Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# A Calibration Framework for Bridging Household-Structured Microscopic and Macroscopic Epidemic Models

**Rex Ndubuisi Ali**

alirex@engineering.upenn.edu

University of Pennsylvania

**Nagesh Subbanna**

Amrita Vishwa Vidyapeetham

**Saswati Sarkar**

University of Pennsylvania

**Article**

**Keywords:** Household transmission dynamics, Deterministic approximation of stochastic models, Microscopic, macroscopic modeling, Model calibration, B-spline regression, Interpolation and Extrapolation

**Posted Date:** November 2nd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11119711/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# A Calibration Framework for Bridging Household-Structured Microscopic and Macroscopic Epidemic Models

Rex Ndubuisi Ali[1*], Nagesh Subbanna[2] and Saswati Sarkar[1]

[1*]Department of Electrical and Systems Engineering, University of Pennsylvania, 200 South 33rd Street, Philadelphia, 19104, Pennsylvania, USA.

[2]School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Bengaluru, India.

*Corresponding author(s). E-mail(s): alirex@engineering.upenn.edu;
Contributing authors: nageshks@am.amrita.edu;
swati@engineering.upenn.edu;

**Abstract**

Modeling the spread of epidemics has had two complementary paradigms: 1) a macroscopic paradigm which divides the populace into few groups of large number of statistically identical individuals and 2) a microscopic paradigm which considers each individual as a separate entity with his own contact network. The macroscopic paradigm achieves fast computation through the solution of ordinary differential equations, but cannot capture essential features of human society such as household structures. In contrast, the microscopic paradigm can capture arbitrary heterogeneities, but its computational cost can become prohibitive as population size increases. This study develops a systematic framework for bridging these paradigms while retaining the computational advantages of the macroscopic and the accuracy of the microscopic paradigm. To the best of our knowledge, this direction of research is novel. We consider a populace comprising adults and children organized into households of different sizes. Members of the same household have different contact patterns and are therefore not statistically identical. Contact rates within households are much higher than those across households, causing clustered epidemic spread. We first demonstrate that the most natural macroscopic formulation can substantially misrepresent the epidemic trajectory. We subsequently show that a specifically designed macroscopic formulation, which captures clustered household spread by calibrating

parent–child interactions through a calibration parameter, can closely approximate the microscopic model. The optimum value of this parameter is obtained by minimizing the discrepancy between the microscopic and macroscopic epidemic trajectories. This discrepancy at the optimum value is small, and it can be predicted well through standard statistical techniques applied on few exact computations. The framework significantly reduces computation time while retaining accuracy of the microscopic, and provides a foundation for obtaining the same for arbitrary social networks.

**Keywords:** Household transmission dynamics, Deterministic approximation of stochastic models, Microscopic–macroscopic modeling, Model calibration, B-spline regression, Interpolation and Extrapolation

# Introduction

Mathematical modeling of infectious diseases is widely used to understand epidemic dynamics and to inform public health policy and intervention strategies. There are two distinct modeling paradigms. One is a macroscopic approach in which the stochastic evolution of the disease process is modeled as a system of deterministic ordinary differential equations (ODEs). The first such macroscopic model is the classical SIR model introduced by Kermack and McKendrick [1] as a special case of their more general theory. Since their inception, these deterministic frameworks have remained central to the development of epidemiological theory and the evaluation of intervention strategies [2, 3], and have been successfully applied to a wide range of infectious diseases [4–9]. This is because ODEs are analytically tractable, easy to compute and as the population size increases to infinity, the temporal evolution of the stochastically varying disease states can be shown to converge to the solution of the ODEs through applications of law of the large numbers under some natural assumptions on the stochastic process [5, 6]. The ease of computation renders them particularly useful for time-sensitive policy decisions during outbreaks [4].

However, the analytical convergence results for the above macroscopic model typically assume homogeneous mixing, meaning that every individual has an equal probability of interacting with every other individual. But in reality individuals exhibit different contact patterns depending on their geographical locations, age, profession, marital and parental status. Consequently, metapopulation models have generalized the classical framework by partitioning the population into multiple interacting subpopulations based on the above characteristics [5, 6, 10, 11]. Individuals within each subgroup are assumed to be statistically identical and there exists a contact rate matrix specifying the contact rate between any two individuals in the given subgroups: specifically the $ij$th entry of the matrix provides contact rates between an individual in subgroup $i$ and an individual in subgroup $j$. Typically, individuals in the same subgroup interact at a higher rate than those across subgroups [12–15]. Each subgroup is assumed to consist of a large number of individuals, then, the temporal evolution of the stochastically varying disease states can be shown to converge to the output of deterministic ODEs through applications of law of large numbers (specifically, as the

subpopulation sizes tend to infinity and the waiting times in each disease state are exponentially distributed, i.e., the transitions between states are Markovian, it has been shown using the law of large numbers that the fraction of individuals in each state at each time in the stochastic metapopulation process converges in probability to the solution of the corresponding deterministic ODE system [5, 6]).

The metapopulation generalization however does not fully capture the contact heterogeneity, specifically the heterogeneity introduced by households. For example, each individual typically interacts far more frequently with the members of their household than with individuals outside [14]. This suggests that households need to constitute the subpopulations of the metapopulation model. But the members of a household are rarely statistically identical. For example, a household may comprise two parents and children, then the children have different contact patterns with individuals outside the household as compared to the parents, and the parents and children cannot be considered statistically identical. In addition, the size of a household rarely exceeds single digits. For example, in 2025, only about 1.1% of households in the United States had seven or more members [16]. So, if a household is considered as a subgroup of a metapopulation model, the stochastic variation of the disease states need not converge to the output of the deterministic metapopulation ODEs. In summary, the limitation of the metapopulation generalization is that it can capture the evolution of disease in populations which can be divided into a few subgroups, each with a large number of statistically identical individuals. Target populations for disease spread rarely satisfy this criterion in practice.

The second paradigm is microscopic modeling in which each individual is represented explicitly as a distinct node in a graph which periodically establishes contact with a distinct set of nodes known as its contact set. The contact sets and contact rates of different nodes can be different, and no two nodes need to have statistically identical contact patterns. Such a contact network can easily capture households and other contact relations. However, Monte Carlo simulations rather than analytical computations are required to capture the spread of a disease. As the population size increases, the computational burden associated with repeated simulation, parameter estimation, sensitivity analysis, and policy evaluation becomes substantial. In practice, population sizes are large for typical target groups. For example, even in small communities such as cruise ships, which constitute fertile grounds for epidemic outbreak, total onboard occupancy including crew can be 8000-9000 [17]. Similarly, U.S. Census Bureau data show that the median American resident lives in a municipality with a population of about 17,500 [18], and approximately 79.6% of the U.S. population lives in urban areas, defined under the 2020 criteria as having at least 5,000 residents [19]. Consequently, simulating stochastic disease transmission in even such small target populations requires tracking millions of potential heterogeneous interactions evolving dynamically across time, households, occupational roles, and social groups. The computations often become very time consuming.

We seek a framework that combines the computational tractability of the macroscopic framework with the accuracy of representation of the microscopic framework. We consider target populations like small towns or cruise ships which comprise a diverse set of households, e.g., those with parent(s) and offsprings, and also those with adults who do not have children.

- We consider a macroscopic metapopulation model that divides the population into adults and children, without considering special relations within households. We consider another microscopic model that considers connections within each household and connections across households. The ODEs for the macroscopic model naturally yield to fast computation even for large size of the target populace, but do not often provide an accurate evolution of diseases. In fact, we show that, for the most natural choice of parameters, the evolution of the disease computed from these ODEs diverges substantially from the actual evolution. In contrast, Monte Carlo simulations for the microscopic model are computationally intensive but accurately capture the evolution of the disease. There is therefore an innate challenge in obtaining a framework that combines the computational tractability of the macroscopic framework with the accuracy of representation of the microscopic framework.
- We provide a principled approach for obtaining an equivalent macroscopic model, given a microscopic model, and therefore redresses the above innate challenge. The equivalence is in the sense that the collective disease statistics closely approximate those for the microscopic model. This retains the computational tractability of the macroscopic model and provides the accuracy of the microscopic model. To the best of our knowledge, this direction of research is novel for the spread of infection in human networks.
- Once the macroscopic equivalent is obtained for a few different contact rates for the microscopic one, macroscopic equivalents of microscopic models for other contact rates can be obtained through interpolation and extrapolation and application of rudimentary statistical techniques. Thus, intensive computations need to be done for a few instances, and accurate representations can be obtained through fast computations for a large number of instances.

We believe that this work lays the foundation for combining the best of microscopic and macroscopic worlds for larger target populaces with more complex contact patterns.

# Methods

We first describe the overall setup we target, outlining the background population structure. We then introduce the microscopic model, in which disease transmission is simulated through individual-level stochastic interactions. Next, we develop the corresponding macroscopic model, which replaces explicit household structure with effective interactions between demographic groups.

## Overall setup

We consider a community comprising households, ranging from those with parent(s) and children to those with one adult each. Each household has either two parents and their children, or single parents with children, or single adults. The contact network is defined by explicit within-household and across-household interactions. Members of the same household interact at a higher rate than those in different households.

Children interact only with their own parents and with other children, but not with other adults. All adults interact with each other, though rates are higher for adults in the same household as stated above. This setting captures interactions in close-knit communities, such as cruise ships, housing societies, etc. In both of the above cases, children mainly interact with other children and their parents, while adults mix more broadly across occupational and social groups. This can approximate interactions in small towns (except for example interactions between teachers and children).

