Electronic Supplementary Material:

The illusion of personal health decisions for infectious disease management: disease spread in social contact networks

Matthew Michalska-Smith, Eva A Enns, Lauren A White, Marie L J Gilbertson & Meggan E Craft

Contents

| | 1 |
|---|--------------------|
| S1 Network data sources | 2 |
| S2 Locality network structure comparison | 3 |
| S3 Alternative background transmission S3.1 No background transmission ($\beta_{background} = 0$) | 3 3 8 |
| S4 Epidemiological outcome quantification | 12 |
| S5 Additional outcome metrics for ($\beta_{background} = 0.001$) | 14 |
| S6 Simulation code accessibility details | 17 |
| S7 Age-structure weighting of interaction networks S7.1 Age-structured simulation results | 17 18 |
| S8 Link removal irrespective of (proximate) vulnerability | 22 |

¹ S1 Network data sources

To construct our plausible human contact networks, we relied on a range of publicly available data sources. 2 For household sizes, we used state-specific data for Florida [S1] and Texas [S2], for the latter, we were 3 only able to obtain accurate numbers for households with three or fewer occupants, so higher values were 4 extrapolated from the overall average occupancy [S3]. Individual vulnerability was assigned based on age-5 specific COVID-19 induced mortality rates [S4]. To construct the underlying age distribution, we used 6 United States American Community Survey data from 2019, specifically publications DP05: Demographic 7 and Housing Estimates [S5]. Each individual in the network was probabilistically assigned into one of thirteen 8 (less than 5 years old, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75-84, or greater 9 than 84 years old) age classes based on state-specific age distributions. The distribution of ages across the 10 network was modified from an initial random allocation to prevent the occurrence of households in which all 11 individuals were children. 12

The school layer was constructed by taking all school-age individuals (5-9, 10-14, or 15-19 years old) and assigning them into age-class-specific classrooms of approximately 20 students per classroom [S6]. Within each classroom, networks were fully connected: all students had the same interaction strength with all other students in the classroom.

The workplace layer similarly considered all working-age adults (20-24, 25-34, 35-44, 45-54, 55-59, or 60-64 years old), subtracted a percentage (10%) based on unemployment rates in the spring-summer of 2020 [S7], and assigned the remainder to workplaces whose size was loosely based on the distribution of workplace sizes in the United States [S8]. This latter distribution was modified to remove especially small (less than 5 workers) and large (greater than 100 workers) work places. Within workplaces, as with classrooms, networks are fully connected: all workers have an equal level of contact with all other workers at the same workplace.



²³ S2 Locality network structure comparison

Figure S1: Household size (left) and age (right) distributions for each of the two locales used in the main text. *N.b.* left and right plots have different vertical axis limits. While Florida tends to have older citizens, Texas tends to have larger families (and consequently more within-household interactions; see also table 1). Asterisks denote values for Texas household size that were inferred to match an overall mean household size in the absence of precise data.

²⁴ S3 Alternative background transmission

In this section, we repeat the analyses from the main text, but with alternative rates of background trans-25 mission. In particular, we consider the case of no background transmission, in which there are only three 26 layers of interactions in the network (within-household, classmate, and co-worker), and a case where the 27 background transmission is fifty times as high, but otherwise structured identically as in the main text. In 28 the absence of background transmission results are similar to those in the main text, though note that in 29 the "Vulnerable Households Avoid Work/School" risk-tolerance regime, each household containing a vulner-30 able individual is completely isolated from the rest of the network, meaning that the only way a member 31 could become infected is if they or a fellow member of their household were randomly chosen as the initial 32 case. When background transmission is sufficiently high, on the other hand, most effects of risk-tolerance 33 regime are washed out, as the classmate and co-worker interaction layers are relatively less important when 34 a pathogen spreads readily by other means. 35

³⁶ S3.1 No background transmission ($\beta_{background} = 0$)



