Journal of Theoretical Biology 581 (2024) 111721

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Who acquires infection from whom? A sensitivity analysis of transmission dynamics during the early phase of the COVID-19 pandemic in Belgium

Leonardo Angeli [a,b,*], Constantino Pereira Caetano [d,h], Nicolas Franco [a,f], Steven Abrams [a,b,g], Pietro Coletti [a,b], Inneke Van Nieuwenhuyse [b,e], Sorin Pop [b], Niel Hens [a,b,c]

[a] Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Hasselt, Belgium
[b] Data Science Institute (DSI), Hasselt University, Hasselt, Belgium
[c] Centre for Health Economics Research and Modelling Infectious Diseases, Vaxinfectio, University of Antwerp, Antwerp, Belgium
[d] Departamento de Epidemiologia, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Lisbon, Portugal
[e] Computational Mathematics, Hasselt University, Hasselt, Belgium
[f] Namur Institute for Complex Systems (naXys) and Department of Mathematics, University of Namur, Namur, Belgium
[g] Global Health Institute (GHI), Department of Family Medicine and Population Health, University of Antwerp, Antwerp, Belgium
[h] Center for Computational and Stochastic Mathematics, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal

ARTICLE INFO

Dataset link: https://github.com/LeoAngeliTR/WAIFW_Angeli2023/tree/main, https://zenodo.org/record/7788684

Keywords:
Sensitivity
Next generation operator
Basic reproduction number
COVID-19
Age-structured models

ABSTRACT

Age-related heterogeneity in a host population, whether due to how individuals mix and contact each other, the nature of host–pathogen interactions defining epidemiological parameters, or demographics, is crucial in studying infectious disease dynamics. Compartmental models represent a popular approach to address the problem, dividing the population of interest into a discrete and finite number of states depending on, for example, individuals' age and stage of infection. We study the corresponding linearised system whose operator, in the context of a discrete-time model, equates to a square matrix known as the next generation matrix. Performing formal perturbation analysis of the entries of the aforementioned matrix, we derive indices to quantify the age-specific variation of its dominant eigenvalue (i.e., the reproduction number) and explore the relevant epidemiological information we can derive from the eigenstructure of the matrix. The resulting method enables the assessment of the impact of age-related population heterogeneity on virus transmission. In particular, starting from an age-structured SEIR model, we demonstrate the use of this approach for COVID-19 dynamics in Belgium. We analyse the early stages of the SARS-CoV-2 spread, with particular attention to the pre-pandemic framework and the lockdown lifting phase initiated as of May 2020. Our results, influenced by our assumption on age-specific susceptibility and infectiousness, support the hypothesis that transmission was only influenced to a small extent by children in the age group $[0, 18)$ and adults over 60 years of age during the early phases of the pandemic and up to the end of July 2020.

1. Introduction

In epidemiological studies, unravelling how population heterogeneity influences disease spread is typically critical. Structured compartmental models tackle this problem by dividing the population into a finite number of discrete groups, referred to as *population state variables* (Metz and Diekmann, 2014). The underlying rationale consists in identifying the individual with the subpopulation to which he or she belongs, defined by the *individual state variable* - typically age, gender, occupation or nationality. Assuming that all individuals are embedded in the same epidemiological context and that individuals in the same states equally participate in transmission, the corresponding population state variables characterise the distribution of the number of individuals in each compartment, i.e. the population distribution. Metz and Diekmann (2014); see Appendix B.1.

In this study, we use a *multistate model* as the reference model, as individuals composing the observed population are identified with a bi-dimensional state (age, progression of the disease). The corresponding system of differential equations and the domain of the parameter functions modelling transmission processes and transition across epidemiological states can be of high dimension. However, under suitable assumptions, compartmental models of arbitrary complexity are effectively linked with a discrete, linear dynamical system matrix while maintaining a consistent epidemiological interpretation (Diekmann et al., 2010). This matrix is the *next generation matrix*, which

* Corresponding author at: Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Hasselt, Belgium.
E-mail address: leonardo.angeli@uhasselt.be (L. Angeli).

https://doi.org/10.1016/j.jtbi.2024.111721
Received 3 May 2023; Received in revised form 23 December 2023; Accepted 28 December 2023
Available online 11 January 2024

models the discrete-time evolution of the subpopulation of newly infected individuals, projecting them towards the next generation, following a demographic analogy that considers new infections as newborns (referred to as the *next generation approach*). Arguably, the main advantage of such an approach lies in the fact that we can mathematically characterise the *basic reproduction number* R_0 as the spectral radius of the next generation matrix (Diekmann et al., 1990). The next generation approach concentrates solely on the infection states immediately after infection, resulting in an operator of relatively manageable size that sheds light on how the transmission potential of the pathogen under study is linked to the interplay of the different ingredients of the model (as quantified by its R_0). Whenever a careful study of a disease's evolution is required, one must account for the impact of demographic heterogeneity. In particular, in the case of COVID-19, a large number of studies worldwide have found a marked age-dependence of epidemiological parameters (e.g., Liu et al., 2020; Davies et al., 2020; Franco et al., 2022), which underlines the importance of including age as a state variable to map the individuals of the host population, resulting in an age-structured model.

In our work, we explore the potential of a formal sensitivity analysis on the next generation operator. In Section 2, we describe how to apply the methodology to the deterministic version of the discrete-time age-structured compartmental model that was developed for modelling the SARS-CoV-2 dynamics in Belgium (Abrams et al., 2021). The proposed approach exploits the functional dependence between the generic entries of the next generation matrix ($\mathbf{K}$) to derive age-specific indices of sensitivity (Caswell, 2000). We discuss the epidemiological significance of the derived quantities, defining the strengths and limitations of this approach in the study of infectious diseases. Next, we briefly introduce the data used to inform the epidemiological parameters of our model, and then, in Section 3, we focus on interpreting the sensitivity analysis results at different time points during the COVID-19 pandemic in Belgium. Our results, revealing that from April to July 2020, children under 18 years and adults over 60 years played only a marginal role in transmission, are consistent with other studies examining the role of age structure in virus spread. (e.g. Davies et al., 2020; Monod et al., 2021; Lovell-Read et al., 2022).

Finally, in Section 4, we discuss the prospects of this approach as a tool to study infectious disease spread, even in real-time, during future pandemics. The ability of the proposed approach to produce robust results depends on the accuracy of the estimated epidemiological parameters and, thus, on the availability of relevant data. A particular challenge we encountered pertains to accurately estimating the susceptible portion of the population across different age groups. This issue is especially pronounced in Belgium, where the absence of comprehensive serological studies during the pandemic has led to a lack of accurate estimates. In general, there is also difficulty in linking serological data to the actual susceptibility of specific age groups. Owing to these limitations, our study relied on a simulation approach, as elaborated in Appendix E. Although valuable, this approach brings forth inherent constraints that necessitate caution, particularly when considering extending our study to different contexts and timeframes. As such, the age-specific conclusions drawn in this article should be interpreted with caution.

2. Materials and methods

2.1. Notation and terminology

The next generation matrix (NGM or $\mathbf{K}$) is a fundamental tool in epidemiological models. It projects one generation of infected individuals to the subsequent generation, as its entries k_{ij} quantify the expected number of secondary infections in group i produced by a single infected individual in group j during their infectious period. In general, the groups are determined by the possible states of a newly infected individual: in our case, these are the n age intervals that we consider to stratify the population. Appendix B.1 provides an intuitive way of deriving $\mathbf{K}$, based on the epidemiological significance of the components of an SIR model (*Susceptibles–Infected–Recovered*). However, for more complex ordinary differential equations (ODE) systems, a more rigorous derivation is necessary (Diekmann et al., 2010, 2012).

This section presents the methodology for an SIR model with a discrete age structure of n age classes and introduces the required notation and terminology. In the first step, we restrict the analysis to the groups of infected individuals (*infected states*), and model the transmission process analogous to a demographic growth process by considering newly infected individuals as the offspring of infectious individuals and replacing the age-specific fertility rates by the parameters modelling virus transmission (see Sections 2.4 and 2.5). We assume that the population is fully susceptible and that the age distribution of susceptible individuals reflects the demographic age distribution. Mathematically, this assumption corresponds to a linearisation around the *infection-free equilibrium* of the original system of ODE (see Eq. (B.1) in Appendix B.1), selecting only the differential equations that describe the dynamics of the infected subpopulation. The resulting *infected subsystem*, composed of $n \times m$ equations (where m is the number of compartments of infected individuals), yields the *NGM with large domain* (Diekmann et al., 2010) given by

$$\mathbf{K_L} := -\mathbf{T}\boldsymbol{\Sigma}^{-1}. \tag{1}$$

$\mathbf{K_L}$ satisfies the following equality

$$\rho(\mathbf{K_L}) = \rho(\mathbf{K}),$$

where $\rho(\cdot)$ is the *spectral radius* of a matrix (see (3)), and $\mathbf{K}$ is the NGM introduced earlier.

Eq. (1) defines the *transmission matrix* $\mathbf{T} = (t_{ij})$ as the matrix of transmission rates (i.e., new case rates) for each of the states that an individual can be in immediately after infection(*states-at-infection*). The *transition matrix* $\boldsymbol{\Sigma} = (\sigma_{ij})$ contains the transition rates of infected individuals through the epidemiological classes. It can be shown that the arbitrary entry in the matrix $-\boldsymbol{\Sigma}^{-1}$ reflects the expected time that an individual who is currently in infected state j will spend in infected state i (Diekmann et al., 2012). In the case of a SIR, the infected states coincide with the potential states-at-infection, so $\mathbf{K_L} \equiv \mathbf{K}$. This is not true in general, yet it is always possible to define a transformation $\pi(\cdot)$ such that $\pi(\mathbf{K_L}) = \mathbf{K}$ (see Eq. (B.12) in Appendix B.2.2).

We examine the growth of the newly infected subpopulation over successive generational steps. This approach enables us to focus on the states-at-infection and to disregard information on transmissions from transient infected states. Hence, $\mathbf{K}$ contains all the required information for our purposes and is the core subject of our analysis. For a detailed derivation of the mathematical operators that support this claim, we refer to Appendix B.2.