We now describe the stochastic model for disease propagation. Each individual is either susceptible (state $S$), or infected (state $I$), or recovered (state $R$). Disease transmission is modeled as a stochastic process driven by interactions between individuals as described in the first paragraph. Thus, an infected individual (adult or child) can infect any other susceptible member of the same household, or members of other households he or she interacts with. Thus, an infected adult can infect susceptible adults of the same or different households, but can infect only his or her susceptible children but not unrelated susceptible children (i.e., children in different households). An infected child can infect his or her susceptible siblings, or unrelated susceptible children (i.e., in different households), but can infect only his or her parents but not other adults. The underlying connectivity graph and transmission pathways defining these permitted interactions are visually represented in Figure 1. An individual remains in the infected state for an exponentially distributed duration with rate $\gamma$ and subsequently recovers or dies. A recovered or dead individual cannot be infected again. Thus, recovered and dead individuals are effectively removed from the infection network, so we do not distinguish between the two states for the purposes of this paper. Each pair of individuals who can interact (i.e., a pair of adults, or a pair of children, or an adult-child pair in the same household) interact after exponentially distributed times. The exponential processes however have different parameters depending on if the pair is in the same household or is across households. Specifically, the exponential process has rate $\hat{\beta}$ for pairs across households, and $x\hat{\beta}$ where $x > 1$, for pairs in the same household. Here $x$ denotes the household amplification factor. Consequently, transmission events occur more frequently between members of the same household than between individuals connected only through community interactions. As depicted in Figure 1, the network structurally distinguishes these localized, high-rate intra-household contacts (solid lines) from the lower-rate global community connections (dashed lines). Also, contacts between any two individuals occur according to an exponential process whose rate depends on whether the individuals are adults, children, belong to the same household. For an individual, the time between successive contacts with other individuals is exponentially distributed. If the individual is an adult, the rate of this exponential process is $x\hat{\beta} \times$ (number of individuals in his or her household) + $\hat{\beta} \times$ (number of adults outside his or her household), if the individual is a child, the rate is $x\hat{\beta} \times$ (number of individuals in the same household) + $\hat{\beta} \times$ (number of children outside the household). These are referred to as contact rates of individuals. Note that the contact rates of adults and children in the same household are different, those of adults and children in different households are also different because household sizes are different. Thus, the individuals are not statistically

identical and individual heterogeneity is captured. The resulting stochastic process is a continuous-time Markov process. We will use different values of $x$ and $\hat{\beta}$ in our evaluation.

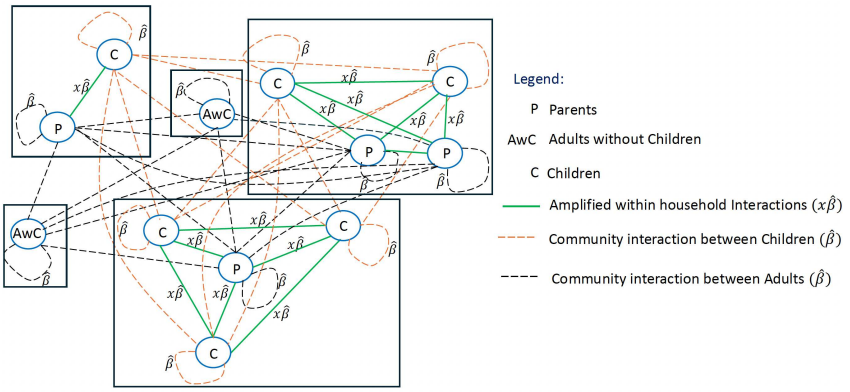



**Fig. 1** Stochastic infection network and household structure. The graph illustrates the heterogeneous population composition – including couple parents, single mothers, single fathers, adults without children, and children – and their corresponding interaction channels. Each rectangular box represents a household. High-frequency localized transmission within a household is governed by the amplified factor $x\hat{\beta}$ (solid lines), whereas background societal transmission across households is driven by the baseline community rate $\hat{\beta}$ (dashed lines).

## Microscopic Evaluation

The number of infected individuals in the given system over time can be evaluated through Monte Carlo simulation. This however consumes significant time, primarily because Monte Carlo simulations need to be averaged over multiple, e.g., 200 runs. We will describe the computational time in the next section.

## Macroscopic ODE Model

We now seek a macroscopic deterministic approximation of the total number of infected individuals as a function of time in the original system that is easy to compute. Toward that end, we first consider an intermediate alternative microscopic model which, on the one hand may provide the number of infected individuals close to that in the original system over time, for appropriate choice of parameters; and on the other hand, the intermediate system lends itself to an accurate computationally simple macroscopic representation. If we can obtain such an intermediate system, then its macroscopic representation will also provide us a good approximate macroscopic representation of the original system.

The intermediate system we consider is as follows. The population is partitioned into three groups: adults who are parents (say $P$), adults who are not parents (say $X$) and children (say $C$). Individuals in each group are statistically identical and interact with each other, say at rate $\hat{\beta}_1$. Individuals in group $P$ interact with those in

groups $C$ and $X$ say at rate $\hat{\beta}_2, \hat{\beta}_3$ respectively. However, children do not interact with adults who are not parents, that is, individuals in group $X$ do not interact with those in group $C$. Broadly, this system differs from the original system in that there is no household comprising one or two individuals from $P$ and $C$, such that the interaction within each household is at a much higher rate than interactions across households. But, still, can the number of infected individuals in the two systems be close as a function of time, for an appropriate choice of $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ given $\hat{\beta}, x$? It is well known that for any $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, the number of infected individuals in the intermediate system can be accurately computed as a function of time by solving a system of ordinary differential equations (ODEs) (when the transmission and recovery events are exponentially distributed). We will provide the ODEs and the well-known mathematical results we rely on subsequently.

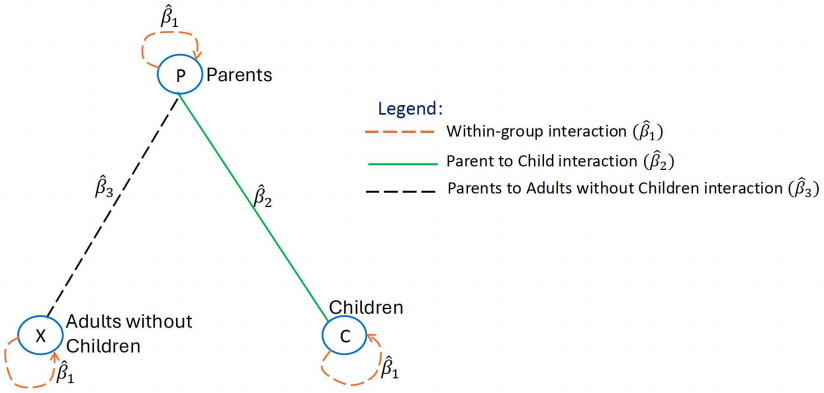



**Fig. 2** Intermediate interaction structure.

The stochastic model for disease propagation in the intermediate system is similar to that in the original system with groups replacing households. As in the original system, each individual is either susceptible (state $S$), or infected (state $I$), or recovered (state $R$). An infected individual (adult or child) can infect any other susceptible individual he or she interacts with. Thus, an infected parent can infect susceptible adults who are not parents and susceptible children. An infected child can infect a susceptible parent, but not a susceptible adult who is not a parent. Similarly, an infected adult who is not a parent can infect a susceptible parent but not a susceptible child. As in the original system, an individual remains in the infected state for an exponentially distributed duration with rate $\gamma$ and subsequently recovers or dies. A recovered or dead individual cannot be infected again. Thus, recovered and dead individuals are again effectively removed from the infection network. Each pair of individuals who can interact (i.e., a pair of adults, or a pair of children, or a parent-child pair) interact after exponentially distributed times. The exponential processes however have different parameters depending on if the pair is in the same group or is across groups. Specifically, the exponential process has rate $\hat{\beta}_1$ for pairs in the same group (i.e., if both are parents, both are children, both are adults who are not parents), and rate

$\{\hat{\beta}_2, \hat{\beta}_3\}$ if members in the pair are in different groups, that is, $\hat{\beta}_2$ for parent-child pair, $\hat{\beta}_3$ for parent-adult who is not parent pair.

The intermediate system differs from the original system in that groups replace households and individuals in each group are statistically identical, though individuals in households were not statistically identical. In the intermediate system a child interacts with not only his or her parent but any adult who is a parent, similarly an adult who is a parent interacts with every child and not only with his or her own child. Also, a parent interacts with every parent at the same rate, so spouses (i.e., parents in the same household) do not interact at higher rates. Thus, the intermediate system is in some sense more homogeneous than the original system, though it is not fully homogeneous because individuals in different groups are not statistically identical.

The evolution of the disease in the intermediate system can be captured by the following system of ordinary differential equations (ODEs), where $S_i(t)$ ($I_i(t)$, respectively) denote the number of individuals who are susceptible (infected, respectively) in group $i$ divided by the total number of individuals in all groups, $N$, with $\beta_i = \hat{\beta}_i/N$.

$$\frac{dS_P}{dt} = -S_P\left(\beta_1 I_P + \beta_3 I_X + \beta_2 I_C\right) \tag{1}$$
$$\frac{dS_X}{dt} = -S_X\left(\beta_3 I_P + \beta_1 I_X\right) \tag{2}$$
$$\frac{dS_C}{dt} = -S_C\left(\beta_2 I_P + \beta_1 I_C\right) \tag{3}$$
$$\frac{dI_P}{dt} = S_P\left(\beta_1 I_P + \beta_3 I_X + \beta_2 I_C\right) - \gamma I_P \tag{4}$$
$$\frac{dI_X}{dt} = S_X\left(\beta_3 I_P + \beta_1 I_X\right) - \gamma I_X \tag{5}$$
$$\frac{dI_C}{dt} = S_C\left(\beta_2 I_P + \beta_1 I_C\right) - \gamma I_C \tag{6}$$
$$\frac{dR}{dt} = \gamma I_P + \gamma I_X + \gamma I_C \tag{7}$$

The terms in red in equations (1)–(6) are quadratic terms representing the transmission of the disease to susceptible individuals via interaction with infectious individuals, resulting in the subsequent transition of susceptibles into the infected class. The terms in blue in equations (4)–(7) govern the recovery process, representing the rate at which individuals are cured or die (and are thus effectively removed from the active infection network). The corresponding total number of infected individuals at time $t$ is obtained by multiplying the sum of the infected fractions across all demographic groups by the total population size $N$, i.e., $I_{\text{total}}(t) = N\big(I_P(t) + I_X(t) + I_C(t)\big)$.

