THE LANCET
Infectious Diseases

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Koo JR, Cook AR, Park M, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis* 2020; published online March 23. hhttps://doi.org/10.1016/S1473-3099(20)30162-6

Supplementary Information

A full description of how the realistic yet synthetic population was generated, and the respiratory infection model incorporated is discussed below.

Generation of the Synthetic Population

In order to create representative household structures, census data was required which detailed the population attributes within the household and as individuals. The latest Census report[1], carried out in 2010, provides 14 register-based tables containing population counts at the planning area and subzone spatial district level further divided by age group, ethnicity and type of dwelling. The remaining 178 tables present other attributes collected through sample enumeration based on a representative sample of approximately 200,000 households. We selected the variables required to build a synthetic population which appropriately represents the age structure and movement dynamics for epidemic spread simulations (Supplementary Table 1). An heuristic search approach was used where 3.77 million individuals with Singaporean citizenship or permanent resident status were created using the three primary attributes; age group, ethnicity and gender. The remaining attributes for each individual were drawn randomly from the attribute's marginal distributions. Summary tables equivalent to the census tables were calculated and the fit compared via Pearson's chi-squared statistic. Zero count cells were avoided by setting them to 0.1. For each attribute, a Monte Carlo swapping algorithm developed in C++ was utilized (hereby called the swapping algorithm, SA). It continually swaps two random individual's data until the improvement in the measure of fit becomes negligible (<0.001% improvement) when comparing the synthetic Census table outputs and actual Census totals. A group of individuals were thus allocated to each spatial unit, planning area or subzone, that have characteristics as represented in the Census tables.

Partners were then allocated using a matching algorithm which was constrained by age group and ethnicity between husband and wife according to the Census tabulation on married couples. Couple matching was first carried out on co-ethnic couples before inter-ethnic couples due to the majority being the former. Location attributes were set for the couple at one spouse's location and the SA was carried out again to ensure the synthetic couples resembled the Census data on spousal characteristics such as education level, religion, working status and income level.

Following partner formation, households were characterized. Attributes including the features of the household heads, household size, dwelling type, number of working people, monthly household income and predominant language, were generated for each synthetic household. The SA was performed before the introduction of further variables which were the number of family nuclei, number of generations and age of the youngest child of the household head, before again implementing the SA. Constraints were applied for this SA on the lower bound of household size, number of generations and number of family nuclei for married household heads with or without children, as represented in the Census tables.

Once partners and households had been created for each geographical spatial unit, individuals and couples were assigned to these synthetic households to form household heads or members. Maternal status was randomly attributed among ever-married females for each age group at each household in a spatial unit. For mother-child relationships, the frequency distributions of age differences between mothers and children by the child's year of birth and ethnic group were derived from the Registry of Births and Deaths, 2010[2]. For widowed and divorced men, ex-spouse ages were generated based on the distribution of the age difference between husbands and wives, which was extracted from the Statistics on Marriages and Divorces, 2011[3], in order to assign each child a widowed or divorced father. Parents of household heads and their partners were assigned in the same manner. During child assignment between parents, household characteristics such as the household size, family nuclei number, number of generations and number of working people were retained as thresholds to retain the generated household configurations.

The 1.14 million households consisting of single or multi-generational parents and children were assigned to 76,834 georeferenced residential buildings with postal codes as listed by the Singapore Land Authority on their portal Onemap[4] whilst preserving the characteristics of the households listed for each spatial unit using the SA. Each of the 3.77 million Singaporean individuals were then allocated further attributes from the Labour Force

Report, 2010, to account for relationships among socio economic attributes. A final SA was employed for each attribute to match the joint distributions within the Census until a 99% match was determined between the actual and synthetic Census tables created from the SA.

Allocation of Workplace and School

The time taken travelling from home to work or school for each individual was used to determine the corresponding address. Smart EZ-Link data, which records the public transport activities for all individuals in one month (1st to 31st August, 2013), was used to determine the distance individuals are able to travel across Singapore. The boarding and terminating ridership data are recorded at the card user level at bus stop and train station locations. A full list of both bus and train lines and nodes was obtained from the LTA datamall[5] as a reference dataset for travel times to be calculated across the public transport network in Singapore.

The EZ-Link dataset was processed to remove noise by eliminating trips with empty alighting stops and infrequent card users with fewer than 30 journeys a month as we assumed individuals who are students or working would utilise public transportation twice a day. For each remaining card ID, we determined the stay duration for each journey, the morning journey (after 12am and before 9am) and evening journey pathways (after 5pm to 12am). Potential home to work or school journeys were assigned to be where the stay duration was greater than 6 hours based on Singapore's Household Interview Travel Study[6] with the most frequent and first morning journey departure point at the home location and most frequent morning alighting stop at work or school. Three EZ-Link card types exist within Singapore; Adult, Child and Senior. The adult and senior pass data were used for work addresses, and child passes for school addresses.

A 250m buffer was created around every bus stop and 500m for train stations, representing a 5 to 10-minute walk for the former and 10 to 15-minute walk for the latter in the dense and vertical structure of Singapore's urban city. The high spatial density of the bus stop network also required the smaller buffer size. All home, work and school addresses were allocated a corresponding bus and/or train station for access with an associated probability,

$$P(a,b) \propto \frac{1}{d^2}$$

Equation 1

where location a is a home, work or school location, b a bus or train node and d, the Euclidean distance between them. The probabilities are normalised for each location.

Those that fell outside of a buffer zone were allocated the nearest train or bus stop. The school directory was obtained from data.gov[7] and workplace postal codes collected from the Greenbook[8] for this analysis. Based on the EZ-Link dataset of origin and destinations, and time taken between nodes, a network analysis was performed in ArcGIS 10.6[9] to compute the full public transportation network and time required to travel between each transportation node.

For each home location of the working population, potential journeys were created from the three highest probable travel nodes to relevant workplaces, as allocated by the occupation type and travel times in the synthetic population process, and highest probable work travel nodes simultaneously. The total probability of the home-work pair was determined to be the product of Equation 1 at the home and work site. A ten-minute leeway was allowed however in travel time which allowed for at least 3 potential journeys for each working individual. Once all journeys were ranked from highest probability to lowest, one of the top three were selected randomly to be representative. The same process was repeated for students except a maximum capacity was allocated to each school as determined by school enrolment sizes[10], therefore schools were iteratively filled as random students were selected with no spatial bias. For each working individual and student, a travel trajectory is thus provided which serves as a contact network for epidemiological modelling.

Transmission dynamic component

The open-source FluTE model was originally developed to model pandemic influenza [33] in an age-structured population capturing workplaces, households and neighbourhoods. We have adapted FluTE here to fully utilise GeoDEMOS with its synthetic population and generation of travel trajectories to estimate disease spread. We called the hybrid model GeoDEMOS-R, reflecting respiratory disease spread in Singapore, which contains highly representative information at the individual level and incorporates the spatial distribution and hierarchy of households, families, colleagues and students for the Singaporean population.

