infectious Diseases* 2021; **73**: 328–331.
- 470 41. **Edwards DA, et al.** Exhaled aerosol increases with COVID-19 infection, age, and obesity.
471 *Proceedings of the National Academy of Sciences* 2021; **118**: e2021830118.
- 472 42. **Caillard S, et al.** SARS-CoV-2 viral dynamics in immunocompromised patients. *American*
473 *Journal of Transplantation* 2021; **21**: 1667–1669.
- 474 43. **Aytillo T, et al.** Shedding of viable SARS-CoV-2 after immunosuppressive therapy for
475 cancer. *New England Journal of Medicine* 2020; **383**: 2586–2588.

- 476 44. **Zhao S, et al.** Inferencing superspreading potential using zero-truncated negative binomial
477 model: exemplification with COVID-19. *BMC Medical Research Methodology* 2021; **21**: 30.
- 478 45. **Taube JC, Miller PB, Drake JM.** An open-access database of infectious disease
479 transmission trees to explore superspreader epidemiology. *PLOS Biology* 2022; **20**:
480 e3001685.
- 481 46. **Endo A, et al.** Implication of backward contact tracing in the presence of overdispersed
482 transmission in COVID-19 outbreaks. *Wellcome Open Research* 2021; **5**: 239.
- 483 47. **Davies NG, et al.** Age-dependent effects in the transmission and control of COVID-19
484 epidemics. *Nature Medicine* 2020; **26**: 1205–1211.
- 485



486

487 **Figure 1.** PRISMA flow diagram for the systematic review and meta-analysis of SARS-CoV-2

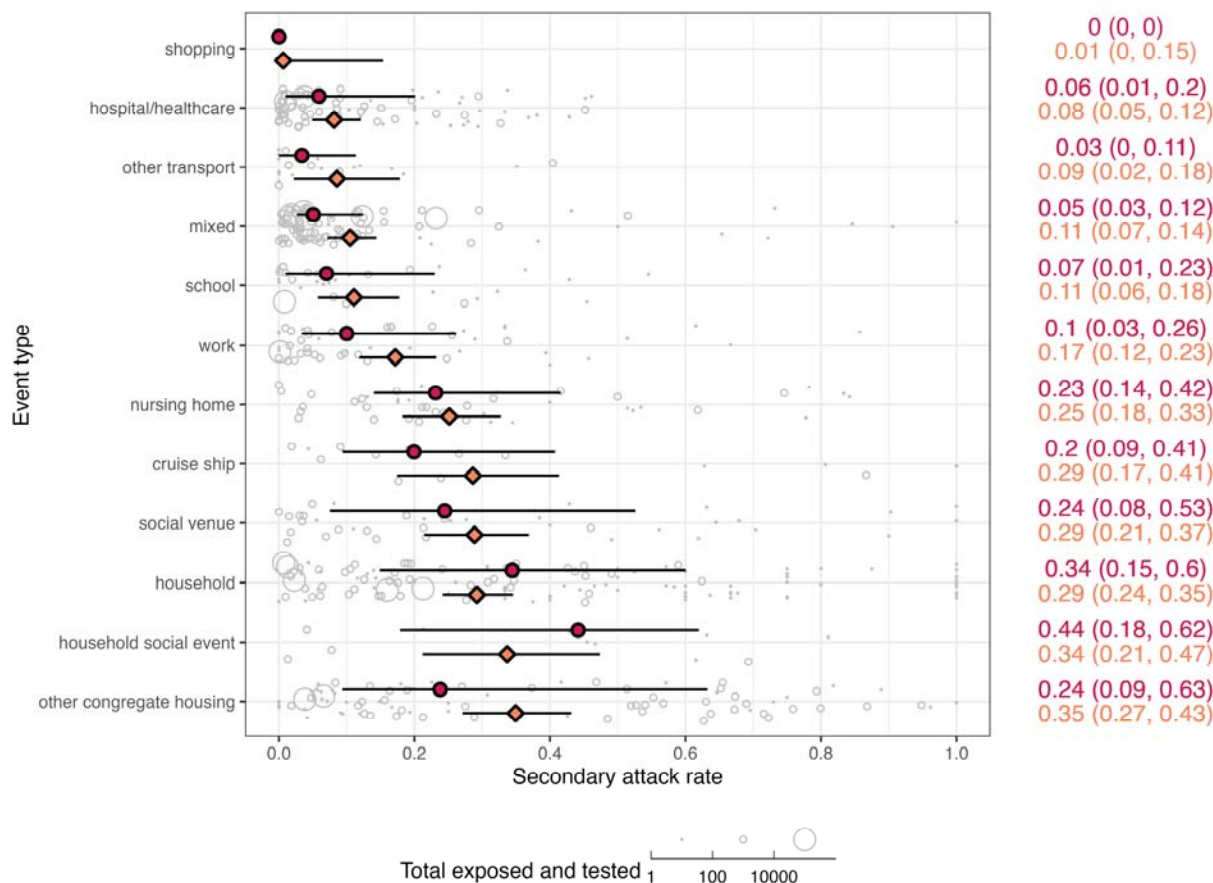
488 superspreading reported in the published literature. *There were 4 types of data that we sought to