Classical results from epidemiology guarantee that $S_i(t)$ (respectively $I_i(t)$) evaluated by solving the ODE converges in probability at each time $t$ to the number of susceptible (infected, respectively) in group $i$ in the intermediate system divided by

the total number of individuals $N$ at the corresponding time, as $N$ grows to infinity, that is, in the asymptotic limit (see [5, 6]). We therefore consider the solution of the ODE as an accurate representation of the fraction of infected individuals in the intermediate system.

The solution of the ODEs can be evaluated fast using modest computational resources, regardless of the value of $N$. Since the ODE constitutes a deterministic system, they do not need to be evaluated over multiple runs. In our implementation, the time to solve the ODE system has been approximately 0.3 seconds using MATLAB R2024a on a MacBook Pro with an Apple M4 processor and 16 GB of memory regardless of the value of $N$.

We seek to assess if for appropriate choices of $\hat{\beta}_i$, given $\hat{\beta}, x$, the number of infected individuals as a function of time in the original system is close to that in the intermediate system, or equivalently to the solution of the ODEs, that is to $N(I_P(t) + I_X(t) + I_C(t))$. For this we first define a measure of proximity.

To quantify this agreement, we define the average trajectory error as the mean percentage difference between the infected populations predicted by the microscopic and macroscopic models (i.e., ODE) over all time points. Since we consider percentage difference, we ignore the times at which the infected fraction in the real system is very low as then the denominator for the difference becomes low inflating the percentage difference. We therefore consider time points at which the infected fraction in the original system exceeds a threshold $\varepsilon$. Specifically,

$$E_{\text{avg}} = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left| \frac{I_{\text{prob}}(t) - I_{\text{ODE}}(t)}{I_{\text{ODE}}(t)} \right| \times 100\%, \tag{8}$$

where $I_{\text{prob}}(t)$ denotes the ensemble-average infected fraction obtained from the original microscopic model, $I_{\text{ODE}}(t)$ denotes the infected fraction predicted by the deterministic ODE model, and $\mathcal{T} = \{t : I_{\text{prob}}(t) \geq \varepsilon\}$.

In fact, we will consider, $\hat{\beta}_1 = \hat{\beta}_3 = \hat{\beta}, \hat{\beta}_2 = yx\hat{\beta}$. That is, all adults, both parents and those who are not parents, interact with each other at rate $\hat{\beta}$, all children interact with each other at rate $\hat{\beta}$, but the interaction between parents and children happen at the rate $yx\hat{\beta}$. Note that the parameter $y$ in the intermediate model is intended to compensate for the loss of explicit household structure in it as compared to the original system. In the special case that $y = 1$, we have a metapopulation model with interaction rate $x\hat{\beta}$ between parents and children, and all other interaction rates (in Figure 2) as $\hat{\beta}$. But $y$ will be selected to minimize the average error.

The $y$ that minimizes the average error for a given $\hat{\beta}, x$, will be denoted as $y_{\text{best}}$ for that $\hat{\beta}, x$. We will determine if the average error attained at $y_{\text{best}}$ is acceptable for a wide range of $\hat{\beta}, x$. If so, then the macroscopic ODEs can substitute the microscopic Monte Carlo simulations for the original system and the computational challenge is redressed. Note that if in addition $\hat{\beta}_1, \hat{\beta}_3$ are also varied, the minimum error can only become lower. But the additional complexity need not be warranted given the errors we obtain considering $\hat{\beta}_1 = \hat{\beta}_3 = \hat{\beta}, \hat{\beta}_2 = yx\hat{\beta}$.

For our computation, we fine tune the intermediate system further. Parents are further classified based on their gender and marital status, specifically into coupled fathers

(CF), coupled mothers (CM), single fathers (SF), single mothers (SM), while children are partitioned into children of coupled parents (CC), children of single mothers (SMC), and children of single fathers (SFC). Coupled fathers (respectively mothers) are those fathers (respectively mothers) whose households comprise both parents and their children, single fathers (respectively mothers) are those fathers whose households comprise father (respectively mother) and his (respectively her) children. This separation allows the model to limit the interaction between parents and children, to that between 1) coupled fathers (respectively mothers) and the children of couples, rather than all children, 2) single fathers (respectively mothers) and the children of single fathers (respectively mothers) rather than all children. The interaction structure therefore approximates interactions in the original system with households better. Refer to Figure 3 to see the connection rates we consider for this fine tuned intermediate system and equations (9) - (11) for the ODEs that capture the evolution of the disease in it. The mathematical result that guarantees the accuracy of the representation in terms of the ODEs is stated in the appendix.

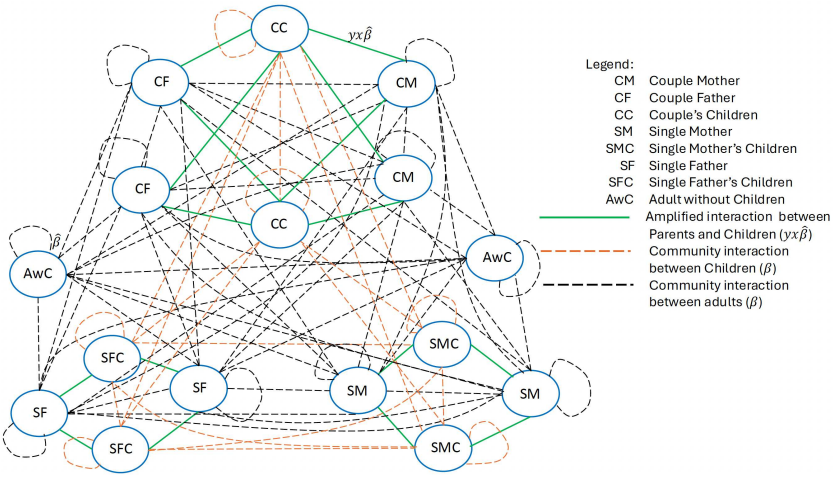



**Fig. 3** Macroscopic interaction network. Solid (green) lines denote calibrated parent–child interactions operating at the effective transmission rate $yx\hat{\beta}$. Dashed lines highlight community-level background transmission pathways split into adult-adult channels (black) and child-child channels (orange), each operating at the unamplified baseline rate $\hat{\beta}$.

$$\frac{dS_i}{dt} = -\begin{cases} S_i\left(\beta\sum_{j\in\mathcal{A}} I_j + yx\beta\, I_{c(i)}\right), & i\in\mathcal{A} \\ S_i\left(yx\beta\, I_{p(i)} + \beta\sum_{j\in\mathcal{C}} I_j\right), & i\in\mathcal{C} \end{cases} \tag{9}$$

$$\frac{dI_i}{dt} = \begin{cases} S_i \left( \beta \sum_{j \in \mathcal{A}} I_j + yx\beta\, I_{c(i)} \right) - \gamma I_i, & i \in \mathcal{A} \\ S_i \left( yx\beta\, I_{p(i)} + \beta \sum_{j \in \mathcal{C}} I_j \right) - \gamma I_i, & i \in \mathcal{C} \end{cases} \tag{10}$$

$$\frac{dR_i}{dt} = \begin{cases} \gamma I_i, & i \in \mathcal{A} \\ \gamma I_i, & i \in \mathcal{C} \end{cases} \tag{11}$$

where

$$\mathcal{P} = \{SF, CF, SM, CM\}, \quad \mathcal{A} = \mathcal{P} \cup \{AwC\}, \quad \mathcal{C} = \{SFC, SMC, CC\}, \quad \mathcal{G} = \mathcal{A} \cup \mathcal{C}.$$

$$c(i) = \begin{cases} SFC, & i = SF, \\ SMC, & i = SM, \\ CC, & i = CF \text{ or } CM, \end{cases} \qquad I_{p(i)} = \begin{cases} I_{SF}, & i = SFC, \\ I_{SM}, & i = SMC, \\ I_{CF} + I_{CM}, & i = CC, \end{cases}$$

and

$$I_{c(i)} = \begin{cases} I_{SFC}, & i = SF, \\ I_{SMC}, & i = SM, \\ I_{CC}, & i = CF \text{ or } CM, \\ 0, & i = AwC. \end{cases}$$

Equation 9 describes the depletion of susceptible individuals through infection. Equation 10 governs the evolution of the infected populations, while equation 11 describes recovery at rate $\gamma$. The terms in red in equations (9) and (10) are quadratic terms representing the transmission of the disease to susceptible individuals via interaction with infectious individuals while the terms in blue in equations (10) and (11) govern the recovery process, representing the rate at which individuals are cured or die (and are thus effectively removed from the active infection network).

For a fixed household amplification factor $x$, the microscopic model is simulated repeatedly to obtain an ensemble-average epidemic trajectory. The deterministic system is then solved for candidate values of the calibration parameter $y$, and the value minimizing the discrepancy between the deterministic and ensemble-average stochastic trajectories is selected and denoted by $y_{\text{best}}$.