Suppose i and j are two individuals in the synthetic population with j becoming infected, we denote the probability of j infecting i on day t in location type g is given by

$$P_{ijg}(t) = \beta_g I\left(t - t_j\right).$$

Equation 2

Here β_g is constant for location type g, defined as the home, workplace or school which both individuals belong to; β_g is obtained from a contact rate study[21] for different social group settings in Singapore where the contact rates serve as the likelihood for individuals to infect one another at the specific group locations and the wider spatial subzone area. Individuals are expected to minimally interact with others outside of their home, work and school by travel transit or in local commercial sites. During the simulation, school and work attendance were assumed to be different where 65% of children school stayed at their home address when infected[11], due to the use of helpers and family network culture which allows parents to continue working. Strong work presenteeism behaviour was assumed among working adults[12] with a withdrawal rate of 13% to the home community.

We assume that no individuals have existing immunity to the novel virus. The infectiousness factor $I(t - t_j)$ is determined by the basic reproduction number, R_0 and an estimated infectivity profile[14]. I is the infectivity profile function, t_j is the day of individual j's onset of symptoms; i.e. the end of t_j's incubation period. Hence $I\left(t - t_j\right)$ gives the infectiousness factor for j at $t - t_j$ days after the onset of symptoms. The calibration of the model's R_0 followed the methods outlined in the original FluTE model[13]. We additionally assumed that an infected individual had a 92.5%[15] chance of becoming symptomatic with a corresponding 7.5% asymptomatic rate, and they became infectious after a mean incubation period of 5.3d[16,17]. A symptomatic individual would be hospitalised within 7d after onset of symptoms, with a mean of 3.5d[14]. Hospitalised individuals are unable to infect any other individual and are considered to have recovered and be completely immune after their hospitalisation period. All individuals, regardless of whether they are symptomatic or not, had an incubation period and the same infectivity duration, but asymptomatic cases were assumed to be 50% less infectious (infectivity factor) within contact events based on Nishiura and Colleagues findings.[18]

An epidemic simulation was run over 80d where each day consisted of a day and night step. During the day step, workers interacted with individuals in the same workgroup and those located in the same subzone at different rates with higher rates allocated to the former. Student interactions were constructed to favour students similar to their own age with lower contact rates among those in the same subzone as their school addresses, including teachers. Individuals who are not working or studying were modelled to interact with people in the same residential subzone during the day, hereby called the home community.

In the night step, individuals had high contact rates with family members and other individuals from their neighbourhood and subzones. The rates of physical contacts were determined by the two interacting individuals' ages and their relationship, with children in the same family having the highest probability of having disease-transmitting contact with his or her family members. We also added weekends which simulated closer interactions between families and the home communities within the subzones. The day steps of the sixth and seventh day for each week were consequently substituted with night steps.

Overall, the probability of individual i getting infected from location type g on a day t is therefore given as,

$$p_G(t) = \frac{1}{|G|}\sum_{j\in G} P_{ijg}(t) = \frac{\beta_g}{|G|}\sum_{j\in G_t} I(t - t_j).$$

Equation 3

Here G is a set of individuals of location type g, G_t is the subset of all individuals who belongs to set G and are infectious on day t. We use $|\cdot|$ to denote the size a set of individuals. Hence, $|G|$ is the total number of people in set G. The number of people in set G that would be infected on day t can be denoted by a random variable $X_G(t)$ and

$$X_G(t) \sim Bin\big(|G| - |G_t| - |G_r|,\ p_G(t-1)\big).$$

Equation 4

where G_r is the subset of G consisting of individuals that have been removed through hospitalisation and subsequent recovery. The total number of people α infected on day t can then be expressed by summing over all the different sets of individuals in the population,

$$\alpha(t) \sim \sum_G X_G(t)\,.$$

Equation 5

Variable	Input/Categories	
Basic Demographic Features of Individual		
Age	0 to 100 Years	
Gender	• Male • Female	
Race	• Chinese • Malay	• Indian • Other
Citizenship	• Singapore Citizen • Permanent Resident	
Religion	• No Religion • Buddhism • Taoism • Islam • Hinduism	• Sikhism • Catholic • Other Christian • Other Religion
Born in Singapore	• Yes • No	
Family and Fertility		
Marital Status	• Single • Married	• Divorced • Widowed
Number of Children	0 to 9 Children (for Ever-married Females Only)	
Work and Education Status		
Current Education Level	• Pre-Primary • Primary • Secondary • Post-Secondary (Non-Tertiary)	• Polytechnic • Professional Qualification and Other Diploma • University
Highest Qualification Attained	• No Qualification • Primary • Lower Secondary • Secondary	• Post-Secondary (Non-Tertiary) • Polytechnic • Professional Qualification and Other Diploma • University
Working Status	• Economically Inactive • Working • Unemployed	
Monthly Salary	• Not Working • Below 1,000 • 1,000–1,499 • 1,500–1,999 • 2,000–2,999 • 3,000–3,999 • 4,000–4,999	• 5,000–5,999 • 6,000–6,999 • 7,000–7,999 • 8,000–8,999 • 9,000–9,999 • 10,000 & Above
Occupation	• Not Working • Senior Officials & Managers • Professionals • Associate Professionals & Technicians • Clerical Workers • Service & Sales Workers	• Agricultural & Fishery Workers • Production Craftsmen & Related Workers • Plant & Machine Operators & Assemblers • Cleaners, Labourers & Related Workers • Workers Not Classifiable by Occupation
Usual Hours Worked	• Below 30 Hours per Week • 30–34 • 35–39 • 40–44 • 45–49	• 50–54 • 55–59 • 60–64 • 65 & Over
Industry	• Manufacturing • Construction • Wholesale & Retail Trade	• Professional Services Administrative & Support Services • Public Administration & Education

	• Transportation & Storage • Accommodation & Food Services Information & Communications • Financial & Insurance Services • Real Estate Services	• Health & Social Services • Arts • Entertainment & Recreation • Other Community • Social & Personal Services • Others
Enrolled in National Service	• Yes • No	
Household Features		
Partner ID	Unique Integer	
Household ID	Unique Integer	
Head of Household	• Yes • No	
Housing Type	• Other HDB Dwellings • HDB 1- and 2-Room Flats • HDB 3-Room Flats • HDB 4-Room Flats • HDB 5-Room Flats and Executive Flats	• Condominiums and Private Flat • Landed Bungalows • Semi-Detached Bungalows • Terrace Houses • Others
Geographical Location		
Subzone	215 Districts	
Planning Area	36 Districts	
Region	• East • North • North-East	• West • Central
Travel Times		
Usual Mode of Transport to School/Work	• Public Bus Only • MRT Only • MRT & Public Bus Only • MRT & Car Only • MRT & Another Mode • Taxi Only	• Car Only • Private Chartered Bus/Van Only Lorry/Pickup Only • Motorcycle/Scooter Only • Others • No Transport Required
Travelling Time to School/Work	• Up to 15 Minutes • 16–30 • 31–45	• 46–60 • More Than 60
Other		
Mobility Status (Elderly of 65 Years and Over)	• Non-ambulant • Semi-ambulant • Ambulant	
Overseas More Than 6 Months	• Yes • No	