489 include: 1) transmission chain; 2) number of index cases, number of contacts, and number of

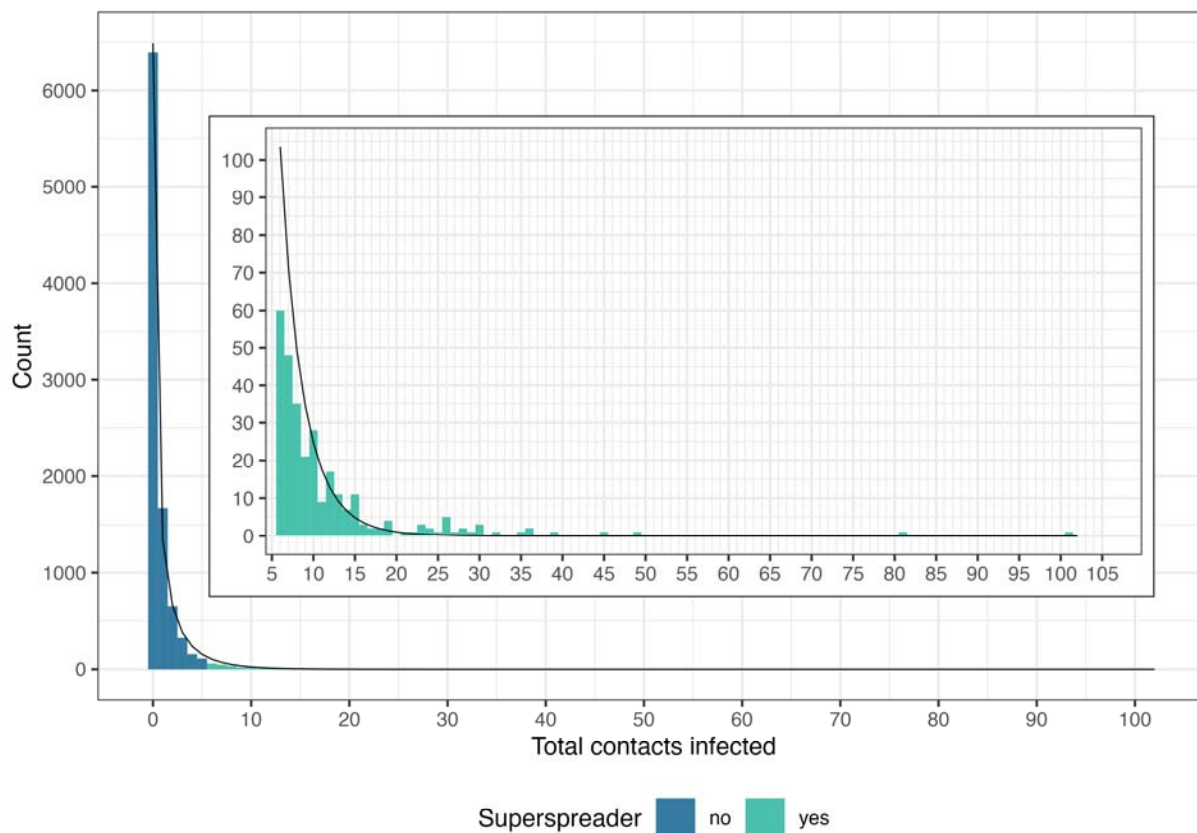
490 infected contacts; 3) number of index cases and number of infected contacts; or 4) secondary

491 attack rate. **Languages other than Spanish, Chinese, French, Turkish, German, and Portuguese.