Figure S2: As fig. 1 in the main text, but with no background transmission and for more epidemiological outcome measures (columns; defined in section S4). Individuals in the network either: did not change behaviour in response to (contact with individuals with) vulnerability status (top), changed behaviour if they were vulnerable themselves (middle), or changed behaviour when a member of their household was vulnerable (bottom). Values were then log-scaled and normalized for each epidemic outcome such that the maximum value is 1 (yellow) and the minimum value is 0 (purple). Multiple points within each hexagon were averaged to produce the plotted value. Each panel consists of a heatmap showing the relative epidemic outcome of simulations spanning various levels of co-worker (vertical axis) and classmate (horizontal axis) transmission.



Figure S3: As fig. 2 in the main text, but with no background transmission and for more epidemiological outcome measures (columns; defined in section S4). Only simulations with no difference in behaviour based on vulnerability and only outcomes from epidemics resulting in greater than 5% of the total population being infected are shown.



point colours distinguish age and household size distribution locales (as in fig. 2). Vertical lines extending beyond the point extents indicate 95% confidence intervals for the slope estimates (some confidence intervals are obscured by the points). To ease interpretation, lines connect coefficient values across interaction types for results from models of the same risk-tolerance regime and locale. Points are slightly offset horizontally to reduce overlap. Only outcomes from epidemics resulting in greater than 5% of the total population being infected were included Figure S4: As fig. 3 in the main text, but with no background transmission and for more epidemiological outcome measures (columns; defined in section S4). The vertical axis indicates the value of the best-fitting coefficient for each transmission rate in a linear model fit to simulation output. Facets distinguish between epidemic outcome measures, point shapes distinguish risk-tolerance regimes (*i.e.* rows in fig. 1), and in the linear models.



Figure S5: As the leftmost column in fig. 1, showing the reproduction number across parametrizations, but with no background transmission.



³⁷ S3.2 High background transmission ($\beta_{background} = 0.05$)

Figure S6: As figs. 1 and S2, but with background transmission set to $\beta_{background} = 0.05/N$. Values are scaled within each outcome measure such that the lowest value is 0 (purple) and the highest is 1 (yellow).











Figure S9: As the leftmost column in fig. 1 and as fig. S5, showing the reproduction number across parametrizations, but with background transmission set to $\beta_{background} = 0.05/N$.

³⁸ S4 Epidemiological outcome quantification

³⁹ We quantify epidemiological outcomes according to:

40 Total Infected The total number of nodes (individuals) infected over the course of a simulated epidemic

Average Infectious The average number of infectious individuals across time (Total Infected / epidemic
 duration)

43 Total Died The total number of individuals that die over the course of the epidemic

44 Time to Peak The timepoint in which the most people are concurrently infectious

45 Time to Peak Death The timepoint in which the most individuals die over the course of the epidemic

46 Maximum Infectious The highest number of concurrently infectious individuals

47 Vulnerable Infected The number of vulnerable individuals infected over the course of the epidemic

48 Time to Vulnerable Infected The timepoint in which the first vulnerable individual is infected

49 Percent Vulnerable Households Infected The percent of households with at least one vulnerable indi vidual that have at least one member infected over the course of the epidemic

⁵¹ Epidemic Rate The average number of infections per timepoint until the peak is reached (Maximum

52 Infectious / Time to Peak)



Figure S10: Correlogram of epidemiological outcomes. Points, correlation coefficients are coloured according to risk-tolerance scenario: No Difference in Behaviour (I, red), Vulnerable Individuals Avoid Work/School (II, purple), and Vulnerable Households Avoid Work/School (III, blue).



S5 Additional outcome metrics for $(\beta_{background} = 0.001)$

Figure S11: As fig. 1, but showing results for all epidemiological outcome measures, as in figs. S2 and S6. Values are scaled within each outcome measure such that the lowest value is 0 (purple) and the highest is 1 (yellow). Note that some metrics related to peak timings tend to be higher in the household-link-reduction scenario (while the other two are largely similar, as in other cases). This is because peak timing depends both on the rate of disease spread and on the size of the peak. The latter tends to be smaller in this scenario, and is consequently reached faster, even with a smaller "Epidemic Rate".