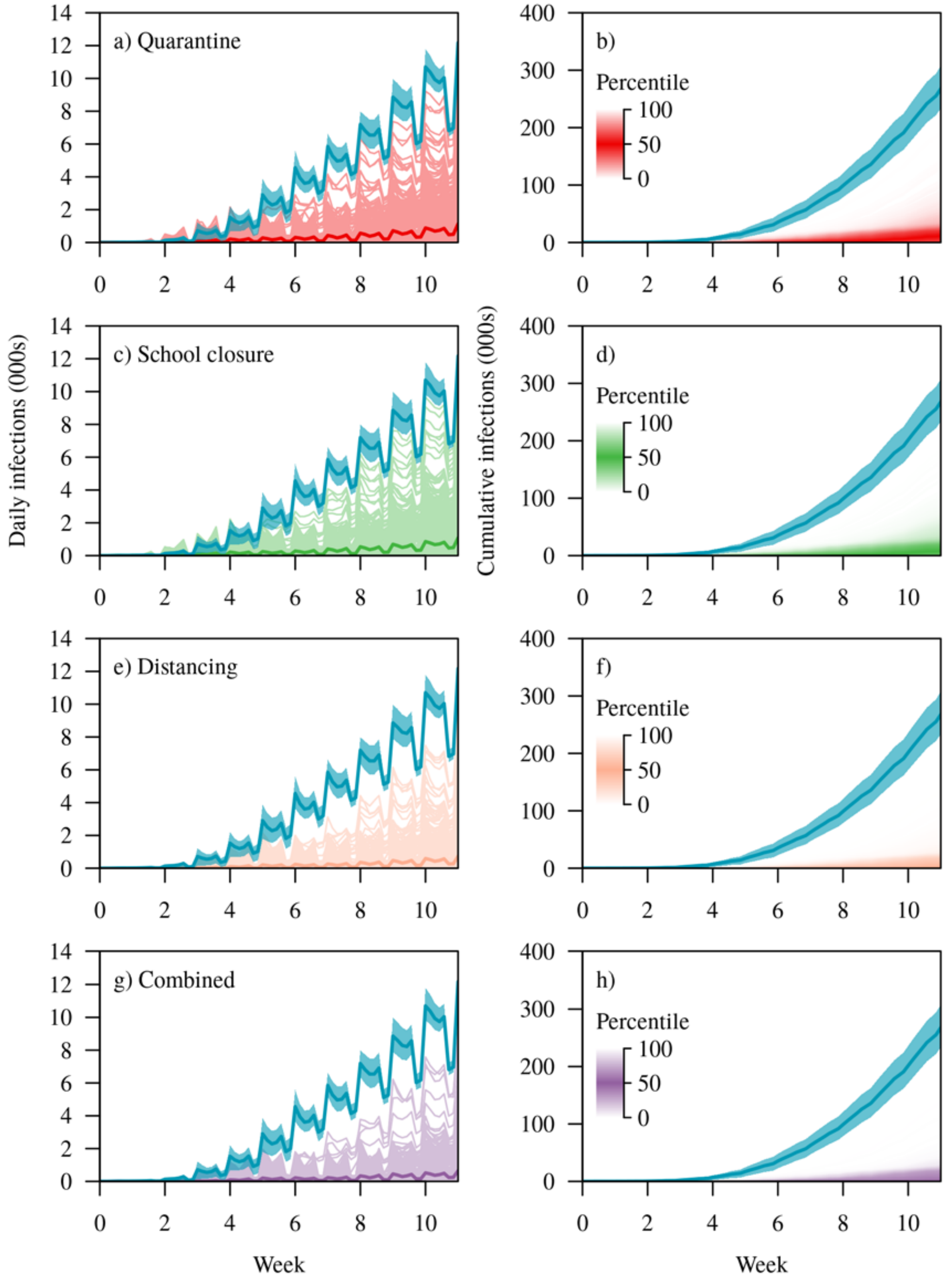
Supplementary Table 1. Variables allocated to individuals within the synthetic population with corresponding values.

R_0	Infection Location	No Control	Quarantine	School Closure	Work Distancing	Combined
1·5	Total	340 000 (311 000-378 000)	121 000 (55 000-180 000)	113 000 (42 000-179 000)	73 000 (19 000-124 000)	58 000 (5 400-123 000)
	Home Community	169 000 (159 000-194 000)	54 000 (26 000-83 000)	51 000 (13 000-89 000)	31 000 (5 400-58 000)	22 000 (2 100-51 000)
	School	1 500 (1 400-1 800)	500 (300-800)	500 (100-800)	400 (51-600)	300 (25-500)
	Work	170 000 (160 000-195 000)	67 000 (37 000-92 000)	61 000 (28 000-97 000)	42 000 (17 000-69 000)	36 000 (8 100-61 000)
2·0	Total	829 000 (789 000-868 000)	540 000 (376 000-671 000)	540 000 (37 7000-661 000)	417 000 (268 000-554 000)	411 000 (223 000-540 000)
	Home Community	428 000 (414 000-448 000)	273 000 (206 000-343 000)	273 000 (216 000-343 000)	209 000 (146 000-268 000)	207 000 (128 000-270 000)
	School	4 500 (4 400-5 000)	2 600 (2 000-3 500)	2 600 (2 100-3 400)	2 000 (1 400-2 700)	2 000 (1 200-2 700)
	Work	396 000 (386 000-413 000)	264 000 (200 000-329 000)	264 000 (209 000-329 000)	206 000 (141 000-259 000)	202 000 (126 000-263 000)
2·5	Total	1 294 000 (1 268 000-1 322 000)	1 129 000 (1 268 000-1 322 000)	1 122 000 (979 000-1 211 000)	1 010 000 (815 000-1 108 000)	991 000 (778 000-1 088 000)
	Home Community	694 000 (687 000-718 000)	596 000 (533 000-645 000)	592 000 (505 000-644 000)	528 000 (454 000-579 000)	517 000 (408 000-575 000)
	School	8 500 (8 300-8 800)	7 000 (6 200-7 700)	7 000 (5 700-7 700)	6 000 (5 100-6 900)	6 000 (4 500-6 800)
	Work	591 000 (587 000-607 000)	526 000 (481 000-562 000)	523 000 (460 000 -561 000)	476 000 (420 000-516 000)	468 000 (382 000-513 000)