## Results

This section assesses the capability of the proposed macroscopic model to approximate the microscopic model. Recall that within the microscopic framework, individuals are grouped in households and individuals contact other individuals within the same

household at rate $x\hat{\beta}$ and with others outside the household, with whom they can contact, at rate $\hat{\beta}$. Then, in the macroscopic framework, individuals are grouped as parents, children, adults without children, and interaction between the parent-children group is at rate $yx\hat{\beta}$, and interaction within and across all other groups in which contact is allowed is at rate $\hat{\beta}$. Here, $y$ is the calibration parameter. The microscopic model is evaluated through Monte Carlo simulation and the macroscopic model through the solution of ODEs. To obtain any data point in the microscopic model, the Monte Carlo simulations are repeated across several runs, and the overall infection counts in the runs are averaged to provide the ensemble average. We average over a sufficient number of runs that the ensemble average converges. This has required 200 or more runs.

This section is organized as follows.

- First, we assess whether the parameter $y$ is redundant, or more generally, if any macroscopic framework, or metapopulation model, can closely approximate the microscopic model. The answer is in the negative – in fact, the epidemic trajectory of the macroscopic model when $y = 1$ can be considerably different from that of the microscopic model.
- Second, we determine whether a single calibration parameter $y$ can compensate for the omission of explicit household structure while accurately reproducing the epidemic trajectory of the microscopic model. That is, given $x, \hat{\beta}$, is the error (as defined in the Methods section) low at the optimal value of $y$, $y_{\text{best}}$? To obtain $y_{\text{best}}$, the macroscopic model is solved for a range of values of $y$. The optimal value, $y_{\text{best}}$, is selected as the $y$ that gives the minimum error between the macroscopic trajectory and the corresponding ensemble average microscopic trajectory.
- Third, we demonstrate that the macroscopic model can be computed much faster than the microscopic one.
- Finally, once $y_{\text{best}}$ is obtained for a few values of $x$, given a $\hat{\beta}$, can values of $y_{\text{best}}$ be predicted for other values of $x$ through interpolation and extrapolation, and are the errors low at the predicted values of $y_{\text{best}}$? Note that interpolations and extrapolations are computationally simple. So, if the answer is yes to both, then the computationally tractable macroscopic model with the values of $y$s computed as above can replace the computationally intensive microscopic model while preserving the aggregate epidemic dynamics. Thus, the macroscopic framework retains the accuracy of the microscopic model while remaining computationally tractable.

For this study we use demographic data from recent U.S. Census Bureau and Current Population Survey (CPS), though the setup can capture data from the census of any other country. Approximately 76.2% of adults live without coresident children under the age of 18, while 23.8% are parents. Among parents, 82.8% live in coupled households, and the remaining 17.2% are single parents, of whom approximately 80% are single mothers and 20% are single fathers [20–24]. For simplicity, we consider that a household can have at most 4 children, because the fraction of US adults with 5 or more children is only 5% [25, 26]. As per US census, for a household with two parents, the number of children is 1, 2, 3, or 4 and above, with probabilities $0.214, 0.419, 0.244$, and $0.123$, respectively. The corresponding probabilities are $0.313, 0.379, 0.195$, and

0.113 for single-mother households, and $0.421, 0.385, 0.149$, and 0.045 for single-father households. For households with four or more children according to the original census distribution, we assign four children. This stochastic assignment produces realistic heterogeneity in household and contact patterns.

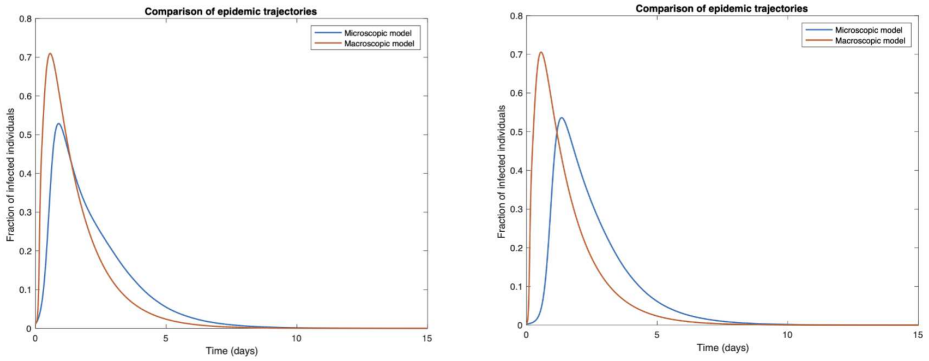
We have evaluated the models for a wide range of values of parameters $x, \hat{\beta}$ and different spreads of initial infections. The answers we obtain for the questions above for all the evaluations are similar. We reproduce the results at a few representative values in the remainder of the section. We reproduce for three different values of $\hat{\beta}$, separated by orders of magnitude. For each we consider several values of $x$ between 2 and 100, and a really high value of $x$, that is, 1000. Recall that $x > 1$ since individuals interact more frequently within households than outside. We consider two distributions for initial infection for each $x, \hat{\beta}$ : 1) infection is initially confined to the adult community, with 2% of adults without children infected at the start of the epidemic (Scenario A); 2) infection is seeded simultaneously in both the community and households, with 1% of children and approximately 0.1% of adults without children initially infected (Scenario B). The initial infection scenarios have been selected so as to capture the frequently posited social phenomena that adults without children are more prone to initial infection as they are more socially active, and children are also prone to initial infection because of their less than fully developed immune system and because they interact closely with other children in their schools. In Scenario A, initial infection is seeded only in households of size one, while in Scenario B, initial infection is seeded in larger households. Since contact rate is higher within households, infection will spread faster in Scenario B. Thus, the two scenarios capture different aspects.

## Can any macroscopic model well approximate the microscopic model?

We examine if a naive macroscopic model can well approximate the microscopic model. For a naive macroscopic model, we consider $y = 1$. Recall from the Methods section that this is the metapopulation model, which ignores the existence of households as in all metapopulation models, but also does not alter the transmission rates between the subgroups to compensate for the above simplification. We plot the evolution of the fraction of infected individuals for the microscopic model and the macroscopic model for Scenarios A, B, and an example value of $x, \hat{\beta}$, namely $\hat{\beta} = 0.002, x = 25$ (Figure 4). We notice a significant mismatch. The macroscopic model substantially overestimates the peak infection level and advances the epidemic peak in both Scenario A (Figure 4a) and Scenario B (Figure 4b). It also produces a steeper decline during the later stage of the epidemic. This also establishes that, sans adaptation, metapopulation models can not capture the evolution of infection in a network with households.

## Can the macroscopic model be selected to well approximate the microscopic model?

We now compare the epidemic trajectories of the microscopic model and the macroscopic model with $y = y_{best}$ for different values of $x$, $\hat{\beta} = 0.002, 0.005$, and 0.01 and




(a) $\hat{\beta} = 0.002$, Scenario A, $x = 25$, Average Error = 91.4751%

(b) $\hat{\beta} = 0.002$, Scenario B, $x = 25$, Average Error= 126.8655%

**Fig. 4** Comparison of epidemic trajectories between the microscopic model and the macroscopic model for $y = 1$.

Scenarios A and B. Table 1 shows that the average error between the two trajectories is below 3.5%. Thus, the macroscopic model closely approximates the microscopic model under a diverse set of conditions. In fact, Scenario A achieves errors below 1% for nearly all values of $x$. Scenario B exhibits slightly larger errors because seeding in larger households produces stronger clustering effects leading to bigger divergence between the macroscopic and microscopic models. But, even in Scenario B, all but one entry is 1.6% or below. Also, all but three entries is below 1.5%. Note that, with $y = y_{best}$, for the parameters given in Figure 5, the average error is now 0.61% and 1.43%. This is a substantial drop from the average error seen when $y = 1$.

Additional computations were performed for household amplification factors as large as $x = 1000$ to capture scenarios where individuals interact at a significantly higher rate within the household than outside the household. The same qualitative behavior was observed across both outbreak scenarios and both transmission rates. At $x = 1000$, the trajectory errors remained below 1.5453% and 2.0406% for $\hat{\beta} = 0.002$, below 0.51921% and 1.0464% for $\hat{\beta} = 0.005$, and below 0.42183% and 0.48301% for $\hat{\beta} = 0.01$ under Scenarios A and B, respectively. These results demonstrate that the proposed framework remains accurate and robust over a wide range of household transmission intensities than that shown in Tables 1.

We also present example trajectories of the number of infected individuals as a function of time for both the microscopic and macroscopic models to visualize the discrepancies between them. For each transmission rate $\hat{\beta}$ and initial infection scenarios, we present the trajectories for the $x$ that gave the maximum percentage trajectory error (Figure 5). The trajectory plots show that the microscopic and macroscopic models match very well throughout the entire duration of the epidemic. Minor discrepancies between the two trajectories are confined primarily to the epidemic peak and the immediate region. For $\hat{\beta} = 0.002$, the largest discrepancy occurs past the peak, though even here the discrepancy is barely visible. Note that Figure A1 in the

**Table 1** Optimal calibration parameter $y_{\text{best}}$ and percentage trajectory error across different values of the household amplification factor $x$ for $\hat{\beta} \in \{0.002, 0.005, 0.01\}$. E (%) denotes Error (%).