The linearisation performed to derive the infected subsystem (and the NGM) breaks down when the *basic reproduction number* $R_0 = \rho(\mathbf{K}) > 1$ (see, e.g., Diekmann et al., 2010). Concurrently, $\rho(\mathbf{K})$ governs the stability around the origin of the discrete dynamical system described by $\mathbf{K}$, enabling the demographic analogy. Denoting each generational step by t_m, and starting at time t_0 from a population of infected $\mathbf{I}_{(t_0)}$, we have:

$$\mathbf{I}_{(t_1)} = \mathbf{K}\,\mathbf{I}_{(t_0)} \quad \text{and} \quad \mathbf{I}_{(t_{m+1})} = \mathbf{K}\mathbf{I}_{(t_m)} \implies \mathbf{I}_{(t_{m+1})} = \mathbf{K}^m\mathbf{I}_{(t_0)}. \tag{2}$$

After m generations, the size of the infected population can be approximated by $\|\mathbf{K}^m\mathbf{I}_{(t_0)}\|$, that is the average per-generation growth factor is $\|\mathbf{K}^m\|^{\frac{1}{m}}$, and on the long run

$$\lim_{m\to\infty} \|\mathbf{K}^m\|^{\frac{1}{m}} = \rho(\mathbf{K}) = \max_{i=1,\dots,n} \{|\lambda_i|\}, \tag{3}$$

where $\rho(\mathbf{K})$ is the spectral radius, i.e. the maximum of the absolute values of $\mathbf{K}$'s eigenvalues. The equality in (3) holds for any bounded linear operator, and any finite-dimensional real-valued matrix defines such an operator. When $\rho(\mathbf{K}) > 1$ the infected population is expected to

undergo an unbounded geometrical growth. When $\rho(\mathbf{K}) < 1$, the origin of the discrete system is asymptotically stable (Caswell, 2000), meaning that the number of infected individuals will rapidly decay, thereby preventing a disease outbreak. This is consistent with the definition of R_0 as the expected number of secondary infections caused by a typical infected individual in a fully susceptible population (Diekmann et al., 1990).

2.2. Mathematical model and NGM

In the present work, we focus on the SARS-CoV-2 pandemic initiated at the end of 2019 in Wuhan, China, and officially declared as a pandemic on the 11th of March 2020 (WHO, 2020). The disease dynamics are described by the deterministic formulation of the SEIR compartmental model developed by Abrams et al. (2021), which is summarised below. We assume that an individual can be in either of the epidemiological states (*Susceptible, Exposed, Infectious and Recovered*). For the infectious compartment, we further distinguish between *pre-symptomatic*, *asymptomatic* and *symptomatic* individuals; the latter are further classified into *mildly symptomatic* and *severely symptomatic*. The resulting system of ODEs is displayed below in (4), along with the corresponding compartmental diagram (Fig. 1).

$$\begin{cases} \frac{d\mathbf{S}(t)}{dt} = & -\boldsymbol{\alpha}(t)\mathbf{S}(t), \\ \frac{d\mathbf{E}(t)}{dt} = & \boldsymbol{\alpha}(t)\mathbf{S}(t) - \gamma\mathbf{E}(t), \\ \frac{d\mathbf{I}_{pre}(t)}{dt} = & \gamma\mathbf{E}(t) - \theta\mathbf{I}_{pre}(t), \\ \frac{d\mathbf{I}_{asym}(t)}{dt} = & \theta\mathbf{p}\mathbf{I}_{pre}(t) - \delta_1\mathbf{I}_{asym}(t), \\ \frac{d\mathbf{I}_{mild}(t)}{dt} = & \theta(\mathbf{1}-\mathbf{p})\mathbf{I}_{pre}(t) - (\boldsymbol{\psi}+\boldsymbol{\delta}_2)\mathbf{I}_{mild}(t), \\ \frac{d\mathbf{I}_{sev}(t)}{dt} = & \boldsymbol{\psi}\mathbf{I}_{mild}(t) - \boldsymbol{\omega}\mathbf{I}_{sev}(t), \\ \frac{d\mathbf{R}(t)}{dt} = & \delta_1\mathbf{I}_{asym}(t) + \boldsymbol{\delta}_2\mathbf{I}_{mild}(t) + \boldsymbol{\omega}\mathbf{I}_{sev}(t). \end{cases} \tag{4}$$

Our study focuses on analysing the transmission dynamics through the NGM. Hence, its formulation solely depends on the compartments of infected individuals explicitly included in the infection process under study. For this reason, we let the recovered class $\mathbf{R}(t)$ absorb individuals who were hospitalised (including those who died after hospitalisation), assuming they cannot interplay with the population causing new infections. Consequently, we do not account for the phenomenon of nosocomial infections observed during the pandemic (e.g., Du et al., 2021). In the above system, the parameters $\gamma, \theta, \delta_1, \boldsymbol{\delta}_2, \boldsymbol{\psi}$ and $\boldsymbol{\omega}$ indicate the transition rates between compartments of the model, while $\mathbf{p}$ corresponds to the probability of being asymptomatic upon SARS-CoV-2 infection.

We refer to Table 1 for an overview of the model's parameters, and the respective values that we assumed in our analysis. The boldface notation refers to age-specific vectors, and any products (or ratio) of two vectors should be interpreted as a component-wise Hadamard product, e.g., $\boldsymbol{\psi}\mathbf{I}_{mild}(t) = (\psi_1 I_{mild,1}(t), \dots, \psi_n I_{mild,n}(t))$. Hence, every differential equation corresponds to an n-dimensional subsystem describing the time evolution of the included epidemiological classes for each age group. The vector-valued function $\boldsymbol{\alpha}(t)$ is the *force of infection*, whose ith component reflects the per capita rate at which a susceptible subject of age i becomes infected (see Section 2.4 for more details). Consequently, it reflects our modelling choices with respect to how the virus is transmitted between different epidemiological compartments. More specifically, the analytical formulation of the NGM depends on how individual characteristics (age group, severity of symptoms) are assumed to influence the probability of virus transmission upon contact. We discuss these assumptions in more detail in the following two sections.

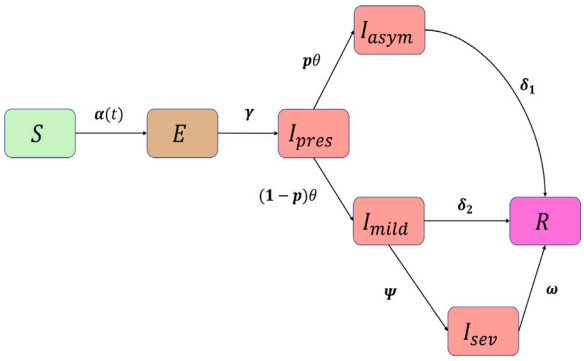


Fig. 1. Model diagram.

2.3. Social contact rates

The fundamental assumption allowing us to define the change in the number of infected hosts due to virus transmission (*transmission term*) is the so-called *social contact hypothesis* (Wallinga et al., 2006), i.e., infected individuals are assumed to generate a number of secondary infections proportional to their social contacts. The proportionality constant (denoted by q) and the type of social contacts relevant to disease spread depend on the pathogen of interest. In our case, the relevant contacts are face-to-face conversations of at least a few words or skin contact (as defined in the large-scale longitudinal survey providing Belgian data on conversational contacts Coletti et al., 2020; Verelst et al., 2021). Self-reported social contact information is cleaned and processed, and social contact rates are estimated by age group using the open-source tool SOCRATES (Willem et al., 2020). More specifically, we estimated the per capita daily contact rate (i.e., the average number of daily contacts) between an individual of age group i and individuals of age group j, for all age groups i and j:

$$m_{ij} = c_{ij} N_j, \quad \text{with } i, j = 1, \dots, n \tag{5}$$

In this expression, n is the number of age intervals, and N_j is the number of individuals of age j in the Belgian population as given by demographic data for Belgium (StatBel, 2020). In Eq. (5), the contact rates c_{ij} refine the model by capturing the inherent correlation between the age of the host and the age distribution of new infections. This goes beyond the simplistic assumption of *separable mixing*, which posits that the age of interacting individuals influences transmission independently.

The proportionality constant q may incorporate various effects, including behavioural aspects such as hand sanitation, physical distancing, susceptibility and infectiousness upon infection, as well as contact type, duration or location. Following Wallinga et al. (2010) and Franco et al. (2022), we further refine q into age-specific quantities, denoted by q_{ij}, for each combination of age groups i and j:

$$q_{ij} = \tilde{q} a_i h_j, \quad \text{with } i, j = 1, \dots, n.$$

Here $\mathbf{a} = (a_i)_{i=1,\dots,n}$ is the vector containing the age-specific probabilities of acquiring the infection given a contact (susceptible side). In contrast, $\mathbf{h} = (h_i)_{i=1,\dots,n}$ is the vector of age-specific probabilities of transmitting infection given a contact (infected side). These may reflect the age-related variability in clinical susceptibility and infectiousness, as well as other behavioural or biological aspects, such as compliance to imposed control measures or viral load of infectious individuals. To avoid confusion with standard definitions of susceptibility and infectiousness, the vectors $\mathbf{a}$ and $\mathbf{h}$ will be referred to as *q-susceptibility* and *q-infectiousness*, respectively, following the terminology suggested by Franco et al. (2022). The factor $\tilde{q}$ accounts for any residual effect, and is consistent across all age groups.

Based on the adopted theoretical definition of contact rates, the *total* number of contacts between two age groups should be uniquely determined (*reciprocal contacts*), meaning that the following conditions should hold

$$N_i m_{ij} = N_j m_{ji} \quad \text{for all } i, j \tag{6}$$

where N_i and N_j represent the individuals in age groups i and j. If the *contact rate matrix* $\mathbf{C} = (c_{ij})$ is symmetric, then the *social contact matrix* $\mathbf{M} = (m_{ij})$ is also symmetric and (6) holds. The inferred contact matrix is not always symmetric, however, due to sampling variation and reporting biases. Therefore, we will use a modified social contact matrix $\tilde{\mathbf{M}}$ with general entries

$$\tilde{m}_{ij} = \frac{m_{ij} N_i + m_{ji} N_j}{2N_i} \quad \text{for all } i, j,$$

satisfying the reciprocity condition above.

2.4. Force of infection

To define the force of infection $\boldsymbol{\alpha}$, we need to specify which infected states are also infectious. We opt for a frequency-dependent transmission as in Begon et al. (2002), where the likelihood of a susceptible person becoming infected is proportional to the disease incidence. We assume heterogeneous contact patterns but constant contact rates within each age group. We do not consider the effect of variations in population density or mixing patterns across different settings. To distinguish between symptomatic ($\mathbf{I}_{mild}$ and $\mathbf{I}_{sev}$) and asymptomatic ($\mathbf{I}_{asym}$ and $\mathbf{I}_{pre}$) cases, we account for a difference in contact rates and infectiousness for the two groups. Estimation of the relative infectiousness between symptomatic and asymptomatic individuals is not trivial though (McEvoy et al., 2021). Several studies suggest that interactions with symptomatic individuals are more likely to lead to transmission than those involving asymptomatic individuals (Li et al., 2020b; Nabi, 2020). Following Abrams et al. (2021), we accommodate such differences by assuming a homogeneous ratio of $\tau_{inf} = 0.51$, which is the ratio of the q-infectiousness of cases not displaying symptoms to those with mild to severe symptoms. We will refer to τ_{inf} as the *infectivity ratio*. Behavioural changes due to the onset of symptoms can be modelled by estimating the location-specific proportional change in contacts of symptomatic cases, indicated here by $\boldsymbol{\xi} = (\xi_{home}, \xi_{work}, \xi_{school}, \xi_{transport}, \xi_{leisure}, \xi_{other})$. We obtain the matrices of symptomatic interaction rates by weighting the location-specific matrices (available from the CoMix survey, see Coletti et al., 2020), namely

$$\begin{aligned} \mathbf{C}_{asym} &= \mathbf{C}_{home} + \mathbf{C}_{work} + \mathbf{C}_{school} + \mathbf{C}_{transport} + \mathbf{C}_{leisure} + \mathbf{C}_{other} \\ \mathbf{C}_{sym} &= \xi_{home}\mathbf{C}_{home} + \xi_{work} \cdot \mathbf{C}_{work} + \xi_{school} \cdot \mathbf{C}_{school} + \dots \\ &\quad \cdots + \xi_{transport} \cdot \mathbf{C}_{transport} + \xi_{leisure} \cdot \mathbf{C}_{leisure} + \xi_{other} \cdot \mathbf{C}_{other} \end{aligned} \tag{7}$$