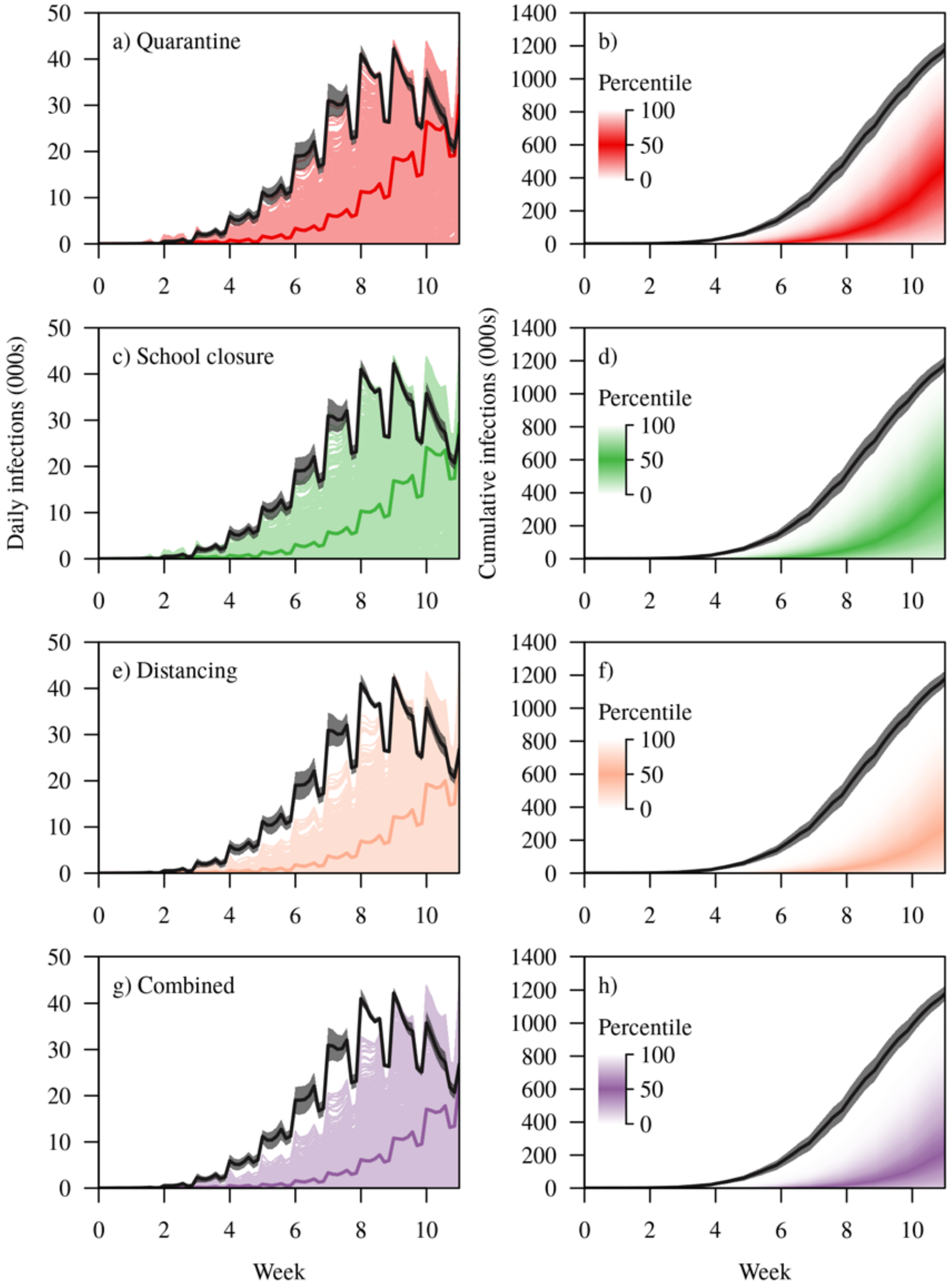
Supplementary Table 2. The cumulative number of infections at day 80 with a higher asymptomatic proportion of 22.7% by location and intervention (quarantine, school closure, work distancing and a combination of all three) for the median simulation at R_0 of 1.5, 2.0 and 2.5. The IQR is presented in brackets. Due to the stochasticity within each simulation, numbers less than 20 indicate near-complete suppression, and should not be compared to assess which is more effective. All numbers above 100 have not been rounded. Those up to 10 000 have been rounded to the nearest hundred, and those above have been rounded to the nearest thousand. Some discrepancies will therefore exist in the summations.

Asymptomatic Proportion (%)	Infection Location	No Control	Quarantine	School Closure	Work Distancing	Combined
30%	Total	365 000 (336 000–394 000)	195 000 (125 000–249 000)	191 000 (113 000–242 000)	131 000 (61 000–190 000)	122 000 (47 000–187 000)
	Home Community	181 000 (166 000–193 000)	94 000 (59 000–118 000)	90 000 (50 000–119 000)	62 000 (27 000–91 000)	55 000 (21 600–89 600)
	School	1 700 (1 500–1 900)	900 (600–1 100)	800 (500–1 100)	600 (300–800)	600 (61–800)
	Work	182 000 (169 000–199 000)	100 000 (65 000–130 000)	100 000 (62 000–122 000)	68 000 (34 000–98 000)	66 000 (25 000–96 100)
40%	Total	395 000 (366 000–416 000)	274 000 (228 000–318 000)	272 000 (226 000–313 000)	213 000 (158 000–269 000)	207 000 (141 000–259 000)
	Home Community	196 000 (181 000–206 000)	134 000 (106 000–153 000)	132 000 (105 000–153 000)	102 000 (75 000–128 000)	99 000 (66 000–123 000)
	School	1 700 (1 500–2 000)	1 200 (900–1 500)	1 200 (1 100–1 400)	1 000 (800–1 300)	1 000 (600–1 200)
	Work	197 000 (183 000–208 000)	139 000 (121 000–163 000)	139 000 (120 000–159 000)	110 000 (82 000–140 000)	107 000 (75 000–135 000)
50%	Total	409 000 (389 000–430 000)	334 000 (299 000–366 000)	333 000 (296 000–363 000)	281 000 (227 000–324 000)	277 000 (226 000–323 000)
	Home Community	203 000 (192 000–212 000)	162 000 (147 000–180 000)	163 000 (146 000–177 000)	136 000 (110 000–160 000)	134 000 (102 000–157 000)
	School	1 800 (1 600–2 100)	1 500 (1 400–1 700)	1 400 (1 200–1 700)	1 300 (1 100–1 600)	1 200 (1 100–1 500)
	Work	204 000 (195 000–216 000)	170 000 (151 000–184 000)	169 000 (149 000–184 000)	144 000 (116 000–162 000)	142 000 (123 000–165 000)

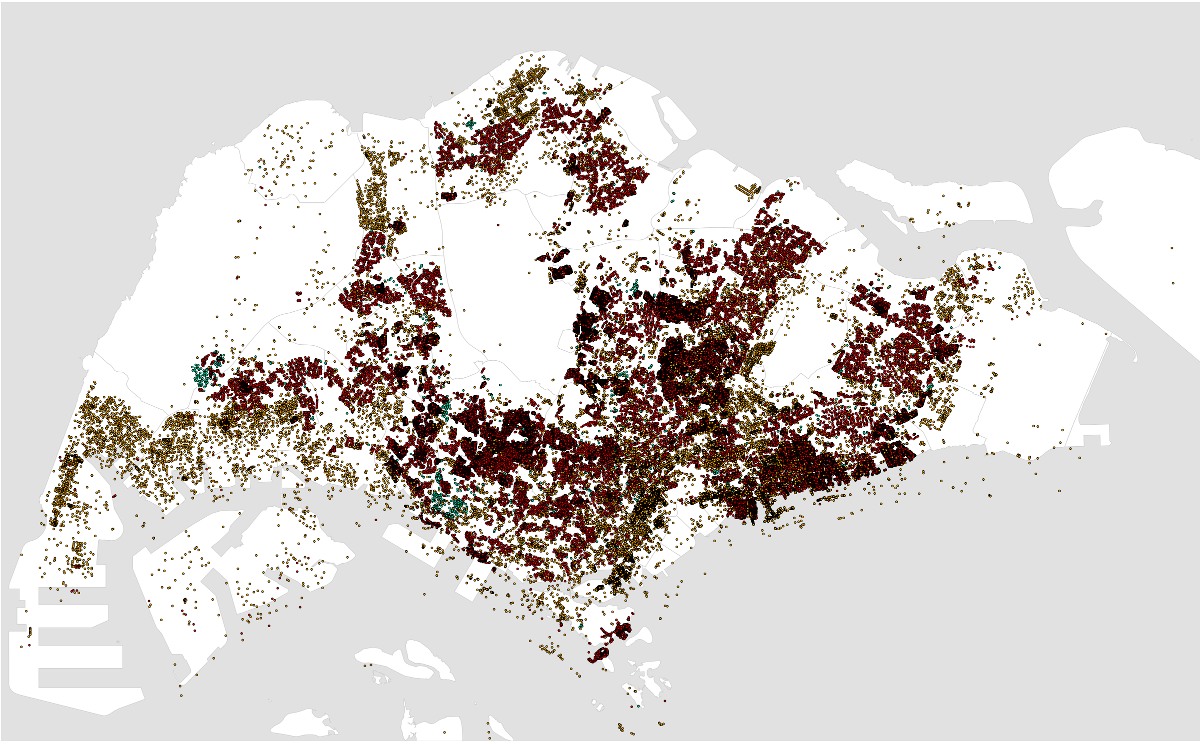
Supplementary Table 3. The cumulative number of infections at day 80 with higher asymptomatic proportions of 30%, 40% and 50% by location and intervention (quarantine, school closure, work distancing and a combination of all three) for the median simulation at R_0 of 1.5. The IQR is presented in brackets. Due to the stochasticity within each simulation, numbers less than 20 indicate near-complete suppression, and should not be compared to assess which is more effective. All numbers above 100 have not been rounded. Those up to 10 000 have been rounded to the nearest hundred, and those above have been rounded to the nearest thousand. Some discrepancies will therefore exist in the summations.