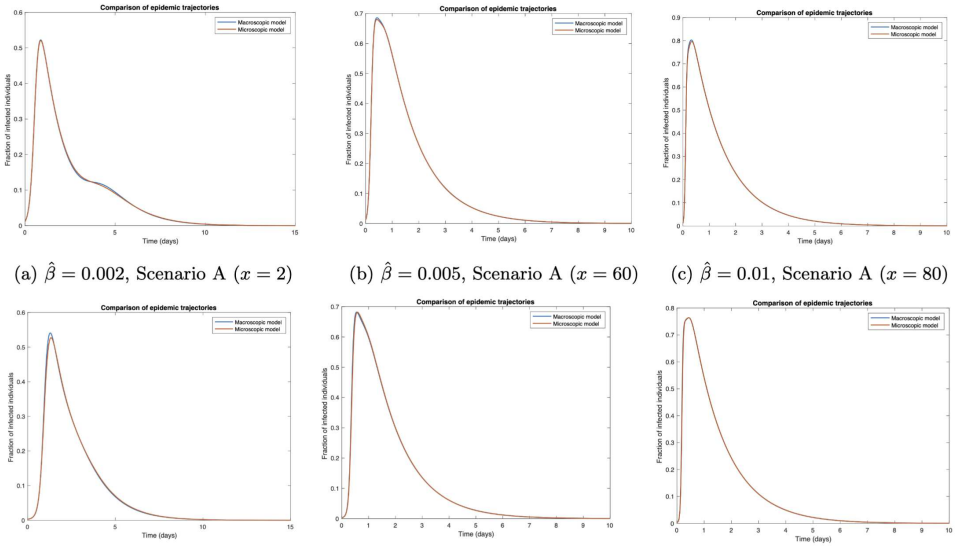
| | $\hat{\beta} = 0.002$ | | | | $\hat{\beta} = 0.005$ | | | | $\hat{\beta} = 0.01$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Scenario A | | Scenario B | | Scenario A | | Scenario B | | Scenario A | | Scenario B | |
| $x$ | $y_{\text{best}}$ | E (%) | $y_{\text{best}}$ | E (%) | $y_{\text{best}}$ | E (%) | $y_{\text{best}}$ | E (%) | $y_{\text{best}}$ | E (%) | $y_{\text{best}}$ | E (%) |
| 2 | 0.00200 | 2.30 | 0.000001 | 3.13 | 0.00215 | 0.30 | 0.00001 | 0.70 | 0.00250 | 0.27 | 0.00001 | 0.55 |
| 5 | 0.00225 | 0.61 | 0.00125 | 1.53 | 0.00225 | 0.42 | 0.00001 | 0.93 | 0.00270 | 0.32 | 0.00050 | 0.26 |
| 10 | 0.00250 | 0.43 | 0.00175 | 0.85 | 0.00265 | 0.31 | 0.00090 | 1.15 | 0.00270 | 0.27 | 0.00090 | 0.24 |
| 20 | 0.00300 | 0.64 | 0.00270 | 1.45 | 0.00265 | 0.35 | 0.00220 | 0.98 | 0.00285 | 0.27 | 0.00115 | 0.50 |
| 30 | 0.00310 | 0.51 | 0.00300 | 1.11 | 0.00285 | 0.22 | 0.00250 | 0.61 | 0.00310 | 0.22 | 0.00180 | 0.29 |
| 40 | 0.00310 | 0.66 | 0.00305 | 1.16 | 0.00300 | 0.25 | 0.00270 | 0.51 | 0.00325 | 0.24 | 0.00180 | 0.26 |
| 50 | 0.00315 | 0.52 | 0.00315 | 1.40 | 0.00325 | 0.37 | 0.00285 | 0.43 | 0.00325 | 0.29 | 0.00240 | 0.34 |
| 60 | 0.00325 | 0.94 | 0.00315 | 1.11 | 0.00330 | 0.53 | 0.002925 | 0.49 | 0.00330 | 0.28 | 0.00245 | 0.30 |
| 70 | 0.00325 | 0.65 | 0.003225 | 1.22 | 0.00330 | 0.44 | 0.00340 | 0.62 | 0.00355 | 0.41 | 0.00255 | 0.37 |
| 80 | 0.00330 | 0.91 | 0.00340 | 1.28 | 0.00340 | 0.45 | 0.003475 | 0.70 | 0.00380 | 0.43 | 0.00300 | 0.34 |
| 90 | 0.00340 | 0.52 | 0.00340 | 1.28 | 0.00355 | 0.28 | 0.00355 | 0.64 | 0.00390 | 0.42 | 0.00300 | 0.45 |
| 100 | 0.003425 | 0.87 | 0.00350 | 1.54 | 0.003625 | 0.42 | 0.003575 | 0.49 | 0.00400 | 0.53 | 0.00315 | 0.33 |

Appendix show that with $y = y_{best}$, for the parameter used in Figure 4, the trajectory for the macroscopic model closely match that of microscopic model, in contrast to the huge discrepancies for $y = 1$ as in Figure 4.

Table 1 show that $y_{\text{best}}$ increases monotonically with increase in $x$. This is intuitive because higher $x$ ensures faster transmission of the disease in households and therefore eventually faster overall spread as members of a household also get in contact with members outside. Increasing $y$ increases the transmission rate between adults and children in the microscopic model, and therefore higher overall spread of the disease. Thus, the best match for $y$ occurs at higher values as $x$ increases.

**Table 2** Computation time required to generate one data point in the microscopic model across 200 runs for $\hat{\beta} = 0.002$ and Scenario A.

| Population ($N$) | Time (s) | Population ($N$) | Time (s) | Population ($N$) | Time (s) |
|---|---|---|---|---|---|
| 5,000 | 16,800 | 12,000 | 55,000 | 19,000 | 138,400 |
| 6,000 | 23,200 | 13,000 | 60,200 | 20,000 | 156,400 |
| 7,000 | 24,400 | 14,000 | 69,200 | 22,500 | 170,400 |
| 8,000 | 28,800 | 15,000 | 76,400 | 25,000 | 205,600 |
| 9,000 | 35,000 | 16,000 | 83,200 | 30,000 | 253,000 |
| 10,000 | 40,400 | 17,000 | 97,800 | 40,000 | 756,000 |
| 11,000 | 53,400 | 18,000 | 104,000 | 50,000 | 933,600 |




(a) $\hat{\beta} = 0.002$, Scenario A ($x = 2$) (b) $\hat{\beta} = 0.005$, Scenario A ($x = 60$) (c) $\hat{\beta} = 0.01$, Scenario A ($x = 80$)

(d) $\hat{\beta} = 0.002$, Scenario B ($x = 2$) (e) $\hat{\beta} = 0.005$, Scenario B ($x = 10$) (f) $\hat{\beta} = 0.01$, Scenario B ($x = 2$)

**Fig. 5** Comparison of infected individual trajectories over time between the microscopic and macroscopic models for all transmission rates, $\hat{\beta} = 0.002, 0.005,$ and $0.01$, and outbreak scenarios (Scenario A and Scenario B), evaluated at the household amplification factor $x$ corresponding to the worst-case trajectory error in each scenario.

## Computation times of Microscopic and Macroscopic models

Recall that each data point for the microscopic model needs to be averaged over a large number of independent realizations to ensure that the sample average is close to the expectation. We have found that for the sample average to converge to any value, 200 or larger number of runs are required. In Table 2, we provide the time required for 200 independent realizations for $\hat{\beta} = 0.002$, Scenario A, and different numbers of individuals. The computation times are similar for other values of $\hat{\beta}$ and other distributions of the initial infection.

In contrast, the macroscopic model requires only a single deterministic ODE solution for a given parameter configuration. In our implementation, this required approximately 0.3 seconds and remained essentially independent of population size because the computational cost is determined primarily by the number of demographic compartments rather than the number of individuals.

In both cases we used MATLAB R2024a on a MacBook Pro with an Apple M4 processor and 16 GB of memory. Consequently, replacing the 200 microscopic realizations with a single calibrated macroscopic solution results in substantial computational savings. For a population of 5,000, the corresponding speedup is approximately 56,000 times, while for a population of 50,000 it is approximately 3.1 million times. The computational savings provide an important practical advantage of the proposed macroscopic framework.

## Using the macroscopic model to predict the outcome of the microscopic model

Determining $y_{\text{best}}$ for every new value of $x$ would largely negate the computational advantages of the macroscopic approximation. We therefore investigate whether the relationship between $x$ and $y_{\text{best}}$ can itself be represented by a smooth functional approximation using few computations for both. We can then predict $y_{\text{best}}$ for all values of $x$ from the function. We obtain this function using B-spline regression (Figures 6). The fitted curves closely match the computed data, with mean squared fitting errors on the order of $10^{-6}$, indicating that the regression accurately captures the nonlinear relationship between $x$ and $y_{\text{best}}$. This smooth relationship between the two parameters indicates that calibration does not need to be repeated for every value of $x$. Instead, intermediate values of $y_{\text{best}}$ can be estimated directly from the fitted spline, ensuring good predictive capability of the fitted regression, which we verify next.

We reproduce the results in two representative cases in which $y_{\text{best}}$ can be assessed through interpolation ($x = 25, 75$) and two others when one needs to resort to extrapolation ($x = 105, 110$). In each case we compare the $y_{\text{best}}$ obtained from the fitted curve, henceforth denoted as $y_{\text{predicted}}$, and the $y_{\text{best}}$ computed by evaluating both the microscopic and macroscopic models for the given parameter values (Tables 3). We continue to denote the latter as $y_{\text{best}}$. Across all scenarios and transmission rates, the percentage difference between the actual $y_{\text{best}}$ and the predicted $y_{\text{best}}$, i.e., $y_{\text{predicted}}$, remains well below 5%, peaking at just 4.43% in the worst-case scenario (Scenario B for $\hat{\beta} = 0.01$ at $x = 25$). For interpolation, the predicted and actual values differ by only $10^{-6}$–$10^{-5}$ across all transmission rates and outbreak scenarios (with a maximum absolute difference of $7.53 \times 10^{-5}$ in Scenario B for $\hat{\beta} = 0.005$ at $x = 75$). Extrapolation remains similarly robust, with absolute differences between predicted and actual parameters bounded by $10^{-4}$ and trajectory error differences remaining small overall. Consequently, $y_{\text{best}}$ for previously unseen values of $x$ can be accurately obtained directly from the regression model, eliminating the need for repeated evaluation of the computationally time consuming microscopic model.