The corresponding component of the force of infection, for each age group i of susceptibles, equals

$$\alpha_i = \tau_{inf} \cdot \tilde{q} a_i \sum_{j=1}^{n} c_{ij}^{asym} h_j \left(I_j^{pre} + I_j^{asym} \right) + \tilde{q} a_i \sum_{j=1}^{n} c_{ij}^{sym} h_j \left(I_j^{mild} + I_j^{sev} \right). \tag{8}$$

Due to the lack of studies on this topic at the beginning of our work, we will assume that the change in contact rates due to symptom onset is the same as observed during the 2009 A/H1N1 pandemic influenza in England (Van Kerckhove et al., 2013), namely $\boldsymbol{\xi} = (1, 0.09, 0.09, 0.13, 0.06, 0.25)$. In our previous assumptions, we did not consider the impact of symptom severity on the contact structure or the chance of transmitting the disease during contact. Therefore, we do not distinguish between pre-symptomatic and asymptomatic nor between mildly symptomatic and severely symptomatic individuals in how we model transmission during a contact. For simplicity in our sensitivity analysis, we will refer to these groups as *asymptomatic* or *symptomatic*, denoted by the superscripts $asym$ and sym, respectively.

2.5. Next generation matrix and reproduction numbers

Appendix B.2 presents a detailed derivation of the next generation operators associated with our system (4). However, we stress that for studies focussing on the analysis of the average per-generation impulse to transmission, as quantified by the basic reproduction number, it is sufficient to focus on the classical next generation operator (Diekmann et al., 1990). The mathematical reason is that we define R_0 relying on the asymptotic convergence of the finite difference equation solution (2) to a quantity proportional to $R_0^m \mathbf{w}$, where $\mathbf{w}$ is $\mathbf{K}$'s right eigenvector corresponding to R_0. The vector $\mathbf{w}$ has positive components and can be rescaled such that $\sum w_i = 1$, that is, its components w_i can be thought of as the asymptotic incidence of the infection in the states-at-infection included in the model. Such convergence is guaranteed in the case of a primitive matrix, which is not necessarily the case for the large domain operator (1) in the context of infectious disease modelling. On the other hand, if we assume that all states-at-infection are potentially contributing to transmission, the matrix $\mathbf{K}$ is positive and thus primitive. We devoted Appendix C.1 to explain the theory supporting the above claims. Hence, to derive our sensitivity measures, we focus on the components of the left and right eigenvectors corresponding to $R_0 = \rho(\mathbf{K})$. A generic entry k_{ij} of the matrix $\mathbf{K}$ is defined as:

$$\begin{aligned} k_{ij} &= k_{ij}^{asym} + k_{ij}^{sym} \\ &= \underbrace{\tau_{\text{inf}} N_i \tilde{q} a_i c_{ij}^{asym} h_j}_{\text{Transmission term}} \overbrace{\left(\frac{\delta_1 + \theta p_j}{\delta_1 \theta} \right)}^{\text{Transition term}} + \underbrace{N_i \tilde{q} a_i c_{ij}^{sym} h_j}_{\text{Transmission term}} \overbrace{\frac{(\Psi_j + \omega_j)(1 - p_j)}{\omega_j (\Psi_j + \delta_{2j})}}^{\text{Transition term}}, \end{aligned} \tag{9}$$

for all $i, j = 1, \dots, n$. The equation for k_{ij} can be interpreted as a weighted sum of two main contributions to viral transmission. The first contribution, denoted k_{ij}^{asym}, comes from interactions involving infectious individuals who are pre-symptomatic or asymptomatic. The second contribution, k_{ij}^{sym}, arises from interactions with infectious individuals who show symptoms (mild or severe). The weights are represented by what we call the *transition term*. This term includes parameters that model disease progression in infected individuals, and the average time of individuals in different compartments. It also incorporates the parameters related to how the virus spreads during contact and accounts for the number of susceptible individuals in each group. This aligns with the force of infection as defined in Eq. (8).

While the interpretation of the spectral radius of $\mathbf{K}$ as a basic reproduction number is sufficiently supported by theory, the perspective of the next generation approach entails a risk of underestimating the effect of the decrease in the susceptible population in stages that are far from the initial stage. Yet, the big advantage of the method is that it exploits the available data to reflect the impact of population heterogeneity on infection routes. In our case, the availability of age-specific estimates of epidemiological parameters (Franco et al., 2022; Abrams et al., 2021) and the evolution of the structure of contacts (Coletti et al., 2020) allows us to update the NGM progressively. Moreover, to analyse specific moments of the SARS-CoV-2 pandemic (Section 2.6), we account for the depletion of susceptibles in each age group using numerical estimates provided by the corresponding stochastic model developed by Abrams et al. (2021), and calibrated on hospital admission data by age (and on serial serological survey data collected during the initial phase of the Belgian pandemic). We then assume $\rho(diag(\mathbf{S})\mathbf{K})$ approximates the *effective reproduction number* R_t. Here $\mathbf{S}$, representing the proportion of susceptible individuals in each of the n age groups, forms the main diagonal of the diagonal matrix $diag(\mathbf{S})$. We will refer to the updated NGM as $\mathbf{K_S} := diag(\mathbf{S})\mathbf{K}$.

Table 1
Parameters used in the model, with their epidemiological interpretation. The components of parameter vectors align with the age brackets: $[0,6), [6,12), [12,18), [18,30), [30,40), [40,50), [50,60), [60,70), [70,\infty)$. The table presents the parameter values used in our analysis and the sources from which these values are obtained.

Notation	Description	Value	Context	Source
τ_{inf}	Infectivity ratio	0.51	COVID-19 - China 2020	Li et al. (2020a)
$\mathbf{p}$	Age-specific probability of an asymptomatic COVID-19 case	$(0.94, 0.92, 0.90, 0.84, 0.61, 0.49, 0.21, 0.02, 0.02)$	COVID-19 - USA 2020	Chin et al. (2021)
$\mathbf{a}$	Age-specific q-susceptibility	$(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74)$	COVID-19 - Multiple Countries 2020	Davies et al. (2020)
$\mathbf{h}$	Age-specific q-infectiousness	$(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99)$	COVID-19 - Belgium 2020	Franco et al. (2022)
c_{ij}	Age-specific daily reported contact rate	–	COVID-19 - Belgium 2020	Coletti et al. (2020)
γ	Exposed removal rate	0.729	COVID-19 - Belgium 2020	Abrams et al. (2021)
θ	Pre-symptomatic removal rate	0.475	COVID-19 - Belgium 2020	Abrams et al. (2021)
δ_1	Asymptomatic recovery rate	0.240	COVID-19 - Belgium 2020	Abrams et al. (2021)
$\boldsymbol{\delta}_2$	Age-specific recovery rate of mildly symptomatic cases	$(0.73, 0.74, 0.75, 0.74, 0.75, 0.74, 0.73, 0.72, 0.70)$	COVID-19 - Belgium 2020	Abrams et al. (2021)
$\boldsymbol{\psi}$	Age-specific rate of transition from mild to severe symptoms	$(0.021, 0.014, 0.006, 0.012, 0.010, 0.017, 0.022, 0.032, 0.050)$	COVID-19 - Belgium 2020	Abrams et al. (2021)
$\boldsymbol{\omega}$	Age-specific removal rate of severe symptoms cases	$(0.167, 0.131, 0.095, 0.099, 0.162, 0.338, 0.275, 0.343, 0.338)$	COVID-19 - Belgium 2020	Abrams et al. (2021)
$\boldsymbol{\xi}$	Location-specific behavioural changes upon symptoms	$(1, 0.09, 0.09, 0.13, 0.06, 0.25)$	H1N1 - England 2009	Van Kerckhove et al. (2013)

2.6. Sensitivity measures and epidemiology

We now illustrate the usefulness of differential sensitivity analysis within epidemiological contexts. Namely, by performing a perturbation analysis on the next generation matrix, we introduce new indices highlighting the age-specific impact on virus transmission. In the following, $\mathbf{v}$ and $\mathbf{w}$ indicate the dominant left and right eigenvectors associated with $R_t \approx \rho(\mathbf{K_S})$, satisfying the conditions $\langle \mathbf{v}, \mathbf{w} \rangle = 1$ and $\sum_i w_i = 1$. This choice simplifies notation and aids in the clear interpretation of the various sensitivity indices.

By definition, adding the entries of the jth column of the next generation matrix $\mathbf{K}$ results in the expected number of secondary infections produced by a primary infectious individual in age group j during their infectious period. Any perturbation in the parameters that constitute the matrix entries k_{ij} (for all $i = 1, \dots, n$) will influence this number. In turn, this leads to variations in the effective reproduction number R_t, affecting the overall transmission dynamics. Using derivation techniques, we can quantify this age-specific perturbation as the gradient of R_t with respect to the vectors $\mathbf{k_j}$, which represent the columns of $\mathbf{K_S}$. This forms the basis for the measures outlined below.

Cumulative sensitivities:

$$\tilde{s}_j := \left\| \frac{\partial R_t}{\partial \mathbf{k_j}} \right\|_1 = \sum_{i=1}^{n} \left| \frac{\partial R_t}{\partial k_{ij}} \right| = \sum_{i=1}^{n} \frac{v_i w_j}{\langle \mathbf{v}, \mathbf{w} \rangle} = \sum_{i=1}^{n} v_i w_j = \sum_{i=1}^{n} s_{ij} \tag{10}$$

By adding the classical sensitivities (Caswell, 2019), denoted as s_{ij} and formally derived in Appendix A.1, we obtain an aggregate measure of the R_t rate of change corresponding to a relatively small perturbation of the expected number of infections initiated by an infected individual of age j. Referring to Eq. (9), $\tilde{s}_j$ reflects the R_t rate of change resulting from (possibly concurrent) shifts in the contact behaviour (c_{ij}), the q-infectiousness (h) or modifications in any of the parameters governing the length of stay of an individual of age j in an infected state (transition term).