492



493
 494 **Figure 2.** SARS-CoV-2 secondary attack rates across 12 event types occurring between
 495 December 2019 and August 2021 reported in the literature across 592 events from 296 studies.
 496 Individual event data secondary attack rates are shown as grey bubbles, varying in size according
 497 to the total number of individuals exposed and tested from the event. Median secondary attack
 498 rate for each event type is shown as red circle with a line representing the interquartile range;
 499 values are in red on the right side of the figure. Meta-analysis estimated secondary attack rate for
 500 each event type is shown as an orange diamond with a line representing the estimated 95%
 501 confidence interval; values are in orange on the right side of the figure. Event types were ranked
 502 by increasing estimated mean secondary attack rate along the left axis.
 503



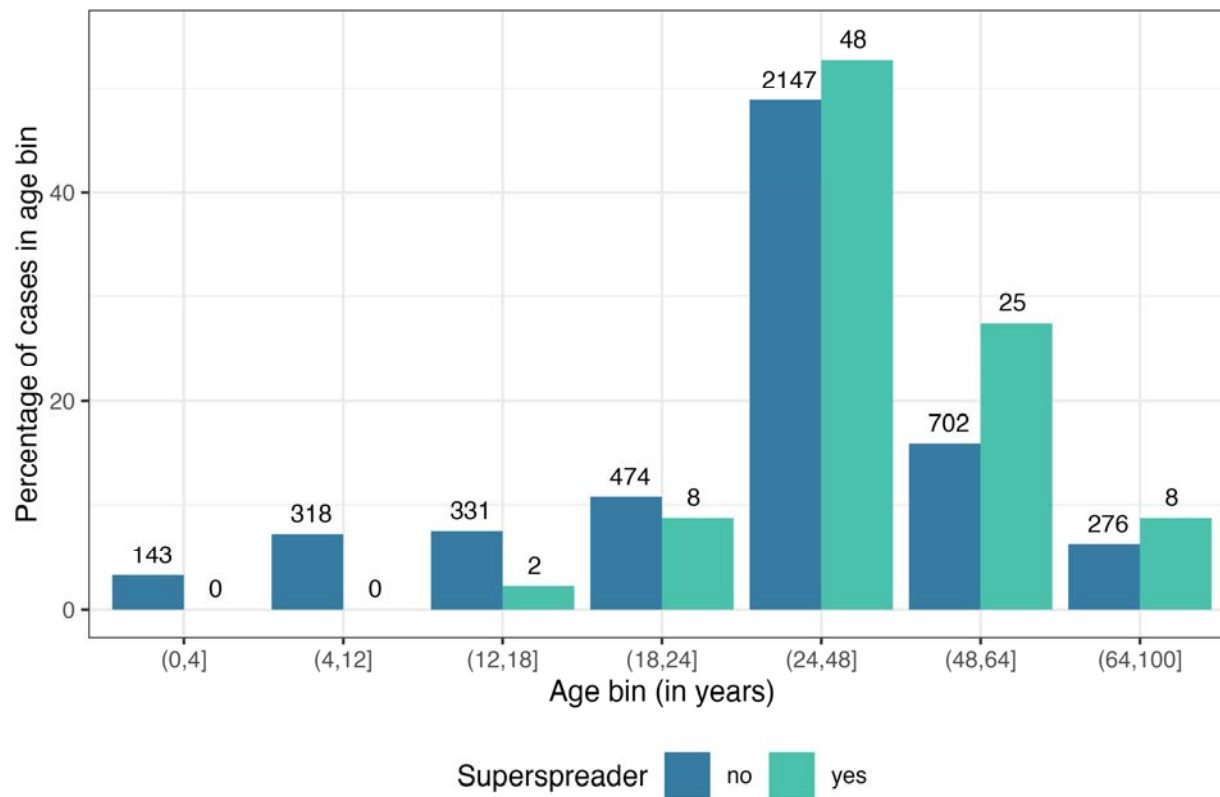
504

505 **Figure 3.** Distribution of secondary contacts infected by individual index cases (N = 9,591) for
506 SARS-CoV-2 cases occurring between December 2019 and July 2021 reported in 259 studies.