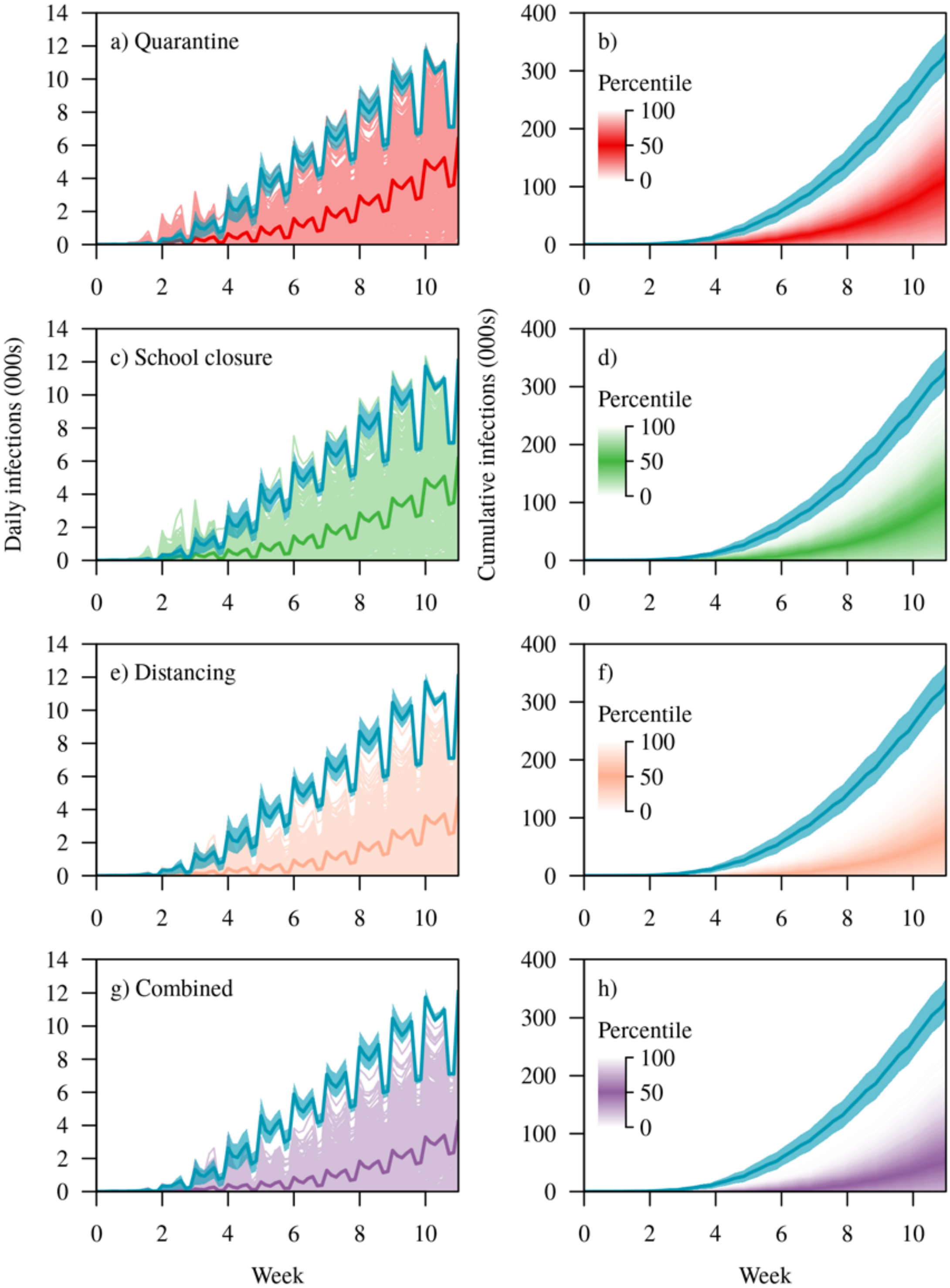
Supplementary Figure 1. The total number of daily and cumulative infections at $R_0 = 1 \cdot 5$ for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in light blue for clarity.