Cumulative elasticities:

$$\tilde{e}_j := \left\| \frac{\mathbf{k_j}}{R_t} \circ \frac{\partial R_t}{\partial \mathbf{k_j}} \right\|_1 = \sum_{i=1}^{n} \frac{k_{ij}}{R_t} \frac{\partial R_t}{\partial k_{ij}} = \sum_{i=1}^{n} \frac{k_{ij}}{R_t} s_{ij} = \sum_{i=1}^{n} e_{ij}. \tag{11}$$

In the above equation, we define the proportional response of R_t to a proportional change in a specific column of $\mathbf{K_S}$ (where $\circ$ denotes the element-wise product). Given the property that R_t—when treated as a function of the entries of $\mathbf{K_S}$—is homogeneous of degree 1 (Diekmann et al., 2012), it follows that $\sum_{j=1}^{n} \tilde{e}_j = 1$. Each of these perturbation measures can be interpreted as the local *proportional contribution* to R_t of age group j. This is *not* an independent age-specific contribution to the reproduction number, as it depends on all values in the NGM. The above expressions lead to the following key insights:

- For each age group j, the sensitivities (as given in (10)) operate as scaling factors. These translate the absolute variation in the age-specific transmission potential (following an explicit age j parameter perturbation) into the absolute change in the overall transmission potential.[1] Mathematically, this is expressed as:

$$\text{Let } \tilde{\mathbf{K}} = \mathbf{K_S} + \Delta \mathbf{K_S} \implies \rho(\tilde{\mathbf{K}}) \approx \rho(\mathbf{K_S}) + \sum_{i,j} s_{ij} \Delta k_{ij}, \tag{12}$$

  where Δk_{ij} is the expected change in the number of secondary infections in age group j generated by a single infected individual of age j. This can be calculated from (9), for instance, when considering reopening primary schools (expected to imply a certain Δm_{ij}), or the emergence of a new virus strain with modified transmissibility (expected to yield a specific Δh_j). Eq. (12) provides an approximation of the effect on R_t (See Fig. D.6(c) in Appendix).
- The elasticities in (11) rescale the sensitivities, directly accounting for the number of infections each group is expected to generate, giving a measure of R_t proportional variation corresponding to a proportional perturbation in the entries of a specific NGM column (e.g. $\mathbf{K}_j$). Such elasticities identify the overall age-specific differences in participation to transmission at a specific point in time. When several measures are in place to counteract overall virus transmission – such as social distancing, quarantine, mandatory mask-wearing, school closure, or teleworking – the optimal strategy is to configure a set of measures that leads to the highest proportional reduction in secondary infections caused by individuals in age groups with the highest elasticity index $\tilde{e}_j$. Based on (12), and considering a relatively small proportional change η_{ij} in k_{ij}, we have:

$$\frac{\Delta R_t}{R_t} \approx \sum_{i,j} e_{ij} \eta_{ij}. \tag{13}$$

[1] The epidemiological interpretation of these expressions is based on the ergodic properties of our problem. The assumption underlying the analysis is that the vector $\mathbf{w}$ aptly approximates the infection's age distribution. The sensitivities s_{ij} reveal that the impact of interactions between age groups i and j on transmission is proportional to the expected disease incidence in group j (w_j) and the potential contribution (v_i) of a single infected individual in age group i to the asymptotic number infections. See Appendix C.1 for more details.

We emphasise that the perturbations to the metrics discussed above should remain "relatively small" for the second-order errors in Eqs. (12) and (13) to be marginal. Yet, in Appendix D (refer to Fig. D.6), we show that the performance of the linear approximation of ΔR_t is robust to significant variations in the column elements of the NGM.

Lower-level epidemiological parameters:

The chain rule enables us to examine variations in R_t at a finer level, by differentiating it with respect to the epidemiological rates constituting k_{ij} (9). Typically, these are termed lower-level parameters, denoted here as l. The corresponding sensitivities are then defined as:

$$s_l = \frac{\partial R_t}{\partial l} = \sum_{i,j} s_{ij} \frac{\partial k_{ij}}{\partial l}; \tag{14}$$

These sensitivities are particularly relevant when analysing the impact of changes in specific epidemiological parameters. Sensitivities associated with different parameters can be validly compared as long as these parameters are on the same scale, e.g. a_i and h_j. Appendix A.2 provides an overview of the cumulative sensitivity measures for each parameter in our model. These aggregated measures are obtained by adding the indices from Eq. (14) across all age groups, allowing for a preliminary parameter comparison. Lower-level sensitivities permit a direct evaluation of the basic reproduction number perturbation when an epidemiological parameter is perturbed. Mathematically, this is expressed via the reformulated Eq. (12), given by $\rho(\tilde{\mathbf{K}}) \approx \rho(\mathbf{K_S}) + \sum_{i,j} s_{ij} s_l \Delta l$. The modelling choices for the transmission term in Eq. (9), which represents the product of independent contributions from interacting individuals and their social contact rate, constrain the insights we can gain from perturbation analysis. Consequently, we exclude lower-level elasticities tied to individual transmission parameters. These would not provide additional insights, as they would align with the general elasticities outlined in (11). However, the term k_{ij} in (9) is derived from a combination of two weighted contributions tied to the symptoms exhibited by the infectious individuals. Assuming that some containment strategies independently affect infected individuals depending on whether they show symptoms or not (e.g., the use of masks will be effective in reducing q-susceptibility and q-infectiousness primarily in contacts of asymptomatic and pre-symptomatic individuals, given the reduced interaction of individuals with symptoms), we can define symptom-specific measures as follows:

$$e_l^* = \frac{l^*}{R_t} \frac{\partial R_t}{\partial l^*} = \sum_{i,j} s_{ij} \frac{k_{ij}^*}{R_t}, \tag{15}$$

with $* = asym, sym$ (we refer to Section 2.4, Eq. (9) for the notation). Using these expressions, we can evaluate the proportional response of R_t to variations in one or many transmission parameters, contingent upon the infectious individuals' symptomatic (mild and severe) or asymptomatic (and pre-symptomatic) status. For instance, asymptomatic individuals, who may be unaware of their infectious status, are likely to have different adherence levels to preventative measures like hand sanitation and mask usage compared to symptomatic individuals. Thus, we can assess the distinct impact of variations in q-infectiousness($l^* = h_j$) through e_l^*. Following Caswell (2000), meaningful elasticities of R_t to lower-level parameters that govern transitions through infected states can be derived. These parameters are components of the transition terms in Eq. (9). Details are provided in Appendix A.2, Table A.4.

Sensitivity ratios:

$$r_{ij} = \frac{s_{ik}}{s_{jk}} = \frac{v_i}{v_j} \tag{16}$$

As detailed in Appendix C.1, the primitivity of $\mathbf{K_S}$ ensures that the distribution of infected individuals converges to the right dominant eigenvector $\mathbf{w}$. This vector is often referred to as *stable age distribution* of new infections and, when normalised, can be interpreted as the relative disease incidence across age groups (Caswell, 2000; Held et al., 2019). In Appendix C.2, we show that this convergence is relatively fast for the next generation matrices analysed. Consequently, $\mathbf{K_S}$'s left eigenvector $\mathbf{v}$, corresponding to R_t can be regarded as *infective values* vector. Each component v_i measures age group i's relative contribution to the asymptotic size of the infected population (see Appendix C.1). Both $\mathbf{w}$ and $\mathbf{v}$ are relative quantities defined up to a constant. Therefore, evaluating the ratios r_{ij} makes sense to quantify the relative impact that new infections in the different age groups have on the overall number of infections. A value of r_{ij} close to one indicates that age groups i and j have a comparable per capita impact on the number of infections. In contrast, a value greater (less) than one implies that infected individuals in age group i (j) have a higher per capita impact on transmission. Sensitivity ratios pinpoint the susceptible age group that, if infected, would most amplify virus transmission, irrespective of the infectee's age.

The interpretation of the introduced metrics hinges on the assumption that variations in each k_{ij} (for all i) are independent. In other words $\frac{\partial k_{ij}}{\partial k_{mj}} = 0$ for any $i \neq m$ in $1, \dots, n$. This is not necessarily true, however: for instance, it is plausible that an individual's daily contact time is finite, leading to a saturating contact rate (Heesterbeek and Metz, 1993) and a potentially negative correlation between contacts across different age groups.

2.7. Assumptions

This subsection lists the assumptions for our sensitivity analysis.

- **Age intervals:** we consider the following age groups (with ages in years):

$$\Omega = \{[0,6), [6,12), [12,18), [18,30), [30,40), [40,50), [50,60), [60,70), [70,\infty)\}.$$

  The age structure mimics the Belgian school system for individuals under 18 years of age. The notation Ω thus denotes the set of age intervals.
- **Contact rates**: for each survey wave, we consider the sample mean of the reported number of contacts as a proxy for the per capita number of contacts per day, denoted as m_{ij}, as defined in Eq. (5). The considered waves of the CoMIx social contact survey, lack data on contacts made by children under 18 years of age. Following (Van Hoang et al., 2021), we estimated these contact rates using pre-pandemic social contact data from Flanders, Belgium, and adjusted them based on the observed relative changes in reported contact rates in other age groups (available from the CoMix study). No social contact survey data were collected between the beginning of September and the beginning of November 2020.
- **Epidemiological parameters estimates:** we assume the q-susceptibility to be age-specific, following (Davies et al., 2020), namely

$$\mathbf{a} = (a_i)_{i=1,\dots,n} = (0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74). \tag{17}$$

  Assessing the age-dependence of q-infectiousness is not trivial (Lau et al., 2020; Franco et al., 2022). Following (Franco et al., 2022), we infer that the average infectiousness of individuals in each age group is proportional to the probability for an infected individual to remain asymptomatic $\mathbf{p}$, with $\mathbf{1} - \mathbf{p}$ being the probability of developing symptoms. For any age group j, we impose that $h_j = \tau_{inf} p_j + 1(1 - p_j)$. The resulting age-specific q-infectiousness profile is:

$$\mathbf{h} = (h_i)_{i=1,\dots,n} = (0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99). \tag{18}$$

- We choose to set $R_0 = 3.4$ (95% C.I. $(3.36, 3.44)$) for SARS-CoV-2 spread in Belgium, in agreement with Coletti et al. (2021), Willem et al. (2021), Abrams et al. (2021). This corresponds, in our model, to a proportionality factor of $\tilde{q} = 0.137$, which will then be rescaled for the dominant eigenvalue of $\mathbf{K_S}$ to coincide with the R_t estimated from positive PCR tests (Sciensano, 2021).

3. Results and discussion

In our analysis, we assess how age-related heterogeneity (potentially) affects virus transmission at different points in time.[2] In this section, we present and discuss the main results.