507 The black line shows the fit of the distribution to the expected negative binomial distribution.

508 The inset shows a portion of the same data to highlight the distribution of superspreaders (index
509 cases with >5 secondary cases).

510



511
512 **Figure 4.** Comparison of the age distribution of superspreading index cases. The bars show the
513 percentage of individuals within an age bin across superspreaders (index cases with >5
514 secondary cases) and non-superspreaders. Numbers above the bars display the raw totals and
515 percentages are shown in Table 3.
516

517 **Table 1.** Description of variables extracted from papers in the systematic review of SARS-CoV-
 518 2 superspreading from December 2019 to July 2021.

All papers	<ul style="list-style-type: none"> • Title • Author(s) • Publication year, volume, and issue • Journal • Study location(s) <ul style="list-style-type: none"> ○ Country ○ Administrative unit(s): state/province, county, city • General study time period (e.g., start and end year/month of data collection) • Diagnostic testing method (PCR, serology, rapid antigen tests, symptom diagnosis, or mixed) • Variant name • Any reported prevention measures implemented in the event (e.g., masking, social distancing) • Number of exposed people with reported demographic characteristics (age, sex) and vaccination status
Events	<ul style="list-style-type: none"> • Type of event/setting (e.g. nursing home residents, household transmission study, or school) • Start and end date of event
Index case	<ul style="list-style-type: none"> • Demographic characteristics (age, sex, occupation) <ul style="list-style-type: none"> ○ Age group: ≤ 4 years, 5–12 years, 13–18 years, 19–24 years, 25–48 years, 49–64 years, ≥ 65 years • Symptom onset (if applicable) and diagnosis dates • Symptoms (text descriptions or presence/absence) • Real-time PCR cycle threshold (Ct) value

	<ul style="list-style-type: none">• Specimen type• Clinical outcome (if applicable)• Setting of contact (e.g., work, social, and school)
--	--

519 **Table 2.** Types of SARS-CoV-2 secondary transmission events occurring between December
 520 2019 and August 2021 reported in the literature. Heterogeneity across event types was assessed
 521 based on the variance and interquartile range of secondary attack rates. Outlier events were
 522 identified for each event type as events that exceeded the estimated upper confidence interval of
 523 the meta-analysis estimated SAR for that event type or were greater than 50%.

Event type	Description of outbreak location	N	Minimum and maximum secondary attack rate
1	cruise ship or other densely populated watercraft (e.g., fishing vessel, aircraft carrier)	16	0.02, 1
2	transport mode other than ships (e.g., airplane, train, car)	20	0, 0.4
3	households, defined as co-living individuals or close contacts who always meet each other but possibly not living together (e.g., couples in romantic relationship)	115	0, 1
4	hospital or healthcare facility, including patients, healthcare workers, and nursing home workers (if worker data was provided separately from nursing home residents)	89	0, 0.46
5	workplace (e.g., office), including correctional officers and teachers and staff at schools	51	0, 0.86
6	school (data on students only)	32	0, 0.54
7	public social venue (e.g., bar, concert, sporting event)	39	0, 1
8	private social event with members of multiple households (e.g., dinner with neighbors or extended family)	12	0.01, 0.81
9	shopping (activities in shops, markets, and department stores)	2	0, 0
10	nursing home or long-term care facility (residents only or residents and healthcare workers if not described separately in the paper)	41	0, 0.84

11	congregate housing other than nursing home or long-term care facility (e.g., homeless shelter, prison, summer camp)	84	0, 1
12	mixed locations, included any combination of the above but not described separately in the paper	91	0, 1

524 **Table 3.** Statistical comparisons of SARS-CoV-2 superspreaders to non-superspreaders based on
 525 features reported in the literature in 259 studies for cases occurring between December 2019 and
 526 July 2021.