The fitted curves also indicate diminishing sensitivity of $y_{\text{best}}$. Although $y_{\text{best}}$ increases monotonically with $x$, the increase becomes progressively smaller for larger values of $x$; a saturation phenomenon is therefore observed.

**(a) Performance for $\hat{\beta} = 0.002$**

**Table 3** Interpolation and extrapolation performance comparing $y_{\text{best}}$ (obtained via optimization) and $y_{\text{predicted}}$ (derived from the fitted response curve). Percentage difference is calculated as $\frac{y_{\text{predicted}} - y_{\text{best}}}{y_{\text{best}}} \times 100$. Actual Error (%) measures the trajectory percentage error obtained for $y_{\text{best}}$, while Predicted Error (%) denotes the corresponding trajectory error generated when substituting $y_{\text{predicted}}$ in the ODE.

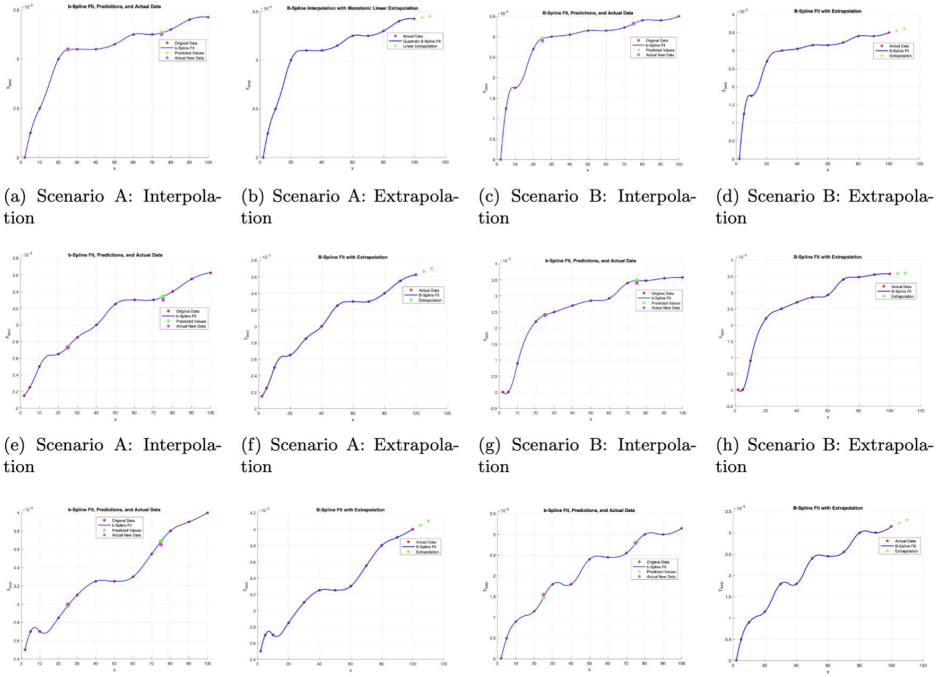
| Scenario | Type | $x$ | $y_{\text{best}}$ | $y_{\text{predicted}}$ | Percentage Difference (%) | Actual Error (%) | Predicted Error (%) |
|---|---|---|---|---|---|---|---|
| A | Interpolation | 25 | 0.00310 | 0.0030963 | 0.12 | 0.61 | 0.61 |
| | | 75 | 0.00325 | 0.0032657 | 0.48 | 0.67 | 0.72 |
| A | Extrapolation | 105 | 0.00350 | 0.0034375 | 1.79 | 0.63 | 0.66 |
| | | 110 | 0.00350 | 0.0034500 | 1.43 | 0.60 | 0.61 |
| B | Interpolation | 25 | 0.00290 | 0.0029600 | 2.07 | 1.43 | 1.45 |
| | | 75 | 0.003325 | 0.0033194 | 0.17 | 1.43 | 1.43 |
| B | Extrapolation | 105 | 0.003575 | 0.0035500 | 0.70 | 1.02 | 1.03 |
| | | 110 | 0.003575 | 0.0036000 | 0.70 | 2.68 | 2.77 |

**(b) Performance for $\hat{\beta} = 0.005$**

| Scenario | Type | $x$ | $y_{\text{best}}$ | $y_{\text{predicted}}$ | Percentage Difference (%) | Actual Error (%) | Predicted Error (%) |
|---|---|---|---|---|---|---|---|
| A | Interpolation | 25 | 0.002725 | 0.0027372 | 0.45 | 0.22 | 0.45 |
| | | 75 | 0.003300 | 0.0033382 | 1.16 | 0.24 | 0.72 |
| A | Extrapolation | 105 | 0.003650 | 0.0036625 | 0.34 | 0.65 | 0.44 |
| | | 110 | 0.003675 | 0.0037000 | 0.68 | 0.37 | 0.80 |
| B | Interpolation | 25 | 0.002400 | 0.0024090 | 0.38 | 0.53 | 0.73 |
| | | 75 | 0.003400 | 0.0034753 | 2.21 | 0.63 | 0.80 |
| B | Extrapolation | 105 | 0.003625 | 0.0035875 | 1.03 | 0.67 | 0.59 |
| | | 110 | 0.003625 | 0.0036000 | 0.69 | 0.54 | 0.86 |

**(c) Performance for $\hat{\beta} = 0.01$**

| Scenario | Type | $x$ | $y_{\text{best}}$ | $y_{\text{predicted}}$ | Percentage Difference (%) | Actual Error (%) | Predicted Error (%) |
|---|---|---|---|---|---|---|---|
| A | Interpolation | 25 | 0.00300 | 0.0029835 | 0.55 | 0.29 | 0.32 |
| | | 75 | 0.00365 | 0.0036879 | 1.04 | 0.15 | 0.44 |
| A | Extrapolation | 105 | 0.00410 | 0.0040500 | 1.22 | 0.29 | 0.30 |
| | | 110 | 0.00410 | 0.0041000 | 0.00 | 0.69 | 0.69 |
| B | Interpolation | 25 | 0.00155 | 0.0014814 | 4.43 | 0.43 | 0.43 |
| | | 75 | 0.00280 | 0.0027839 | 0.58 | 0.45 | 0.49 |
| B | Extrapolation | 105 | 0.00315 | 0.0032250 | 2.38 | 0.24 | 0.76 |
| | | 110 | 0.00340 | 0.0033000 | 2.94 | 0.45 | 0.42 |



(a) Scenario A: Interpolation (b) Scenario A: Extrapolation (c) Scenario B: Interpolation (d) Scenario B: Extrapolation

(e) Scenario A: Interpolation (f) Scenario A: Extrapolation (g) Scenario B: Interpolation (h) Scenario B: Extrapolation

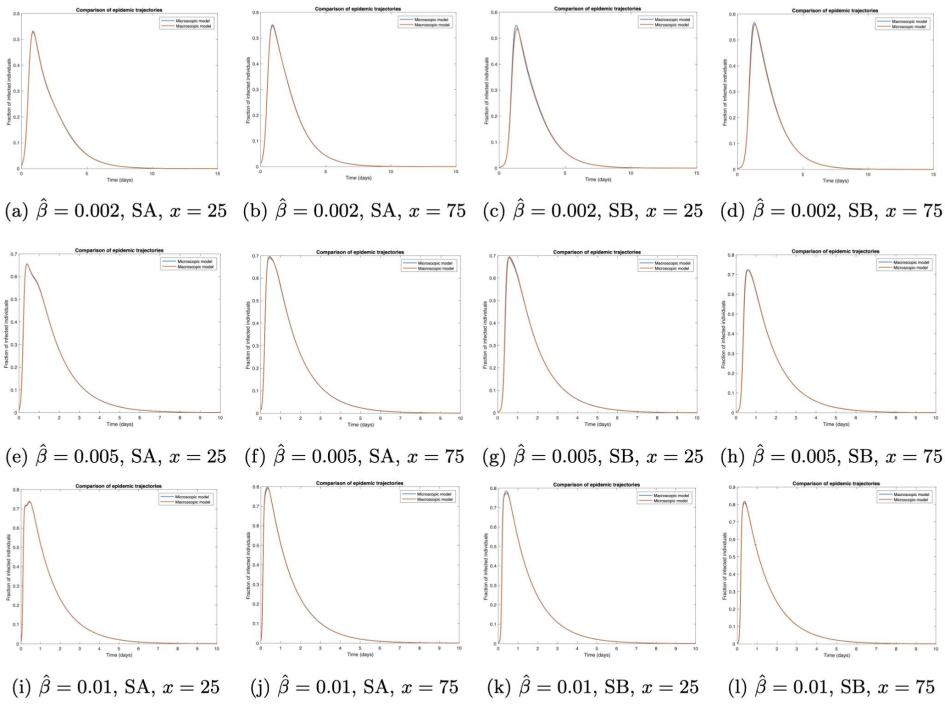
(i) Scenario A: Interpolation (j) Scenario A: Extrapolation (k) Scenario B: Interpolation (l) Scenario B: Extrapolation

**Fig. 6** B-spline regression relating the household amplification factor $x$ to the optimal calibration parameter $y_{\text{best}}$ for $\beta = 0.002$, $\beta = 0.005$, and $\beta = 0.01$. Each row corresponds to a value of $\beta$, while the columns show interpolation and extrapolation for Scenarios A and B, respectively.