3.1. Initial phase

This phase focuses on the epidemiological setting before the start of the pandemic in March 2020 (pre-pandemic scenario). Fig. 2(a) graphically depicts the importance of interactions between individuals in the age interval $[18, 60)$. The percentage composition of R_0, corresponding to the per age-group cumulative elasticities $\tilde{e}_i$ (see Fig. 2(d)), discloses more in-depth knowledge about age-specific engagement in transmission, identifying the age groups in the range $[18, 50)$ as primary contributors to R_0. We briefly explore the effect of the variability in social contact rates, based on the CoMix survey data, by evaluating the sensitivity measures on 1000 bootstrap samples. The bootstrap-based 95% percentile confidence interval (CI) for the highest elasticity indices $\tilde{e}_{[18,30)}, \tilde{e}_{[30,40)}$ and $\tilde{e}_{[40,50)}$, are $(0.18, 0.26), (0.23, 0.30)$ and $(0.22, 0.29)$, respectively and do not allow to observe significant differences between these groups. Despite taking part in around 35% (CI $(32\%, 37\%)$) of the contacts, children under 18 years contribute to a small extent to R_t composition, i.e. 6.6% (CI $(4.3\%, 9.7\%)$); see Fig. 2(d). The sensitivity indices provide indications in line with the above observations (see Fig. F.8 in the Appendix). In Fig. 2(b), we focus on the sensitivity ratios r_{ij}, as defined in (16), thereby adding information about the susceptibles contacted by an infectious person. In particular, the age group $[40, 50)$ shows the highest reproductive value, indicating that a new COVID-19 case in this group would contribute to the asymptotic number of infections approximately 2.9 times more than a new case in the age group $[60, 70)$, and 3.8 times more than a new case in age group $[70, \infty)$. We obtain ratios close to one when comparing the age intervals in $[18, 50)$, namely $\min r_{ij} = 0.87$ and $\max r_{ij} = 1.15$ with i, j in $\{[18, 30), [30, 40), [40, 50)\}$, indicating a comparable impact of these age groups on the long term number of infections. New infections in age groups $[12, 18)$ and $[50, 60)$ have comparable infective values, which seems to contradict the conclusions from the elasticity measures. This is mainly explained by the different q-susceptibility ($a_{[12,18)}/a_{[50,60)} = 0.46$) and q-infectiousness ($h_{[12,18)}/h_{[50,60)} = 0.62$) values in the two groups, which counterbalance other epidemiological differences in demography, contact recovery rates, the probability of developing symptoms, etc. The result is that the contribution to R_t from the group $[50, 60)$ is higher than that of the group $[12, 18)$, underlining that elasticities efficiently account for all the epidemiological parameters and should be preferred over sensitivities to analyse the overall transmission pattern. The assumption expressed by (7) implies that pre-symptomatic individuals will change their contact behaviour immediately after the onset of symptoms. While this may seem unrealistic, especially in the early phases of the virus spread, the assumption helps to emphasise the weight of asymptomatic and pre-symptomatic infections in SARS-CoV-2 transmission. Calculating $\sum_j e_l^*$ for $*$=asym,sym, with e_l^* as defined in (15), we obtain an overall value of 0.90 (CI: $(0.85, 0.96)$) and 0.073 (CI: $(0.069, 0.078)$), respectively, revealing the predominant role of pre-symptomatic and asymptomatic cases in spreading the virus. The age-dependent relative index $e_{h_j}^{asym}/e_{h_j}^{sym}$ for all $j \in \Omega$, as shown in Fig. F.9 in the Appendix, highlights a big difference between individuals under 30 and over 30 years old, regardless of the distribution of contacts between age groups and the change in contact behaviour at symptoms onset. The above index is exceptionally high in the age band $[0, 18)$ with an average value of 70.5, which means that, among children, the variation in epidemiological aspects related to the infectiousness (q-infectiousness) of pre-symptomatic and asymptomatic cases impacts R_t variation 70.5 times more than the same proportional variation in symptomatic cases. Despite their limited contribution to transmission ($\tilde{e}_j$), we interpret this last result as a clear indication that efficient testing policies to detect and trace infections among children under the age of 18 (in particular, those in primary and secondary school) should have been implemented. Given the static assumptions on behavioural changes at symptoms onset, this insight remains valid throughout the observation period (April–July 2020). The symptom-specific elasticity index will follow the contact structure variation (see Fig. F.10 in the Appendix). However, aspects such as risk perception, social and physical distancing recommendations, contact tracing and quarantine of close contacts of a confirmed case would undoubtedly impact the above index.

3.2. Sensitivity analysis

We focus on the period corresponding to the start of the deconfinement strategy, which took place in Belgium as of the 4th of May 2020 (Government(BE), 2020). The data from the corresponding CoMix survey waves are used to account for changes in social contact behaviour (Coletti et al., 2020). Figs. 3 and F.13 in Appendix, show that both sensitivities and elasticities are highest for individuals between 18 and 50 years of age, thereby identifying them as the main actors of virus transmission. During April–July 2020, the cumulative elasticities are dominated by the group $[18, 30)$, with an average $\tilde{e}_j$ of 24.4% (determined as the arithmetic mean of the indices corresponding to CoMix survey waves 1 to 8, April–July 2020). In particular, Fig. 3(b) elucidates how the age groups in the interval $[18, 50)$, from April to May 2020, are sustaining transmission, as they consistently show the highest elasticity values and highest number of daily contacts (this is also confirmed by Fig. F.13(b) Appendix F). This observation gains significance in connection with the age-specific increase in daily contacts detected at the end of May, corresponding to the reopening of industries, business-to-business, healthcare-related businesses and a relaunch of public transport (Government(BE), 2020) (see Table 2 and Fig. B.5 in the Appendix). Abrams et al. (2021) thoroughly explored the effect of different exit strategies, showing how similar percentage increases in daily per capita contact rates result in a resurgence of COVID-19 infections.

The elasticity measures suggest that children under 18 years and adults over 60 years had a marginal role in transmission. The sensitivity measures for adults over 60 are, however, 2.35 times higher on average than those for individuals under 18 years old.

From this perspective, the gradual resumption of school activities (starting on May 18) has likely had a limited impact on COVID-19 transmission. This finding is supported by several studies, which estimated that school closure policies reduce transmission by 5.6% (95% CI 4.1–6.9) (Viner et al., 2020) and would have had a minor impact on the chances of an outbreak as compared to interventions aimed at reducing contacts in the work environments or in other context (Lovell-Read et al., 2022). Other studies highlighted the low proportion of paediatric cases originating from the school environment (Rajmil, 2020; Zhu et al., 2021). On the other hand, the relatively high sensitivity values for individuals older than 60, mainly determined by relatively high susceptibility and infectiousness, would have called for caution in implementing deconfinement strategies, especially in light of the high number of deaths and hospitalisations in these age groups (Abrams et al., 2021).

[2] More precisely, we describe the average age-specific transmission pattern over a few projections of the NGM, from the moment we update the NGM structure. Assuming an approximately steady (no major perturbations) epidemiological framework, our analysis has the highest possible accuracy when the age-specific incidence of the disease is close to convergence to $\mathbf{w}$. In Appendix C.2 we explain how this is related to the ratio between the largest $\rho(\mathbf{K_S})$ and the second largest eigenvalue of the NGM, the *damping ratio*. As evident from Fig. B.5 and Table C.5, three iterations (three generation intervals, approximately corresponding to 15 days Ganyani et al., 2020) of the next generation matrix are sufficient in the early phases of the pandemic to achieve a satisfactory convergence (≈1% error).

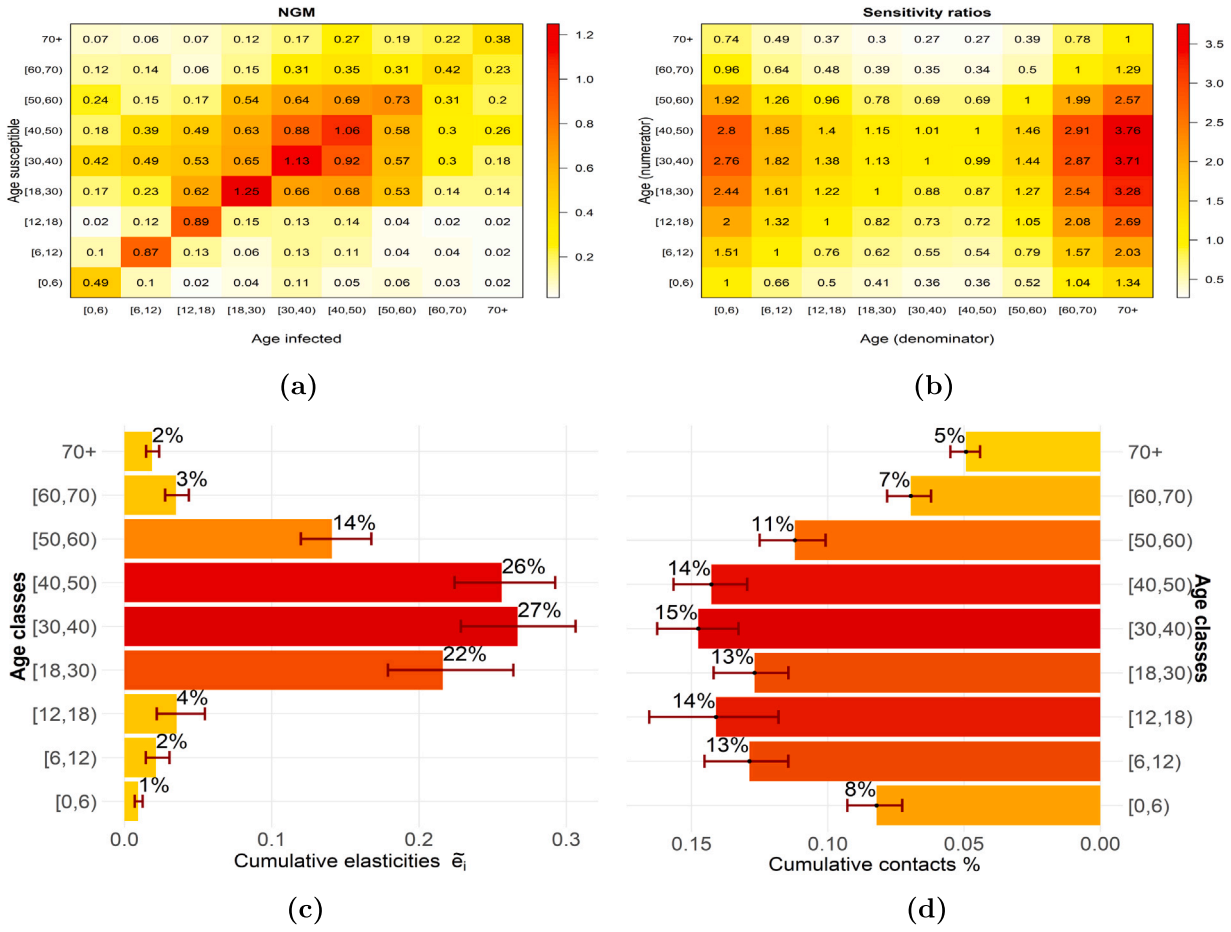


Fig. 2. Results corresponding to pre-pandemic social contacts data (Van Hoang et al., 2021). In panel (a), we display the next generation matrix and in panel (b) the sensitivity ratio matrix $(r_{ij})_{i,j} = (v_i/v_j)_{i,j}$. In panel (c), the elasticity measures $\tilde{e}_i$ show the percentage contribution of each group to R_0's absolute value; the percentage of the overall daily contacts reported on average by an individual in each age group is presented in panel (d). The bar's length corresponds to the mean of the reported social contacts, while the error bars indicate the 95% percentile interval obtained from 1000 bootstrap samples of the contact matrix.