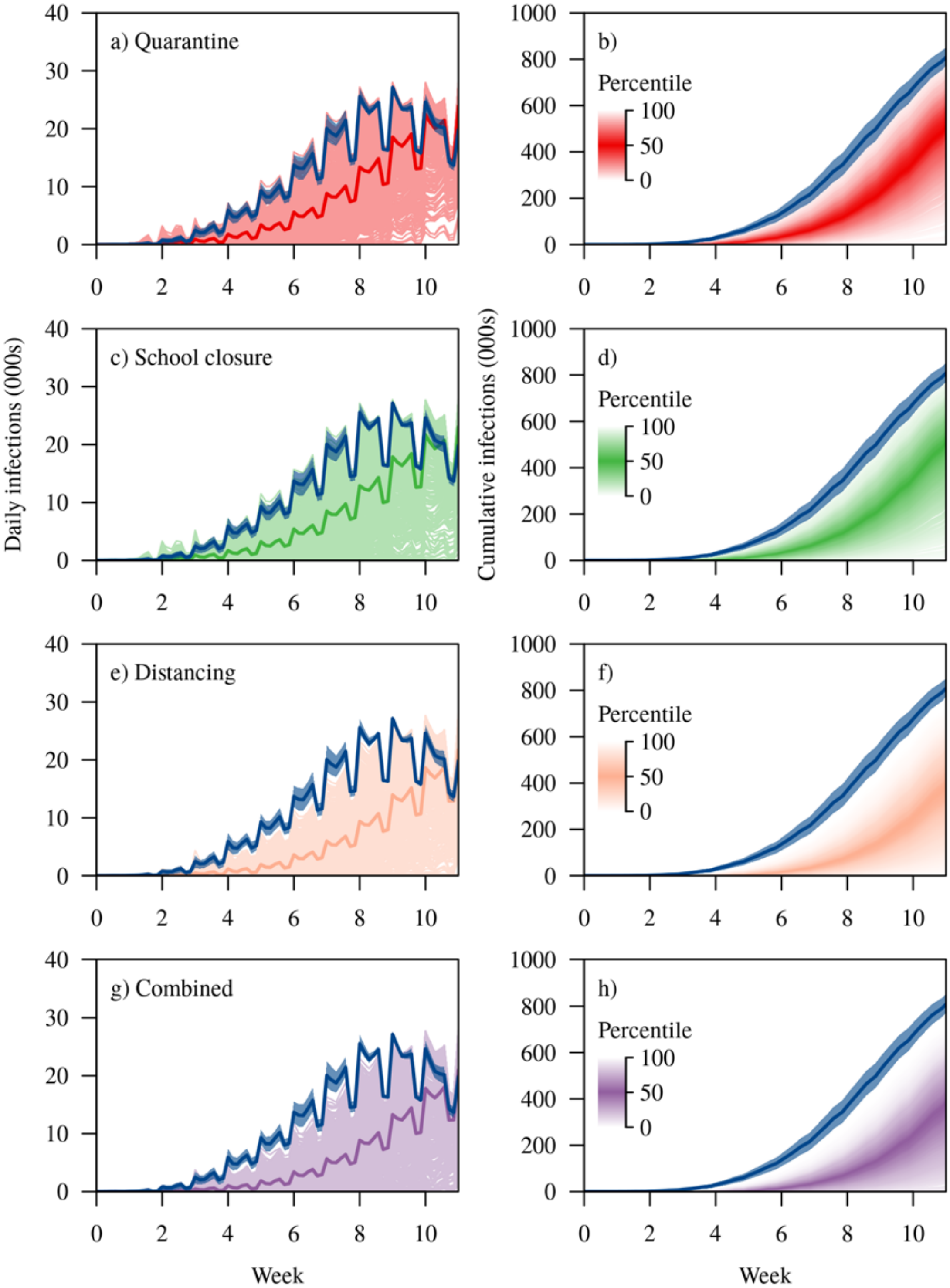
Supplementary Figure 2. The total number of daily and cumulative infections at $R_0 = 2 \cdot 5$ for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in black for clarity.



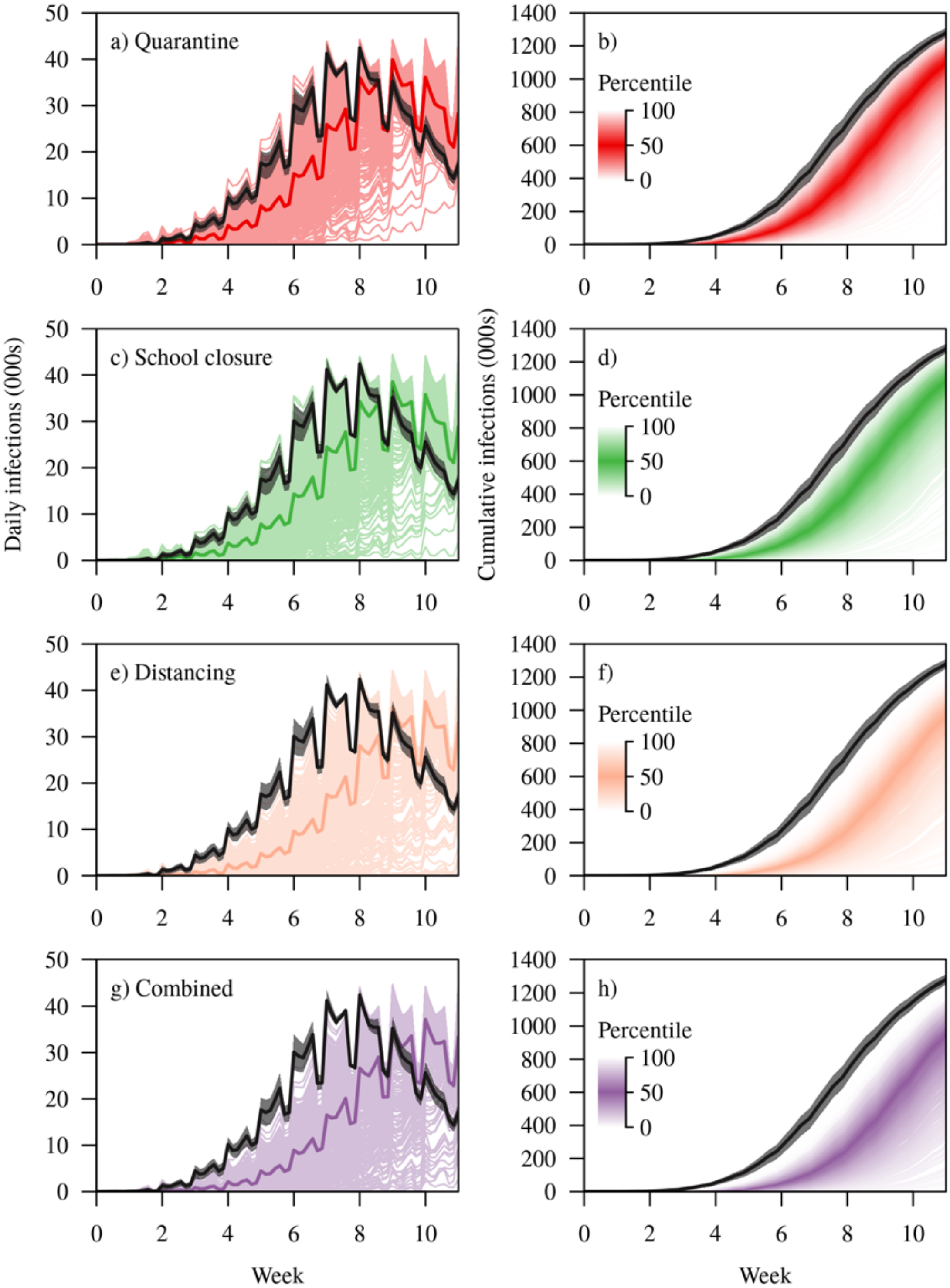
Supplementary Figure 3. Map of all household (red dots), workplace (orange) and school (blue) locations in Singapore used within the analysis.