We now plot the trajectories of the macroscopic model with $y_{\text{predicted}}$ obtained from interpolation (Figure 7) and extrapolation (Figure 8), and the microscopic model for $x = 25$ and $\hat{\beta} = 0.002$. A close match is observed. Thus, the small differences between $y_{\text{predicted}}$ (from interpolation and extrapolation) and $y_{\text{best}}$ have little effect on the epidemic dynamics.

# Discussion

This study establishes a quantitative and computationally efficient correspondence between a household-structured microscopic probabilistic SIR model and a macroscopic deterministic ODE model. The microscopic model explicitly represents the exact network of connection between individuals within and across households, whereas the macroscopic model aggregates individuals into broader demographic groups and aggregate connections between groups rather than between individuals and therefore can not capture the household organization. The evolution of infection in the macroscopic model is much more computationally tractable to compute than that in the



(a) $\hat{\beta} = 0.002$, SA, $x = 25$ (b) $\hat{\beta} = 0.002$, SA, $x = 75$ (c) $\hat{\beta} = 0.002$, SB, $x = 25$ (d) $\hat{\beta} = 0.002$, SB, $x = 75$

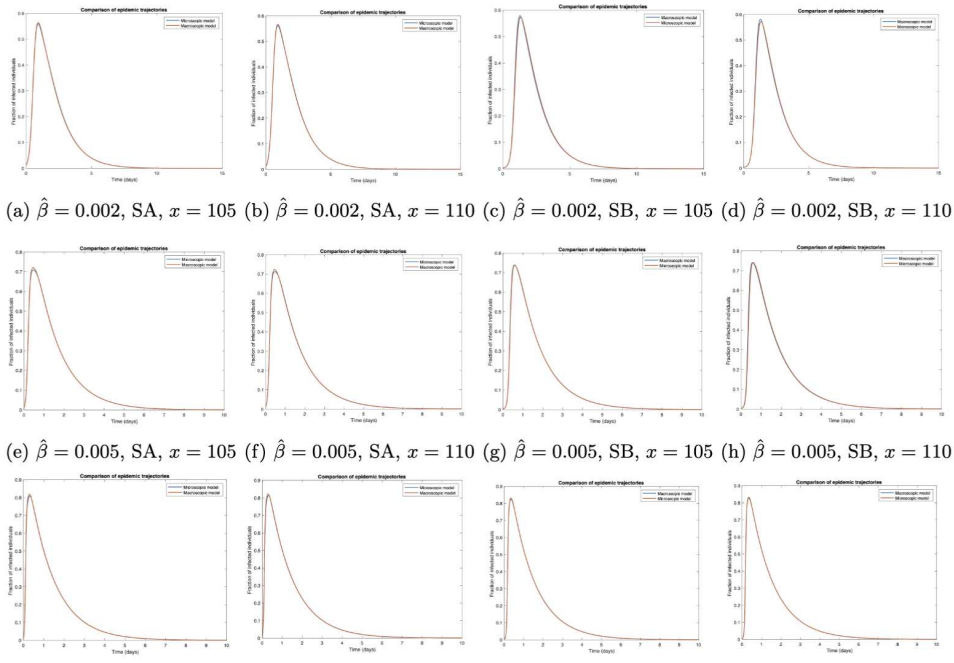
(e) $\hat{\beta} = 0.005$, SA, $x = 25$ (f) $\hat{\beta} = 0.005$, SA, $x = 75$ (g) $\hat{\beta} = 0.005$, SB, $x = 25$ (h) $\hat{\beta} = 0.005$, SB, $x = 75$

(i) $\hat{\beta} = 0.01$, SA, $x = 25$ (j) $\hat{\beta} = 0.01$, SA, $x = 75$ (k) $\hat{\beta} = 0.01$, SB, $x = 25$ (l) $\hat{\beta} = 0.01$, SB, $x = 75$

**Fig. 7** Comparison of epidemic trajectories between the microscopic model and the macroscopic model using the interpolated calibration parameter $y_{\text{predicted}}$ under Scenarios A and B. The macroscopic trajectories are generated using $y_{\text{predicted}}$ obtained from the B-spline interpolation.

microscopic one. The central objective was therefore to determine whether the structural effects omitted by the macroscopic formulation could be incorporated through a calibrated interaction scaling parameter, $y$, in the macroscopic one, while preserving the aggregate epidemic dynamics generated by the microscopic model. The results demonstrate that this is feasible across a broad range of household amplification factors, baseline transmission rates, and initial infection scenarios, thereby establishing an efficient bridge between detailed household-based microscopic and computationally tractable macroscopic epidemic models.

Specifically, a major finding is that the macroscopic model closely reproduces the epidemic trajectories of the microscopic model across a diverse set of parameters involving a wide range of household amplification factors, baseline transmission rates, and initial infection scenarios. The average error across the entire trajectory has been small for this diverse set of parameters, and comparison of the trajectories reveal that the maximum error is also low. For each transmission rate and initial infection scenario, the microscopic and macroscopic trajectories corresponding to the worst observed average trajectory error were compared directly. The resulting curves exhibit close agreement over the course of the epidemic, with the largest discrepancies generally concentrated around the epidemic peak or its immediate vicinity. Even



(a) $\hat{\beta} = 0.002$, SA, $x = 105$ (b) $\hat{\beta} = 0.002$, SA, $x = 110$ (c) $\hat{\beta} = 0.002$, SB, $x = 105$ (d) $\hat{\beta} = 0.002$, SB, $x = 110$

(e) $\hat{\beta} = 0.005$, SA, $x = 105$ (f) $\hat{\beta} = 0.005$, SA, $x = 110$ (g) $\hat{\beta} = 0.005$, SB, $x = 105$ (h) $\hat{\beta} = 0.005$, SB, $x = 110$

(i) $\hat{\beta} = 0.01$, SA, $x = 105$ (j) $\hat{\beta} = 0.01$, SA, $x = 110$ (k) $\hat{\beta} = 0.01$, SB, $x = 105$ (l) $\hat{\beta} = 0.01$, SB, $x = 110$

**Fig. 8** Comparison of epidemic trajectories between the microscopic model and the macroscopic model using the extrapolated calibration parameter $y_{\text{predicted}}$ under Scenarios A and B. The macroscopic trajectories are generated using $y_{\text{predicted}}$ obtained from the B-spline extrapolation.

in these worst-case comparisons, the differences remain visually small. This agreement is important because the calibration does not merely match a single summary statistic, such as the epidemic peak or final epidemic size; rather, the calibrated macroscopic model reproduces the temporal evolution of the aggregate infected population. The worst case comparison also reveals that the maximum error is low. The errors are relatively higher when the parameters induce greater household clustering (such as in Scenario B of initial infection) but even those are small. Another major finding is that the relation between the value of the calibration parameter yielding the minimum error and the household amplification factor is continuous; thus calibration parameters that produce low errors for different values of the household amplification factor can be obtained through interpolation and extrapolation using some baseline values of the optimum calibration parameters and corresponding household amplification factors and B-spline regression. In fact, we show that with the interpolated and extrapolated values of the calibration parameter, the macroscopic trajectory is close to the microscopic one. This reduces the overall computational burden significantly as determining the calibration parameter that yields the minimum error requires the

computation of the microscopic trajectory, and now because of interpolation, extrapolation the above determination needs be done only for a few values of the household amplification factor.

The computational savings provide an important practical advantage of the proposed framework. Generating an epidemic trajectory from the microscopic model requires repeated Monte Carlo simulations, and the computational cost increases substantially with population size. For example, generating one ensemble-averaged data point using 200 Monte Carlo realizations required approximately 16,800 seconds for a population of 5,000 individuals and 40,400 seconds for a population of 10,000. The computational cost increased to approximately 156,400 seconds for 20,000 individuals and 933,600 seconds for 50,000 individuals. In contrast, the macroscopic deterministic ODE system required approximately 0.3 seconds to solve in our implementation and was approximately independent of population size because the number of ODE state variables is determined by the fixed number of demographic compartments rather than by the number of individuals. For the 5,000-individual case, replacing the microscopic calculation with a single macroscopic ODE solution corresponds to an approximately 56,000-fold reduction in computation time. These results demonstrate that the benefit of the proposed microscopic–macroscopic bridge is not merely a reduction in model complexity, but can translate into substantial computational savings when many epidemic scenarios must be evaluated.

The computational advantage becomes particularly important when parameter sweeps, uncertainty analyses, sensitivity analyses, or repeated intervention scenarios are considered. In such settings, performing hundreds or thousands of microscopic simulations for every parameter combination may become prohibitively expensive. The proposed "calibrate once, analyze many" strategy instead concentrates the computational burden in a limited number of microscopic simulations and subsequently transfers the resulting structural information to the deterministic model. Once the relationship between $x$ and $y_{\text{best}}$ has been established, effective macroscopic parameters for previously untested values of $x$ can be estimated without rerunning the microscopic model. The resulting ODE system can then be solved rapidly for a large number of parameter combinations while maintaining close agreement with the microscopic epidemic dynamics.

The practical significance of this approach lies in its ability to retain important effects of household structure without requiring household-level stochastic simulation for every analysis. The microscopic model remains valuable as a high-fidelity representation of household-level transmission and as a reference for determining the effective calibration parameter. The calibrated macroscopic model provides a complementary representation: it does not explicitly reproduce every microscopic contact, but instead incorporates their aggregate effects through an effective interaction parameter. The resulting framework therefore provides a principled bridge between structural realism and computational tractability.