Table 2
Age-specific percentage increase in daily contacts ($\Delta_i\% = \frac{\sum_j m_{ij}^{wave3} - \sum_j m_{ij}^{wave2}}{\sum_j m_{ij}^{wave2}}$) as reported by CoMix survey participants during wave 3 (May 22, 2020), as compared to the average number of contacts reported during survey wave 2 (May 5, 2020). We indicate the difference between the sample means and the correspondent 95% percentile interval obtained from 1000 bootstrap samples (in brackets).

Age	[0, 6)	[6, 12)	[12, 18)	[18, 30)	[30, 40)	[40, 50)	[50, 60)	[60, 70)	70+
$\Delta_i\%$	53% (30, 80)	48% (27, 72)	40% (20, 62)	34% (13, 59)	47% (26, 70)	42% (22, 64)	39% (16, 64)	80% (43, 122)	192% (77, 330)

3.3. Sensitivity ratios

The above sensitivity measures help summarise the epidemiological framework and quantify the overall contribution of specific age groups to transmission. With the term *contribution*, we summarise the ability to infect (infected side) and the chances to get infected (susceptible side) of an individual, as well as the number and the pattern of interactions in the population (social contacts). We can disentangle this information on both sides of transmission by separately looking at the components of the right and left dominant eigenvectors of $\mathbf{K_S}$, $\mathbf{w}$ and $\mathbf{v}$ respectively. While $\mathbf{w}$ can be used to estimate the relative incidence (Held et al., 2019), we focus on the susceptible side, exploring the patterns returned by the sensitivity ratios r_{ij} (16). These give a relative measure of the age-specific infective value, i.e., the relative impact of a single infection in a specific age group, on the asymptotic size of the infected population Appendix C.1.

By collecting the various r_{ij} in a matrix $\mathbf{R}^{(t_k)} \in \mathbb{R}^{n\times n}$, for a specific observation time t_k corresponding to survey wave k, we obtain a snapshot of the infective value of a single new infection in age group i compared to that of a newly infected person of age j (Fig. 4). Consistently high values (red) on the columns (resp. rows) identify groups having the lower (resp. higher) impact on the size of the infection at convergence. During the lockdown, low infective values are recorded for age groups $[0, 6)$ and 70+, while the remarkable increase in contacts observed in late May 2020 (Table 2) progressively causes the new infections in age band $[18, 50)$ to have the biggest influence on the asymptotic size of the infection.

Sensitivity ratios tell us which susceptible individual group would increase virus transmission the most when infected, regardless of the age of the infectee. One can thus use the $\mathbf{R}^{(t_k)}$ matrices to identify desirable targets for susceptibility reduction interventions, such as the use of personal protective equipment and, of course, vaccination, in view of diminishing the overall number of infections. Further insights can be obtained by considering the sensitivity ratios alongside the contact structure returned by the CoMix surveys, Fig. F.11 in Appendix. For example, a hypothetical increase in infections among children aged under 18, would be less concerning given the low infective value, the prevalent intra-specific contact patterns (meaning individuals mostly

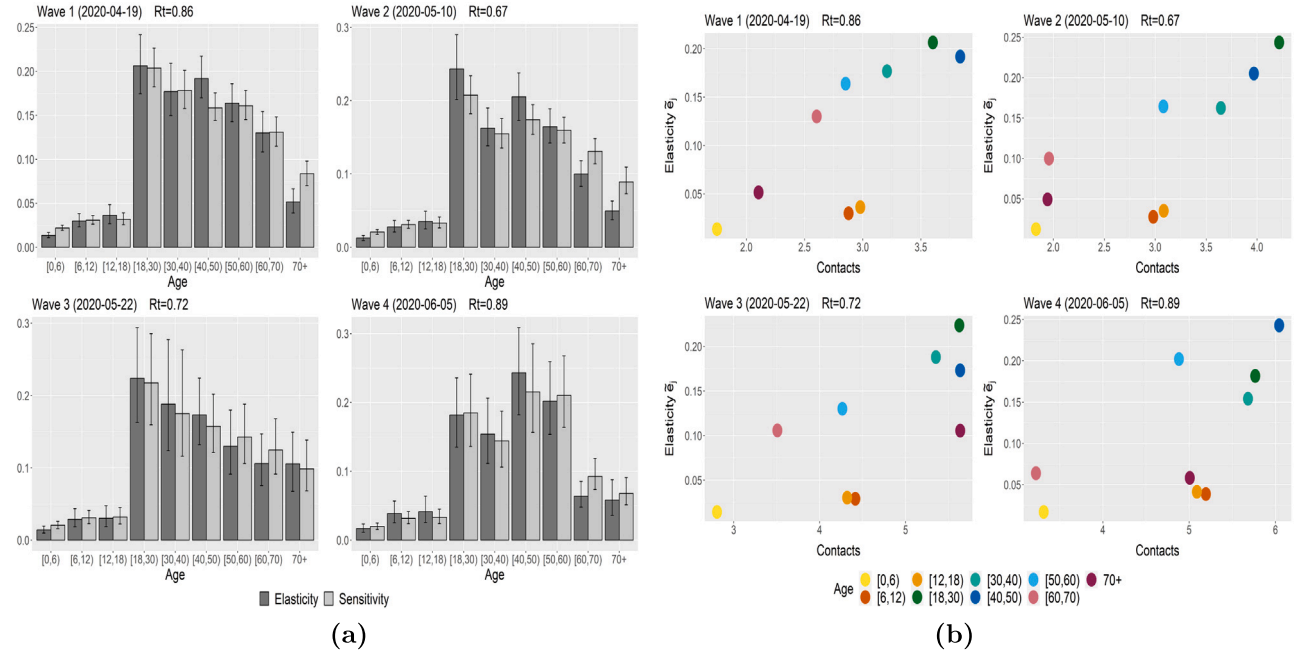


Fig. 3. Age-specific sensitivity and elasticity evolution over the period April–June 2020. Panel (a) represents the evolution of the cumulative indices $\bar{s}_j$ and $\bar{e}_j$ for all $i, j = 1, \ldots, n$. The error bars reflect the 95% percentile interval obtained from 1000 bootstrap samples of the social contact matrices. Panel (b) consists of consecutive plots comparing cumulative contacts (x-axis) and cumulative elasticities (y-axis). The age groups in the top-right portion of the graphs contribute the most to transmission at the observation time. This corresponds to different CoMix survey waves (Coletti et al., 2020), for which we indicate the calendar time and the R_t obtained from Sciensano (2021).

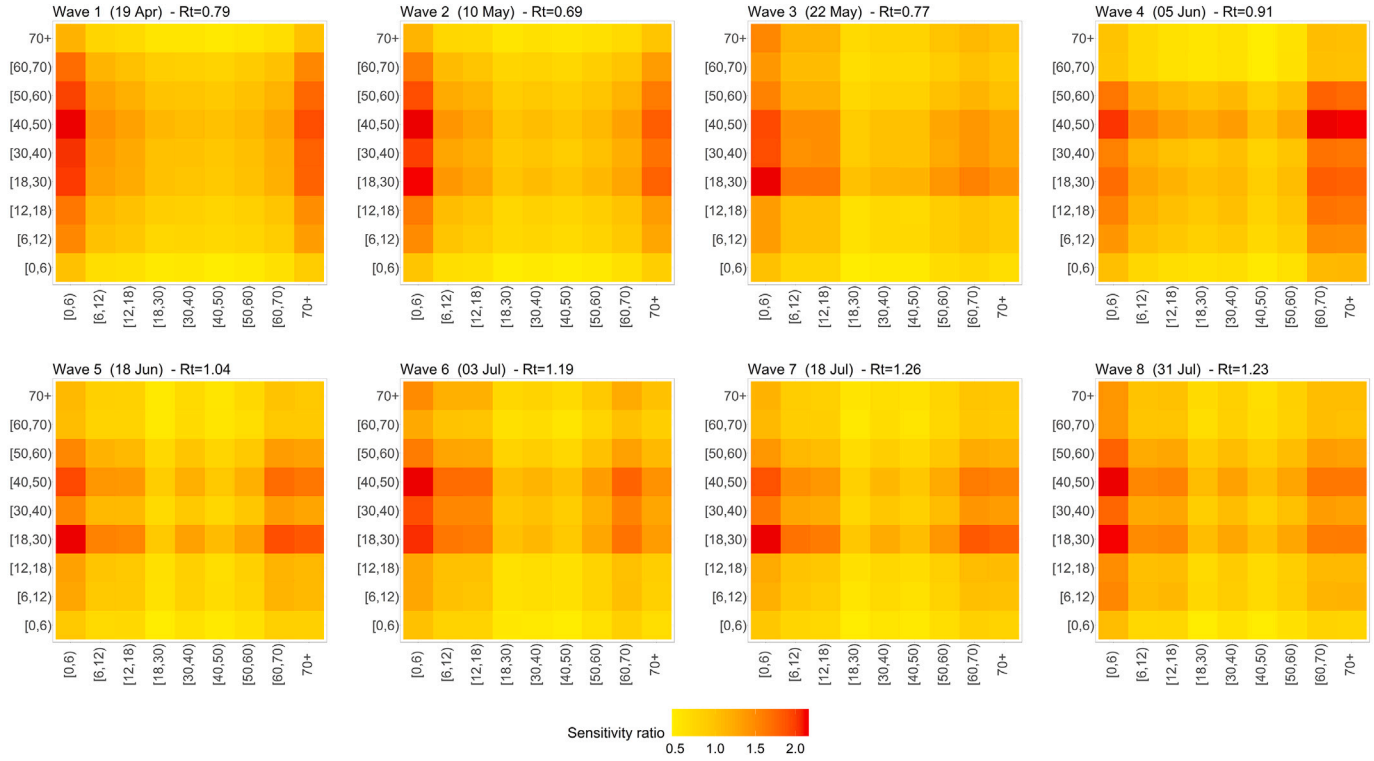


Fig. 4. Sensitivity ratio matrices $\mathbf{R}^{t_k}$. The figure displays how the matrix of sensitivity ratios changes after each data collection interval (or survey wave). The associated calendar dates are indicated, as well as the effective reproduction number shown above each matrix.

interact within their own age group) and the high rate of asymptomatic cases in this age range. However, this could represent an issue if contact patterns shifted to be more inter-specific (increased interactions across different age groups) perhaps due to a relaxation in control measures.

4. Conclusions

This paper presents the mathematical foundations to analyse the relationship between a sophisticated compartmental model and its

corresponding next generation operator. We accomplish this through formal perturbation analysis, aiming at exploring the inherent complexities of temporal disease dynamics. Our work is inspired by the theory developed by Caswell (2000) and Keyfitz and Caswell (2005) in demography and biology. However, coupling formal sensitivity analysis with the next generation approach in epidemiology is relatively novel and holds significant unexplored potential. To our knowledge, the sensitivity and elasticity indices we introduce here are new in their epidemiological interpretation, offering further insights into age-specific impacts on disease transmission. In the case of COVID-19 in Belgium, these sensitivity measures serve as a powerful lens to summarise the intricate relationships among age subgroups and their influence on SARS-CoV-2 transmission. Our approach is versatile enough to assimilate a large volume of data and offers a weighing mechanism for age-specific contributions to viral spread. We focus on the estimated impacts on the reproduction number, R_t, and, more broadly, on the eigenstructure of the Next Generation Matrix (NGM).