54 S6 Simulation code accessibility details

All simulations were conducted in C++ version 8.1.0, with data manipulation and plotting done in R version
4.2.0 [S9], with the use of R packages: assertthat [S10], ggbeeswarm [S11], kableExtra [S12], patchwork [S13],
Rcpp [S14, S15, S16], tidygraph [S17], tidyverse [S18], and scales [S19].

An application for visualizing our synthetic community network structure and simulating disease spread (including the manipulation of disease parameters) is available online: https://z.umn.edu/LINCS [S20]. Code to replicate all aspects of these analyses is available online: https://github.com/mjsmith037/ Layered_Interactions_COVID_Model.

⁶² S7 Age-structure weighting of interaction networks

In this section, we explore the sensitivity of our results to the assumption of uniform, fully-connected networks. Specifically, we construct layered interaction networks in which each layer's interactions are weighted according to the age-class (16 5-year classes for individuals 0-80 years old) of the two individuals involved. To do this, we utilize published information on social interaction age-structure in each of our four network layers: home, work, school, and others [S21].

Practically, each layer's transmission rates are now constrained by a "maximum transmission rate" which is then multiplied by symmetrized, normalized age-structured contact matrices from [S21], rather than a constant rate for all interactions. Household and co-worker interactions are weighted (using the "home" and "work" age-structured interaction matrices, respectively), but otherwise constructed as for the main text.

For the classmate interaction layer, we shift from modelling individual, age-class-homogenous classrooms are now modelled at the school level (rather than classroom, as classrooms were previously constructed of a single age-class), with internal school interactions being weighted according to the ages of each individual using the "school" interactions layer from [S21]. Each school is assembled to hold approximately 500 students, in line with the average school size in the United States [S22].

This means that the number of classmate interactions is substantially higher in these simulations than in those presented in the main text. While the new school configuration has the advantage of considering interactions between students in different classrooms, it lacks the fragmentation typical of the internal structure of American public schools.

Background interactions differ from the main text simulations in that they are explicitly modelled as a fully-connected, weighted interaction network in which the strength of interaction between each pair of

| Metric | "United States" mean $(sd)^1$ |
|--|-------------------------------|
| Number of individuals | 998 |
| Number school-age | 193(12.37) |
| Number employed | 528 (15.8) |
| Number vulnerable | 17 (4.08) |
| Number of households | 411 |
| Number households with children | 170(8.64) |
| Number of households with vulnerable | 17(3.98) |
| Total number of edges (no isolation) | 24723 (2312.91) |
| Household edges | 1101 |
| Classmate edges | $18664\ (2388.91)$ |
| Co-worker edges | 4958 (407.33) |
| Edges when vulnerable isolate | 24646 (2314.37) |
| Edges when vulnerable households isolate | 23051 (2285.19) |

Table S1: Summary statistics for networks generated utilizing data on age-structure of interactions.

¹ Values are presented with both mean and standard deviation except when there was no variance, in which case the constant value is presented.

² "Background" transmission links are omitted from this (and other edge) count(s). Because they connect every individual to every other, there are always N(N-1)/2 such edges, where N is the number of individuals in the network.

individuals depends on the ages of the two participating individuals. The particular transmission rates are the product of a baseline "maximum background transmission rate" ($\beta_{background}^{max} = 0.0001$) multiplied by a scaling factor from the "others" interactions layer from [S21].

There are two main limitations of the age-structure data from [S21] for our analyses. First, that it is resolved at the country-level, so we are unable to replicate the locale-comparison (fig. 2) from the main text. Rather, in these analyses, we utilize United States level census data on household sizes [S2] and age [S5] distributions.