We now acknowledge the limitations of our work. First, the network topology we have considered has a diameter of 2, that is, the maximum distance between two nodes (representing individuals) in the network graph is 2. This arises in closely knit communities such as small towns, neighborhoods, apartment buildings or cruise ships,

but can not model large cities where individuals are spread out in various communities and as such the distance between individuals can be several hops. Thus, our findings need not extend to networks with large diameters. Second, B-spline regression does not provide low errors for extrapolation over large ranges. So, other regression techniques have to be explored for those. Third, the present study considers a simple SIR disease process without reinfection, vaccination, behavioral or opinion dynamics. Allowing reinfection for example would introduce an additional transition from the recovered state back to the susceptible state and could substantially alter the long-term epidemic dynamics, potentially producing recurrent waves of infection. Consequently, results may not extend to such evolved scenarios. But, death can be easily incorporated by introducing a separate dead compartment and individuals transitioning to either recovered or dead from the infected compartment. The dead individuals are removed from the system like the recovered ones as neither infect other individuals. So the dynamics do not fundamentally change.

Despite these limitations, the consistency of the results provide a computationally tractable framework for evaluating the evolution of epidemic in close knit communities with households which is of practical interest. The limitations of the current framework provide topics of future research. Last but not the least, close knit communities may have groups of individuals which have all the characteristics of households, but are not deemed households in common parlance. In common parlance the term household is used to denote families with parents and offsprings. But, broadly, the characteristic of such households we use is that each such household is a small group of individuals who interact among themselves at a much higher rate than with other individuals. That characteristic extends to small groups in other settings such as in prisons where a small group of prisoners share the same cell and they are in close proximity to each other for long durations as also with the guards of their cell. Then such small groups can be adult roommates sharing a dormitory room in University housings. We believe our results would extend to these scenarios as well though the assessment constitutes a topic of future research.

More broadly, the results provide a foundation for extending this calibration strategy to increasingly complex structured epidemic systems. Future work can investigate whether the microscopic–macroscopic calibration framework can be integrated with additional layers of heterogeneity, evolving contact patterns, age-dependent susceptibility, partial immunity, vaccination, or intervention-driven changes in behavior. Such extensions could allow the computational efficiency of deterministic epidemic models to be retained while systematically incorporating structural effects that are otherwise accessible only through substantially more expensive microscopic simulations.

***Data availability:***

Key summary metrics are reported directly within the published article. The demographic information used in the current study is publicly available from the U.S. Census Bureau, including data on families and living arrangements from 2020 available at https://www.census.gov/data/tables/2020/demo/families/cps-2020.html and 2023 available at https://www.census.gov/data/tables/2023/demo/families/cps-2023.html, as well as interactive visualizations on parents' living arrangements available at https:

//www.census.gov/library/visualizations/interactive/parents-living-arrangements.html, historical reports on adults living without children available at https://www.census.gov/newsroom/blogs/random-samplings/2016/12/no_kids_in_the_house.html, and analyses of household structures available at https://www.census.gov/library/stories/2017/08/more-adults-living-without-children.html. All other datasets generated and analyzed, along with the computer code used to run the simulations and generate the figures, are available from the corresponding author upon reasonable request.

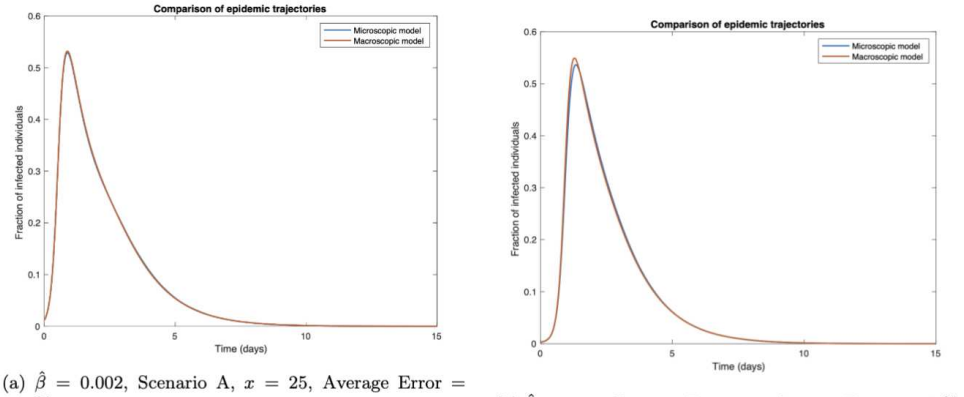
***Conflict of interest:***

The authors declare no conflict of interest

***Funding:***

This research received no specific funding or financial support.

# Appendix A Additional Figures




(a) $\hat{\beta} = 0.002$, Scenario A, $x = 25$, Average Error = 0.61%

(b) $\hat{\beta} = 0.002$, Scenario B, $x = 25$, Average Error= 1.43%

**Fig. A1** Comparison of epidemic trajectories between the microscopic model and the macroscopic model for $y_{best}$.

# References

[1] Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character **115**(772), 700–721 (1927)

[2] Kermack, W.O., McKendrick, A.G..: Contributions to the mathematical theory of epidemics–i. 1927. Bulletin of mathematical biology **53**(1-2), 33–55 (1991)

[3] Sharkey, K.J.: Deterministic epidemiological models at the individual level. Journal of Mathematical Biology **57**(3), 311–331 (2008)

[4] World Bank Document: An Introduction to Deterministic Infectious Disease Models. World Bank. https://documents1.worldbank.org/curated/en/888341625223820901/pdf/An-Introduction-to-Deterministic-Infectious-Disease-Models.pdf. Accessed: 31 Dec. 2025

[5] Ali, R.N., Rubin, H., Sarkar, S.: Countering the potential re-emergence of a deadly infectious disease—information warfare, identifying strategic threats, launching countermeasures. Plos one **16**(8), 0256014 (2021)

[6] Ali, R.N., Sarkar, S.: Impact of opinion dynamics on the public health damage inflicted by covid-19 in the presence of societal heterogeneities. Frontiers in Digital Health **5**, 1146178 (2023)

[7] Berge, T., Lubuma, J.-S., Moremedi, G., Morris, N., Kondera-Shava, R.: A simple mathematical model for ebola in africa. Journal of biological dynamics **11**(1), 42–74 (2017)

[8] Bjørnstad, O.N., Finkenstädt, B.F., Grenfell, B.T.: Dynamics of measles epidemics: estimating scaling of transmission rates using a time series sir model. Ecological monographs **72**(2), 169–184 (2002)

[9] Coburn, B.J., Wagner, B.G., Blower, S.: Modeling influenza epidemics and pandemics: insights into the future of swine flu (h1n1). BMC medicine **7**(1), 30 (2009)

[10] Apolloni, A., Poletto, C., Ramasco, J.J., Jensen, P., Colizza, V.: Metapopulation epidemic models with heterogeneous mixing and travel behaviour. Theoretical Biology and Medical Modelling **11**(1), 3 (2014)

[11] Colizza, V., Vespignani, A.: Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. Journal of theoretical biology **251**(3), 450–467 (2008)

[12] House, T., Keeling, M.J.: Household structure and infectious disease transmission. Epidemiology & Infection **137**(5), 654–661 (2009)

[13] Ball, F., Mollison, D., Scalia-Tomba, G.: Epidemics with two levels of mixing. The Annals of Applied Probability, 46–89 (1997)

[14] Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., *et al.*: Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine **5**(3), 74 (2008)

[15] Toth, D.J., Sheets, T.R., Beams, A.B., Ahmed, S.M., Seegert, N., Love, J., Keegan, L.T., Samore, M.H.: Model-based estimates of age-structured sars-cov-2 epidemiology in households. BMC Public Health **24**(1), 2965 (2024)

[16] Historical Households Tables. Households by Size: 1960 to Present. United States Census Bureau. https://www.census.gov/data/tables/time-series/demo/families/households.html. Accessed: 14 May 2026

[17] Cruise Ship Passenger Capacity. CruiseMapper. https://www.cruisemapper.com/wiki/761-cruise-ship-passenger-capacity-ratings. Accessed: 14 May 2026

[18] Three Things Small Cities Teach Us About Civic Engagement. National Leagues of Cities. https://www.nlc.org/article/2019/07/19/three-things-small-cities-teach-us-about-civic-engagement/. Accessed: 16 May 2026 (2019)

[19] Ratcliffe, M.: Redefining Urban Areas following the 2020 Census. United States Census Bureau. https://www.census.gov/newsroom/blogs/random-samplings/2022/12/redefining-urban-areas-following-2020-census.html. Accessed: 16 May 2026 (2022)

[20] America's Families and Living Arrangements: 2020. United States Census Bureau. https://www.census.gov/data/tables/2020/demo/families/cps-2020.html. Accessed: 16 Jan. 2026

[21] America's Families and Living Arrangements: 2023. United States Census Bureau. https://www.census.gov/data/tables/2023/demo/families/cps-2023.html. Accessed: 16 Jan. 2026

[22] Parents' Living Arrangements. United States Census Bureau. https://www.census.gov/library/visualizations/interactive/parents-living-arrangements.html. Accessed: 16 Jan. 2026

[23] Schondelmyer, E.: No Kids in the House: A Historical Look at Adults Living Without Children. United States Census Bureau. https://www.census.gov/newsroom/blogs/random-samplings/2016/12/no_kids_in_the_house.html. Accessed: 16 Jan. 2026

[24] Schondelmyer, E..: Fewer Married Households and More Living Alone. United States Census Bureau. https://www.census.gov/library/stories/2017/08/more-adults-living-without-children.html. Accessed: 16 Jan. 2026

[25] Yau, N.: Oldest, Youngest, and Middle Children, in Differently Sized U.S. Households. FlowingData. https://flowingdata.com/2022/03/09/kid-order/. Accessed: 13 June 2026 (2022)

[26] Brenan, M.: Americans' Preference for Larger Families Highest Since 1971. Gallup, Inc. https://news.gallup.com/poll/511238/americans-preference-larger-families-highest-1971.aspx. Accessed: 13 June 2026 (2023)