We developed different sensitivity indices that shed light on age-specific participation in the early spread of COVID-19. The cumulative sensitivities defined in (10) combine the expected incidence of the disease in a specific group j (w_j) and the impulse to the long-term size of the infection exerted by a single infected individual from each other age group i (v_i). This results in an index that gauges the effect of behavioural, epidemiological, and environmental changes on the reproduction number. Such indices can assist in quantifying the effect of precise changes in the behaviour of a specific group, e.g. a targeted school closure. The elasticities $\tilde{e}_j$ rescale the sensitivities, directly factoring in the number of infections each group is expected to generate, giving the proportional variation of R_t. Elasticities allow comparing the age-specific contribution to transmission at the observation time and in the subsequent generation intervals. In designing a strategy to lower overall virus transmission, the focus should be on the age groups with the highest elasticity indices, as these are most effective in converting the imposed changes into significant R_t reductions.

By combining the above indices with the available information on the number and structure of contacts, we observed that from April to the end of July 2020, the main contribution to the spread of the virus came from the 18–50 age group. The age group $[18, 30)$ marks a significant contribution to transmission at the start of the three-phase deconfinement strategy implementation in May 2020, consistently reporting higher values of both elasticity and cumulative contacts. Furthermore, using lower-level derivatives (as defined in (15)) emphasises the significant importance of infections generated by pre-symptomatic and asymptomatic cases, with the recovery rate of asymptomatic infections (δ_1) being the one with the highest overall sensitivity index in modulus (and negative), see Appendix A.2. This underscores the significance of an effective screening and contact tracing campaign at this stage. Finally, sensitivity ratios still point to the age group $[18, 50)$ as the desirable target for interventions to reduce susceptibility to the virus (e.g., through vaccination) to control the total number of infections.

Our method evidently has limitations. Two successive linearisations reduce the accuracy of the resulting measures. The first linearisation occurs when deriving the NGM and assuming that the entire population is susceptible to the virus, i.e., that the number of infections is negligible across all the age classes considered. The second one occurs when approximating the linear response of R_t with respect to the perturbation of a group of NGM entries (k_{ij}), assuming that they vary independently from each other, i.e., the R_t is considered as a multilinear function of the NGM entries. The first linearisation becomes invalid when the reduction in susceptibles causes a significant change in age-specific transmission rates, causing differences in disease incidence among various age groups compared with the early stages of infection. Nonetheless, in the present study, we dealt with the problem, considering the variation of susceptible subjects due to the natural course of the pandemic through numerical simulations (Abrams et al., 2021). The susceptibility profile remains relatively stable throughout the observation period, even if several aspects undermine the accuracy of our simulations. We refer to Appendix E and Fig. E.7 of Appendix. The second linearisation becomes invalid when the perturbations are large, and when the change in the NGM entries affects its eigenstructure. In such cases, the second-order terms might cause severe changes in the relative contributions of the different age groups to transmission, as defined by the linear analysis. This limits the relevance of our derivations to the moment of the observation, assuming that no major perturbations of the epidemiological parameters occur. Furthermore, as pointed out at the end of Section 2.6, the assumption of independent perturbations of the NGM's entries may not always hold true. For example, a policy aimed at reducing social contacts could simultaneously affect contact frequencies across different age groups, leading to correlated changes.

In addition, the sensitivity analysis yields more accurate results when the distribution of the infected in the different age classes is closer to that described by the dominant right eigenvector of the NGM. As explained in Appendix C.2, this depends on the convergence speed related to the absolute value of the second largest eigenvalue of $\mathbf{K_S}$. In our case, we observe a satisfactory convergence after three generation intervals (about 15 days). Moreover, the NGM of a compartmental system may be non-normal; that is, perturbations in the direction of the subdominant eigenvalues could produce significant oscillation while converging to the stable age distribution of the infected population, and slow down convergence.

Another limitation of our model is its dependence on the quality of data informing the next generation matrix, particularly regarding age-specific susceptibility. As detailed in Appendix E, accurately estimating age-specific susceptibility, especially for children, poses significant challenges. The initial susceptibility estimates in our model, derived from the stochastic model by Abrams et al. (2021), were based on early serological data. This data had limited representation of younger ages, leading to considerable uncertainty in these estimates. Further calibrations of the model were based on the number of hospitalisations and deaths, with minimal involvement of children in these metrics, making it challenging to capture their susceptibility accurately. Moreover, correlating seroprevalence directly with infection susceptibility adds complexity to the analysis. These factors highlight the necessity for comprehensive, longitudinal serological data to enhance the reliability and applicability of epidemiological models like the one we employed. Notwithstanding the robustness of the results, we advise caution in interpreting the age-specific findings of this study, particularly concerning younger age groups.

In conclusion, combining the next generation approach and sensitivity analysis provides valuable insights into the dynamics of infectious diseases. Its ability to generate interpretable results and accommodate a large amount of heterogeneity makes it a valuable instrument for decision-making processes. We encourage further exploration of its capabilities using more complex compartmental models, such as those that include vaccinated and hospitalised compartments or regional variations. Additionally, sensitivity analysis can be used to study the response of other epidemiological quantities to perturbations in transmission patterns. For example, we could examine the impact of perturbations on the long-term incidence of infection and the speed at which the infected subpopulation converges to the stable age distribution ($\mathbf{w}$). Overall, this tool has great potential to enhance our understanding of infectious disease dynamics and to inform public health policies.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme - project EpiPose (Grant agreement number 101003688). This work reflects only the authors' view. The European Commission is not responsible for any use that

may be made of the information it contains. This project was supported by the ESCAPE project (101095619), funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them. The authors acknowledge funding from the Special Research Fund through the Methusalem project BOF08M01 - phase III. Constantino Pereira acknowledges the funding by Fundaçãoi para a Ciência e a Tecnologia (FCT), Portugal with grant number 2020.10172.BD.

CRediT authorship contribution statement

Leonardo Angeli: Conceptualisation, Formal analysis, Methodology, Project administration, Software, Visualisation, Writing – original draft, Writing – review & editing. **Constantino Pereira Caetano:** Conceptualisation, Methodology, Writing – original draft, Writing – review & editing. **Nicolas Franco:** Writing – original draft, Writing – review & editing. **Steven Abrams:** Writing – original draft, Writing – review & editing. **Pietro Coletti:** Conceptualisation, Data curation, Formal analysis, Software, Writing – original draft, Writing – review & editing. **Inneke Van Nieuwenhuyse:** Conceptualisation, Supervision, Writing – original draft, Writing – review & editing. **Sorin Pop:** Conceptualisation, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Niel Hens:** Conceptualisation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All computer code used in this paper was written in R version 4.3.0 (2023-04-21 ucrt), and is available at the following GitHub repository: https://github.com/LeoAngeliTR/WAIFW_Angeli2023/tree/main. The social contact data utilised in this survey can be found on Zenodo: https://zenodo.org/record/7788684.

Acknowledgements

The authors thank Lander Willem for his expert assistance with the numerical simulations and for providing valuable insights. We also thank the reviewers for their constructive feedback, which helped improve the quality of the manuscript.

Appendix A. Sensitivity indices: further details

A.1. General sensitivity measures

In Section 2.6, we discuss the aggregated sensitivity measures (10). These measures are founded on the established result that, any eigenvalue λ of a matrix $\mathbf{K}$ is associated with left and right eigenvectors. Denoted here as $\mathbf{v}$ and $\mathbf{w}$ respectively, these vectors are essential in understanding the existing relationship between $\mathbf{K}$ and λ. These vectors solve the following equations:

$$\mathbf{K}\mathbf{w} = \lambda\mathbf{w} \tag{A.1a}$$

$$\bar{\mathbf{v}}\mathbf{K} = \lambda\bar{\mathbf{v}} \tag{A.1b}$$

where $\bar{\mathbf{v}}$ indicates the complex conjugate of $\mathbf{v}$. In Appendix C.1, we explain that for a matrix whose entries are strictly positive, like our NGM ($\mathbf{K}$ or $\mathbf{K_S}$), the dominant eigenvalue (R_t) is real, positive, algebraically simple and admits eigenvectors $\mathbf{w}$ and $\mathbf{v}$ with strictly real and positive components. Indicating $\lambda = R_t$ and following Keyfitz and Caswell (2005), we differentiate Eq. (A.1a) to derive

$$\mathbf{K}(d\mathbf{w}) + (d\mathbf{K})\mathbf{w} = R_t d\mathbf{w} + (dR_t)\mathbf{w}.$$

From the scalar product $\langle , \rangle$ of both sides of the above equation with the eigenvector $\mathbf{v}$, using Eq. (A.1b), one obtains

$$dR_t = \frac{\langle (d\mathbf{K})\mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{w}, \mathbf{v}\rangle} = \frac{\mathbf{v}d\mathbf{K}\mathbf{w}}{\langle \mathbf{w}, \mathbf{v}\rangle}. \tag{A.2}$$

The matrix $d\mathbf{K}$ contains elements dk_{ij}. Hence, assuming only k_{ij} is perturbed, the matrix $d\mathbf{K}$ consists of all zeros except for the entry (i, j). This simplifies Eq. (A.2) to:

$$dR_t = \frac{v_i w_j dk_{ij}}{\langle \mathbf{w}, \mathbf{v}\rangle},$$

leading to the sensitivity metric s_{ij} after dividing both sides by dk_{ij}:

$$s_{ij} := \frac{\partial R_t}{\partial k_{ij}} = \frac{v_i w_j}{\langle \mathbf{w}, \mathbf{v}\rangle}. \tag{A.3}$$

A.2. Lower level parameters

In Table A.3, we present aggregated sensitivity measures, which are sums of the indices discussed in Section 2.6. These measures are derived by differentiating R_0 with respect to all NGM entries and epidemiological parameters in the model (4). Calculations are based on pre-pandemic social contact data (Hoang et al., 2021), assuming a stable parameter space. The table includes summed sensitivities across age groups (column "Sum"), the perturbed parameter(s), the Frobenius norm of the sensitivity matrix for each parameter, and the mean value of each perturbed parameter for comparative purposes. Following Caswell (2000), we derived the elasticity indices for the transition parameters appearing in Eq. (9) of the main text and reported them in Table A.4. The formula employed is

$$e_l = \frac{l}{R_0}\frac{\partial R_0}{\partial l} = \frac{l}{R_0}\sum_{i,j} s_{ij}\frac{\partial k_{ij}}{\partial l}. \tag{A.4}$$

Sensitivities are notably high for δ_1, $\mathbf{a}$, and $\mathbf{h}$, highlighting the potential effectiveness of identifying and isolating asymptomatic cases and interventions targeting individual susceptibility and infectiousness in lowering R_t. Elasticities, derived from formula (A.4), further illustrate the influence of variations in the duration individuals spend in the asymptomatic and pre-symptomatic states on virus transmission. This pattern does not incur significant variations during the observation period.