⁹⁰ Second, the provided interaction matrices only consider individuals up to 80 years old, while, for COVID-

⁹¹ 19, infection fatality ratios (IFR) increase substantially for individuals over 80 [S23]. In order to reconcile

these data in our new simulations, we treat all individuals over 75 as being like the 75-80 years old individuals

⁹³ in Prem et al. [S21] and average IFR for individuals over 75 from O'Driscoll et al. [S23].

⁹⁴ S7.1 Age-structured simulation results



Figure S14: As fig. 1 in the main text, but using age-structured interaction network structure. Relative epidemic outcomes are indicated by the columns. Individuals in the network either (rows): did not change behaviour in response to (contact with individuals with) vulnerability status (top), changed behaviour if they were vulnerable themselves (middle), or changed behaviour when a member of their household was vulnerable (bottom). Multiple points within each hexagon were averaged to produce the plotted value. Mean values were then log-scaled and normalized for each epidemic outcome such that the maximum value is 1 (yellow) and the minimum value is 0 (purple). Each panel consists of a heatmap showing the relative epidemic outcome of simulations spanning various levels of co-worker (vertical axis) and classmate (horizontal axis) transmission.



rates on epidemic outcomes. The vertical axis indicates the value of the best-fitting coefficient for each transmission rate in a linear model of the form $y \sim m_1 log(\beta_{classmate}^{max}) + m_2 log(\beta_{co-worker}^{max}) + b$, where y indicates an epidemic outcome measure, m_n is a fitted slope coefficient, β_x^{max} represents the maximum transmission rate for interaction type x, which is then scaled according to interaction age-structure, and b is a fitted intercept coefficient. The horizontal axis distinguishes between the two coefficients (m_1 : "classmate", or m_2 : "co-worker"). Facets distinguish between epidemic outcome measures, point shapes distinguish risk-tolerance regimes (*i.e.* rows in fig. 1), and point colours distinguish age and household size distribution locales (as in fig. 2). Vertical lines extending beyond the point extents indicate 95% confidence intervals for the slope estimates (most confidence intervals are obscured by the points). To ease interpretation, lines connect coefficient values across interaction Figure S15: As fig. 3 in the main text, but using age-structured interaction network structure. Quantifying the effect of changes in transmission types for results from models of the same risk-tolerance regime and locale. Points are slightly offset horizontally to reduce overlap. Only simulations resulting in at least 5% of the population becoming infected were considered.



Figure S16: As the leftmost column in fig. 1 and as figs. S5 and S9, showing the reproduction number across parametrizations, but using age-structured interaction network structure.



⁹⁵ S8 Link removal irrespective of (proximate) vulnerability

Figure S17: As fig. 1 in the main text, but rather than removing links in association with an individual's contact with vulnerable individuals, the same number of links were removed randomly from the same interaction types as in fig. 1. Epidemiological outcome measures (columns) were scaled to be between 0 (purple) and 1 (yellow).