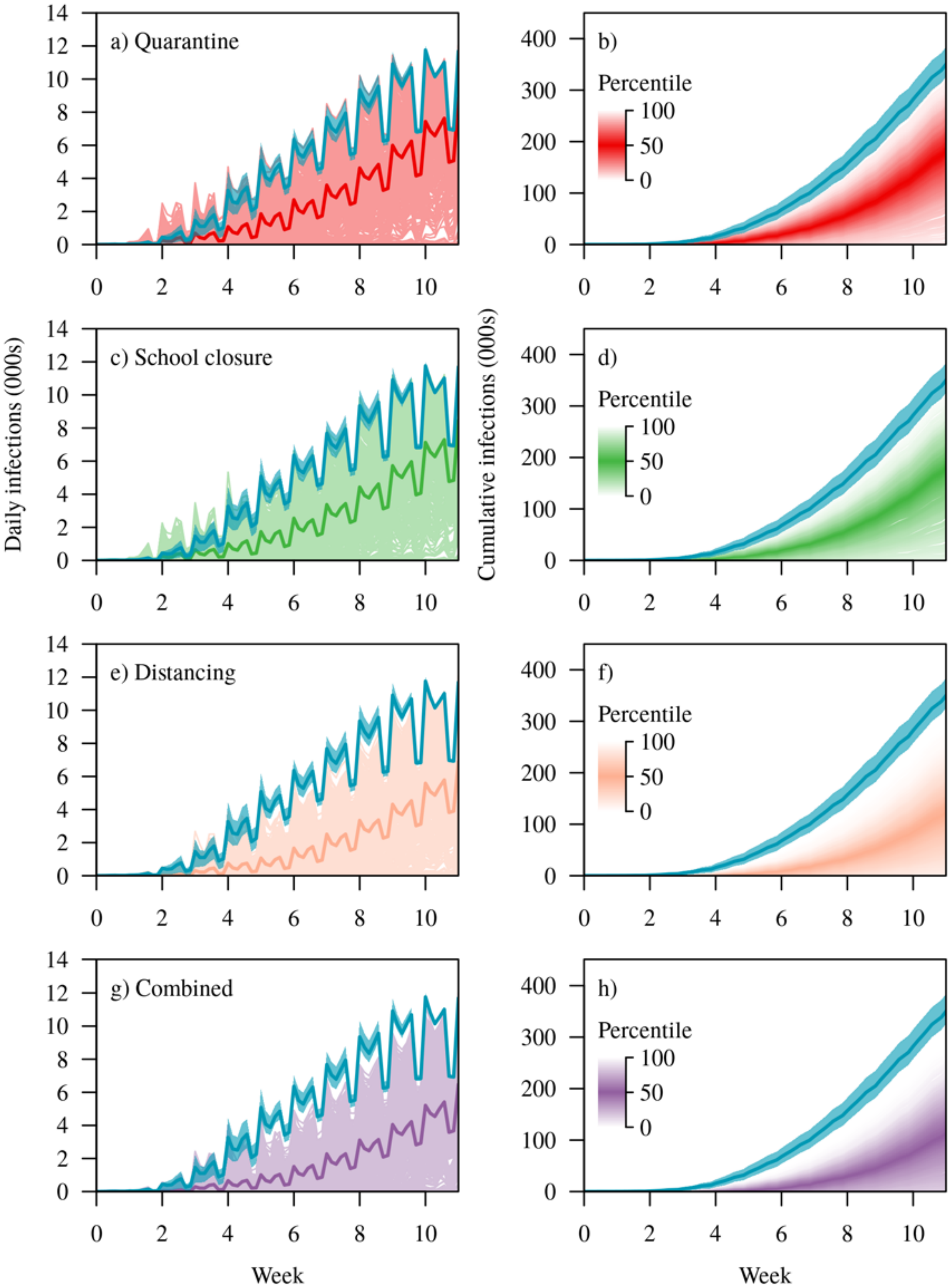
Supplementary Figure 4. The total number of daily and cumulative infections at $R_0 = 1 \cdot 5$ with a 22.7% asymptomatic proportion for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in light blue for clarity.



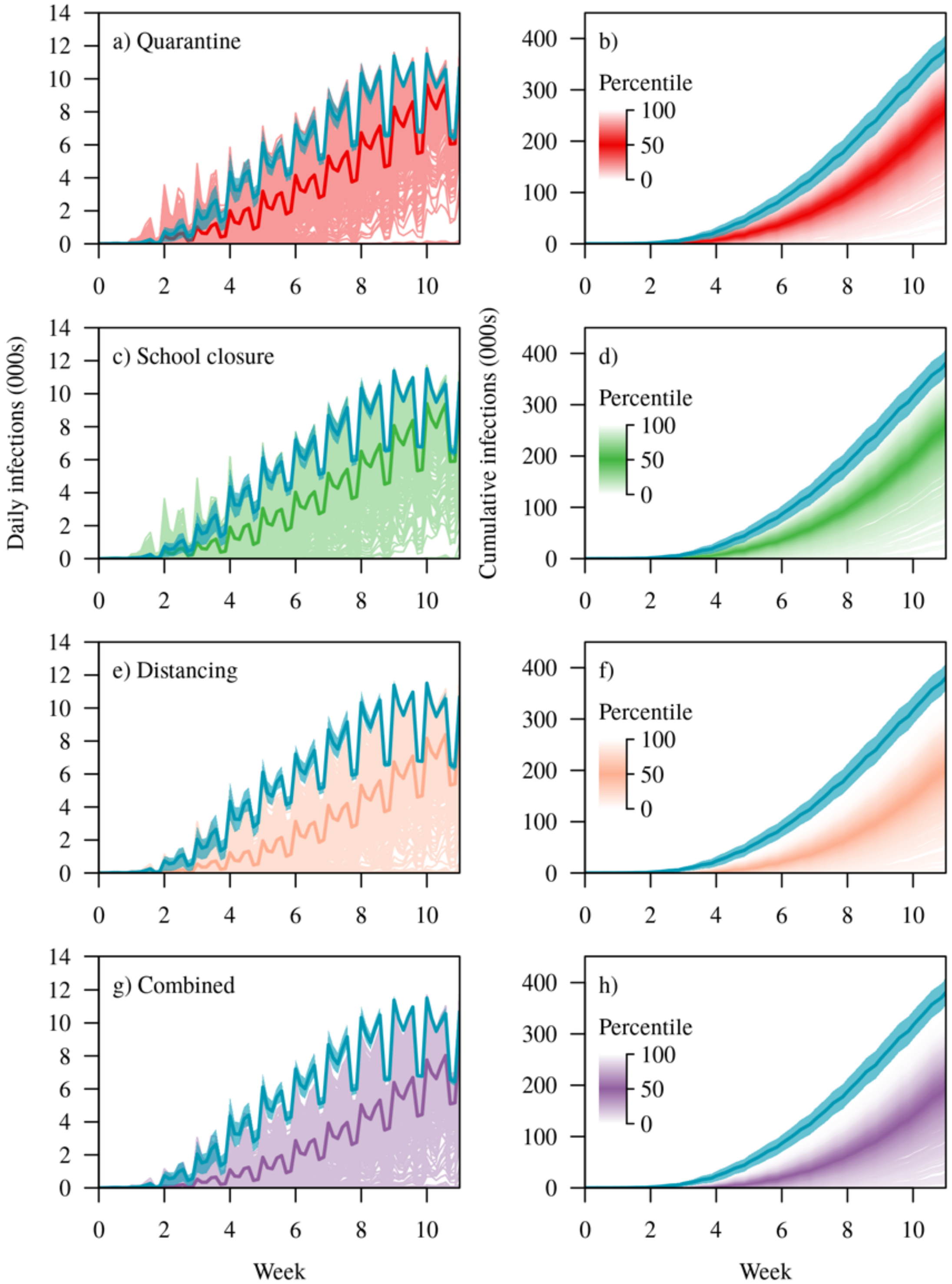
Supplementary Figure 5. The total number of daily and cumulative infections at $R_0 = 2 \cdot 0$ with a 22.7% asymptomatic proportion for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in dark blue for clarity.