Feature of comparison	Percentage or estimated mean for non-superspreaders (total observations)	Percentage or estimated mean for superspreaders (total observations)	Statistical test results
Female	40% (N = 4,543)	38% (N = 102)	$\chi^2_1 = 0.09, P = 0.76^a$
Presence of symptoms (symptomatic)	76% (N = 841)	89% (N = 70)	$\chi^2_1 = 5.4, P = 0.02^a$
Age (in bins)	(N = 4,391)	(N = 91)	$\chi^2_6 = 21.7, P = 0.001^a$
≤4 years	3%	0%	
5–12 years	7%	0%	
13–18 years	8%	2%	
19–24 years	11%	9%	
25–48 years	49%	53%	
49–64 years	16%	27%	
≥65 years	6%	9%	
Age (≥18 years)	84% (N = 4,391)	99% (N = 91)	$\chi^2_1 = 14.1, P < 0.0001^b$
Age (in years)	34.8 (N = 4,391)	43.8 (N = 91)	$t_{94.4} = 5.2, P < 0.0001^b$
Ct value	26.7 (N = 140)	24.8 (N = 10)	$t_{2.69} = -0.5, P = 0.62^b$
Total contacts	36 (N = 471)	79 (N = 59)	$\chi^2_1 = 56.6, P < 0.0001^c$

527 **Table 4.** Summary statistics describing the distribution of secondary cases among individual SARS-CoV-2 index cases occurring
 528 between December 2019 and July 2021 reported in the literature across 259 studies.

Data	Sample size	Percentage with 0 contacts infected	Percentage with 1–5 contacts infected	Percentage with >5 contacts infected	Maximum contacts infected	Estimated mean contacts infected (95% CI)	Estimated overdispersion, k (95% CI)
All rows	9,591	67%	30%	3%	101	0.88 (0.84–0.92)	0.27 (0.25–0.28)
Female	1,866	75%	22%	2%	30	0.63 (0.56–0.71)	0.18 (0.16–0.21)
Male	2,779	74%	24%	2%	101	0.76 (0.69–0.84)	0.17 (0.15–0.19)
Asymptomatic	214	79%	17%	4%	25	0.75 (0.43–1.08)	0.11 (0.07–0.16)
Symptomatic	697	44%	47%	9%	81	2.06 (1.8–2.3)	0.43 (0.36–0.49)
Age ≤4 years	143	90%	10%	0%	3	0.2 (0.07–0.32)	0.09 (0.01–0.17)
Age 5–12 years	318	97%	3%	0%	3	0.04 (0.006–0.07)	0.03 (–0.006–0.06)
Age 13–18 years	333	94%	6%	1%	26	0.23 (0.08–0.38)	0.03 (0.01–0.05)
Age 19–24 years	482	88%	10%	2%	19	0.38 (0.23–0.52)	0.06 (0.04–0.09)
Age 25–48 years	2,195	76%	22%	2%	101	0.73 (0.65–0.82)	0.15 (0.13–0.17)
Age 49–64	727	54%	43%	3%	35	1.16 (1.01–1.31)	0.43 (0.36–0.51)
Age ≥65 years	284	58%	39%	3%	18	1 (0.79–1.21)	0.43 (0.3–0.57)

Ct value ≤ 25	67	42%	52%	6%	12	1.48 (1–1.96)	0.86 (0.31–1.42)
Ct value > 25	83	52%	41%	7%	26	1.45 (0.87–2.02)	0.4 (0.2–0.6)
0–10 total contacts	281	49%	48%	2%	9	1.12 (0.93–1.31)	0.83 (0.53–1.14)
11–100 total contacts	195	47%	34%	19%	39	3.1 (2.29–3.91)	0.32 (0.23–0.41)
101–1000 total contacts	54	35%	37%	28%	101	8 (3.92–12.07)	0.28 (0.16–0.4)

529