96 References

- S1. Miami Matters. 2022 Demographics. 2022. Available from: https://www.miamidadematters.org/
 demographicdata?id=12§ionId=936#sectionPiece_65
- S2. U.S. Census Bureau. S2501: Occupancy Characteristics. American Community Survey 1-Year Esti mates. 2019. Available from: https://data.census.gov/cedsci/table?q=texas%20household%
 20size%20distribution&g=0100000US_0400000US12&tid=ACSST1Y2019.S2501
- S3. Texas Demographic Center. Texas Household Trends and Projections, 2010-2050. 2021. Available from:
 https://demographics.texas.gov/Resources/publications/2021/20210415_HouseholdsTrendsProj_
 Brief.pdf
- S4. Centers for Disease Control & Prevention. Morbidity and Mortality Weekly Report, 18 March 2020.
 2020. Available from: http://www.ecie.com.ar/images/paginas/COVID-19/4MMWR-Severe_
- 107Outcomes _ Among _ Patients _ with _ Coronavirus _ Disease _ 2019 _ COVID 19 United _ States _108February _ 12-March _ 16 _ 2020.pdf
- S5. U.S. Census Bureau. DP05: ACS Demographic and Housing Estimates. American Community Survey
 1-Year Estimates. 2019. Available from: https://data.census.gov/cedsci/table?q=age%
 20distribution&g=0100000US%240400000&tid=ACSDP1Y2019.DP05
- S6. U.S. National Center for Education Statistics. Average class size in public schools, by class type and
 state: 2017-18. 2018. Available from: https://nces.ed.gov/surveys/ntps/tables/ntps1718_
 fltable06_t1s.asp
- S7. U.S. Bureau of Labor Statistics. Seasonally Adjusted Unemployment Rate. 2020. Available from:
 https://www.bls.gov/data/#unemployment
- S8. North American Industry Classification System Association. U.S. Business Firmographics—Company
 Size. 2020. Available from: https://www.naics.com/business-lists/counts-by-company-size/
- S9. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical
 Computing. Vienna, Austria, 2022. Available from: https://www.R-project.org/
- S10. Wickham H. assertthat: Easy Pre and Post Assertions. R package version 0.2.1. 2019. Available from:
 https://CRAN.R-project.org/package=assertthat
- S11. Clarke E and Sherrill-Mix S. ggbeeswarm: Categorical Scatter (Violin Point) Plots. R package version
 0.6.0. 2017. Available from: https://CRAN.R-project.org/package=ggbeeswarm

S12. Zhu H. kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4. 125

2021. Available from: https://CRAN.R-project.org/package=kableExtra 126

- S13. Pedersen TL. patchwork: The Composer of Plots. R package version 1.1.1. 2020. Available from: 127 https://CRAN.R-project.org/package=patchwork 128
- S14. Eddelbuettel D and François R. Rcpp: Seamless R and C++ Integration. J. Stat. Softw. 2011; 40 (8) 129 :1-18. DOI: 10.18637/jss.v040.i08 130
- S15. Eddelbuettel D. Seamless R and C++ Integration with Rcpp. ISBN 978-1-4614-6867-7. New York: 131 Springer, 2013. DOI: 10.1007/978-1-4614-6868-4 132
- S16. Eddelbuettel D and Balamuta JJ. Extending R with C++: A Brief Introduction to Rcpp. The Am. 133 Stat. 2018; 72 (1) :28-36. DOI: 10.1080/00031305.2017.1375990 134
- S17. Pedersen TL. tidygraph: A Tidy API for Graph Manipulation. R package version 1.2.1. 2022. Available 135 from: https://CRAN.R-project.org/package=tidygraph 136
- S18. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry 137
- L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, 138
- Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, and Yutani H. Welcome to the tidyverse. J. 139 Open Source Softw. 2019; 4 (43) :1686. DOI: 10.21105/joss.01686 140
- S19. Wickham H and Seidel D. scales: Scale Functions for Visualization. R package version 1.2.0. 2022. 141 Available from: https://CRAN.R-project.org/package=scales 142
- S20. Michalska-Smith M, White L, Gilbertson M, and Craft M. Layered Interaction Network COVID-19 143 Simulator. 2021. Available from: https://z.umn.edu/LINCS 144
- Prem K, Zandvoort Kv, Klepac P, Eggo RM, Davies NG, Mathematical Modelling of Infectious Dis-S21. 145 eases COVID-19 Working Group C for the, Cook AR, and Jit M. Projecting contact matrices in 177 146 geographical regions: an update and comparison with empirical data for the COVID-19 era. PLoS 147 computational biology 2021; 17 (7) :e1009098. DOI: 10.1371/journal.pcbi.1009098
- Public School Review. Average Public School Student Size. 2023. Available from: https://www. S22. 149
- publicschoolreview.com/average-school-size-stats/national-data 150
- S23. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DA, Azman AS, Paireau J, Fontanet A, 151 Cauchemez S, and Salje H. Age-specific mortality and immunity patterns of SARS-CoV-2. Nature 152
- 2021; 590 (7844) :140-5. DOI: 10.1038/s41586-020-2918-0 153

148