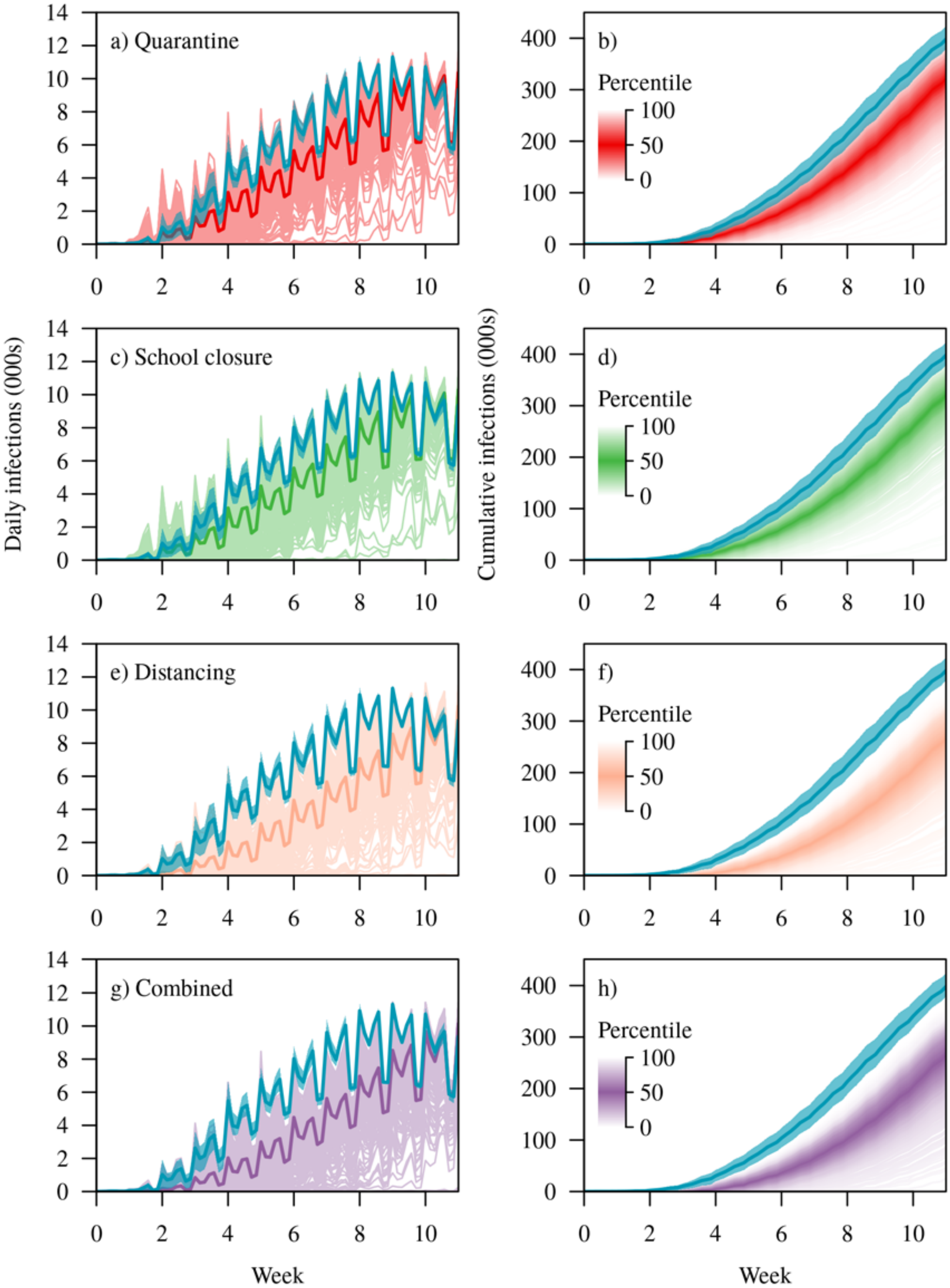
Supplementary Figure 6. The total number of daily and cumulative infections at $R_0 = 2 \cdot 5$ with a 22.7% asymptomatic proportion for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in black for clarity.



Supplementary Figure 7. The total number of daily and cumulative infections at $R_0 = 1 \cdot 5$ with a theoretical 30% asymptomatic proportion for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in light blue for clarity.



Supplementary Figure 8. The total number of daily and cumulative infections at $R_0 = 1 \cdot 5$ with a theoretical 40% asymptomatic proportion for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in light blue for clarity.



Supplementary Figure 9. The total number of daily and cumulative infections at $R_0 = 1 \cdot 5$ with a theoretical 50% asymptomatic proportion for 1000 simulations across 80 days with the median presented as a darker line for each panel. The number of daily infections is shown in panel a) with quarantining, c) school closure, e) workplace distancing, g) all three interventions combined, and corresponding cumulative infections in panel b), d), f) and h). The baseline no control strategy is presented as the IQR in light blue for clarity.

1. Department of Statistics, Singapore. *Singapore Census of Population 2010, Statistical Release 3: Geographic, Distribution and Transport*. https://www.singstat.gov.sg/publications/cop2010/census10_stat_release3.
2. Immigration and Checkpoints Authority, Singapore. *Report on registration of births and deaths 2010*. https://www.ica.gov.sg/docs/default-source/ica/stats/annual-bd-statistics/stats_2010_annual_rbd_report.pdf.
3. Department of Statistics, Singapore. *Statistics on Marriages and Divorces, 2011*. https://www.singstat.gov.sg/-/media/files/publications/population/smd2016.pdf.
4. Singapore Land Authority. *LandQuery*. https://www.onemap.sg/.
5. Land Transport Authority. *Land Transport Datamall*. https://www.mytransport.sg/content/mytransport/home/dataMall/.
6. Chakirov, A. & Erath, A. Use of Public Transport Smart Card Fare Payment Data for Travel Behaviour Analysis in Singapore. in *ETHzurich Research Collection* (2011).
7. Data.gov. *School Directory and Information*. https://data.gov.sg/dataset/school-directory-and-information.
8. The Green Book. *Industrial, Commercial and Consumer directory*. http://www.thegreenbook.com/.
9. ESRI. *ArcGIS Desktop: Release 10.6. Redlands, CA: Environmental Systems Research Institute.*
10. Data.gov. *Number of Classes and Class Size by Level*.
11. Willis, G. A. *et al.* The impact of influenza infection on young children, their family and the health care system. *Influenza Other Respi Viruses* **13**, 18–27 (2019).
12. Quazi, H. *Presenteeism*. (Palgrave Macmillan UK, 2013). doi:10.1057/9781137275677.
13. Chao, D. L., Halloran, M. E., Obenchain, V. J. & Jr, I. M. L. FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation Model. *PLOS Computational Biology* **6**, e1000656 (2010).
14. McBryde, E. S. *et al.* Bayesian modelling of an epidemic of severe acute respiratory syndrome. *Bulletin of Mathematical Biology* **68**, 889–917 (2006).
15. Wilder-Smith, A. *et al.* Asymptomatic SARS Coronavirus Infection among Healthcare Workers, Singapore. *Emerg. Infect. Dis.* **11**, 1142–1145 (2005).
16. Meltzer, M. I. Multiple Contact Dates and SARS Incubation Periods. *Emerg. Infect. Dis.* **10**, 207–209 (2004).
17. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med* NEJMoa2001316 (2020) doi:10.1056/NEJMoa2001316.
18. Nishiura, H. *et al. Estimation of the asymptomatic ratio of novel coronavirus (2019-nCoV) infections among passengers on evacuation flights*. http://medrxiv.org/lookup/doi/10.1101/2020.02.03.20020248 (2020) doi:10.1101/2020.02.03.20020248.