Appendix B. NEXT generation OPERATORS

B.1. The SIR case

We present the theory and the assumptions required to derive the next generation matrix (in the following referred to in short as NGM or $\mathbf{K}$) in the case of a simple SIR model, with the host population subdivided into n age intervals. Drawing from state theory developed by Metz and Diekmann (1986) in population ecology, the epidemiological life of an individual is described by a bi-dimensional state variable recording the progress of the infection in the host (*disease state*, or *d-state*), in this case, *susceptibles, infected and recovered*, and the correspondent age-interval (*heterogeneity state*, or *h-state*). Each individual will be fully described by the couple $(d, h) \in \Delta \times \Omega$, where Δ and Ω are the d-state space and h-state space, respectively. The h-state space dimension can easily increase when including population traits such as gender, home region, or occupation.

Table A.3
Parameter-specific summed R_0 sensitivity. The column "Norm" contains the Frobenius norm value of each sensitivity matrix. The "Value" indicates the value we set for the model parameter, corresponding to the posterior mean of the estimates obtained in Abrams et al. (2021).

Symbol	Sum	Norm	Index	Age-specific	Value	Perturbed parameter
$\tilde{s}_j$	7.31	1.07	Sensitivity	TRUE	0.32	NGM entries
a	4.46	0.91	Sensitivity	TRUE	0.67	q-susceptibility
h	4.81	1.06	Sensitivity	TRUE	0.73	q-infectiousness
m_{ij}	1.46	0.23	Sensitivity	TRUE	1.45	Reported daily contacts
m_{ij}^{asym}	1.08	0.17	Sensitivity	TRUE	1.94	"" (asym)
m_{ij}^{sym}	0.39	0.07	Sensitivity	TRUE	0.57	"" (sym)
θ	−3.28	0.7	Sensitivity	FALSE	0.475	Pre-symptomatic removal rate
δ_1	−6.61	1.57	Sensitivity	FALSE	0.24	Asymptomatic recovery rate
ψ_j	0.64	0.13	Sensitivity	TRUE	0.02	Severe birth rate
δ_2^j	−0.34	0.07	Sensitivity	TRUE	0.74	Mildly symp. recovery rate
ω	−0.08	0.02	Sensitivity	TRUE	0.22	Severe symp. removal rate
p	2.57	0.57	Sensitivity	TRUE	0.55	Probability remaining asymptomatic

Table A.4
Parameter-specific indices of R_0 elasticity.

Symbol	Sum	Norm	Index	Age-specific	Perturbed parameter
$\tilde{e}_j$	1	0.22	Elasticity	TRUE	NGM entries
e_l^{asym}	0.92	0.21	Elasticity	TRUE	k_{ij}^{asym} (9)
e_l^{sym}	0.08	0.02	Elasticity	TRUE	k_{ij}^{sym} (9)
θ	−0.46	0.1	Elasticity	FALSE	Pre-symptomatic removal rate
δ_1	−0.47	0.11	Elasticity	FALSE	Asymptomatic recovery rate
ψ_j	0.003	0.001	Elasticity	TRUE	Severe birth rate
δ_2^j	−0.07	0.01	Elasticity	TRUE	Mildly symp. recovery rate
ω	−0.01	0.001	Elasticity	TRUE	Severe symp. removal rate
p	0.39	0.1	Elasticity	TRUE	Probability remaining asymptomatic

The evolution in time of the correspondent population states, *i.e.* the vectors of the number of individuals in each *d-state*, is then described by the following system of nonlinear differential equations:

$$\frac{dS_i}{dt} = -\alpha_i(t)S_i(t)\,; \quad \frac{dI_i}{dt} = \alpha_i(t)S_i(t) - \gamma I_i(t)\,; \quad \frac{dR_i}{dt} = \gamma I_i(t)\,, \tag{B.1}$$

with i varying in $\{1,\dots,n\}$ and $S(t), I(t)$ and $R(t)$ indicate the number of susceptible, infected and removed individuals at time t. The term $\alpha_i(t)$ is the (age-specific) rate at which susceptible individuals of age i are converted to infected ones and is referred to as *force of infection*, see Section 2.4. It models transmission and determines the formulation of the next generation matrix. For instance, we consider the following formulation

$$\boldsymbol{\alpha}(t) = [\mathbf{Q}\circ\mathbf{C}]\mathbf{I}(t) \quad \text{that is} \quad \alpha_i(t) = \sum_{j=1}^{n} q_{ij}c_{ij}I_j(t) \text{ for all } i \in \{1,\dots,n\}, \tag{B.2}$$

where $\mathbf{Q} = (q_{ij})$ and $\mathbf{C} = (c_{ij})$ are $n\times n$ matrices in $(0,1)^n \times (0,1)^n$, and the symbol $\circ$ indicates the element-wise product between to matrices (or Hadamard product). The underlying assumption is that transmission events between two individuals in age i and j are proportional to their contact rate c_{ij} through a constant q_{ij} (see Section 2.3). In this case, the epidemiological definition of the next generation matrix $\mathbf{K}$, as provided by Diekmann et al. (1990), is sufficient to obtain its analytical formulation. Namely, for all $i,j \in \{1,\dots,n\}$, the general entry of $\mathbf{K}$ is defined as the expected number of newly infected individuals of age i caused by an infectious individual of age j, during its infectious period (of length $\frac{1}{\gamma}$) in a fully susceptible population. In the case of system (B.1), this number is given by:

$$\left[\frac{diag(\mathbf{S}(t))[\mathbf{Q}\circ\mathbf{C}]}{\gamma}\right]_{ij} = S_i(t)q_{ij}\frac{c_{ij}}{\gamma}. \tag{B.3}$$

If our observation happens at a time t such that we can assume $S_i(t) \approx N_i$ (with N_i the overall number of individuals in age-interval i), then Eq. (B.3) defines the NGM corresponding to system (B.1). More precisely,

$$\mathbf{K} = \frac{diag(\mathbf{N})[\mathbf{Q}\circ\mathbf{C}]}{\gamma}, \tag{B.4}$$

where $diag(\mathbf{N})$ indicates the diagonal matrix whose principal diagonal elements are N_i.

B.2. Model operators

The methodology presented in Section 2.1, developed by Diekmann et al. (2010), allows us to derive the NGM for the system described by Eq. (4). We consider only the infected states in the system, corresponding to the five different compartments of infected individuals included in the model.

The system is linearised around the infection-free equilibrium, represented by $\mathbf{P}(t) = (S_1(t),\dots,S_n(t), I_1(t),\dots,I_n(t)) = (N_1,\dots,N_n,0,\dots,0)$, where $S_i(t)$ and $I_i(t)$ denote the number of susceptible and infected individuals, respectively, in age group i at time t. This equilibrium corresponds to the beginning of the virus spread when the incidence of infection is uniformly negligible across all age groups. We get to the following subsystem:

$$\begin{aligned}
\tfrac{d\mathbf{E}(t)}{dt} &= \boldsymbol{\alpha}(t)\mathbf{N} - \gamma\mathbf{E}(t)\\
\tfrac{d\mathbf{I}_{pre}(t)}{dt} &= \gamma\mathbf{E}(t) - \theta\mathbf{I}_{pre}(t)\\
\tfrac{d\mathbf{I}_{asy}(t)}{dt} &= \theta\mathbf{p}\mathbf{I}_{pre}(t) - \delta_1\mathbf{I}_{asy}(t)\\
\tfrac{d\mathbf{I}_{mild}(t)}{dt} &= \theta(\mathbf{1}-\mathbf{p})\mathbf{I}_{pre}(t) - (\boldsymbol{\psi}+\boldsymbol{\delta}_2)\mathbf{I}_{mild}(t)\\
\tfrac{d\mathbf{I}_{sev}(t)}{dt} &= \boldsymbol{\psi}\mathbf{I}_{mild}(t) - \boldsymbol{\omega}\mathbf{I}_{sev}(t)
\end{aligned} \tag{B.5}$$

We decompose the Jacobi matrix of the system (B.5) into the sum of a *transmission matrix* $\mathbf{T} = (t_{ij})$, including the rates of emerging new cases in the exposed classes in $\mathbf{E}$, caused by each individual in the *infectious states* at time t, and a *transition matrix* $\boldsymbol{\Sigma} = (\sigma_{ij})$ describing the transition of infected individuals through the epidemiological classes. The next generation matrix with large domain is defined as $\mathbf{K_L} = -\mathbf{T}\boldsymbol{\Sigma}^{-1}$ and has the following important property

$$R_0 = \rho(\mathbf{K}) = \rho(-\mathbf{T}\boldsymbol{\Sigma}^{-1}). \tag{B.6}$$

Below, we derive the two matrices $\mathbf{T}$ and $\boldsymbol{\Sigma}$ and ultimately the next generation matrix $\mathbf{K}$.

Table C.5
Convergence test: for several survey waves, we report the speed of convergence to the stable distribution. Each column corresponds to the mean distance to the stable distribution $Err(i)$ after i iteration of the NGM. These are evaluated considering 1000 different random initial conditions ($i = 0$).

Wave	d_r	Err0	Err1	Err2	Err3	Err4	Err5	Err6	Err7	Err8
0	3.27	0.32	0.06	0.016	$3.6 * 10^{-3}$	$1.0 * 10^{-3}$	$3.8 * 10^{-4}$	$8.9 * 10^{-5}$	$2.5 * 10^{-5}$	$7.2 * 10^{-6}$
1	2.61	0.32	0.08	0.023	$7.7 * 10^{-3}$	$2.6 * 10^{-3}$	$9.2 * 10^{-4}$	$3.3 * 10^{-4}$	$1.2 * 10^{-4}$	$4.6 * 10^{-5}$
2	3.11	0.33	0.08	0.023	$7.2 * 10^{-3}$	$2.2 * 10^{-3}$	$6.7 * 10^{-4}$	$2.1 * 10^{-4}$	$6.5 * 10^{-5}$	$2.0 * 10^{-5}$
3	3.09	0.32	0.08	0.024	$8.1 * 10^{-3}$	$2.4 * 10^{-3}$	$7.8 * 10^{-4}$	$2.5 * 10^{-4}$	$8.1 * 10^{-5}$	$2.6 * 10^{-5}$
4	3.45	0.27	0.05	0.013	$3.8 * 10^{-3}$	$1.0 * 10^{-3}$	$2.9 * 10^{-4}$	$8.3 * 10^{-5}$	$2.4 * 10^{-5}$	$6.8 * 10^{-6}$
5	2.96	0.29	0.07	0.018	$5.5 * 10^{-3}$	$1.8 * 10^{-3}$	$5.9 * 10^{-4}$	$2.0 * 10^{-4}$	$6.8 * 10^{-5}$	$2.3 * 10^{-5}$
6	3.11	0.27	0.07	0.022	$6.8 * 10^{-3}$	$2.1 * 10^{-3}$	$6.8 * 10^{-4}$	$2.1 * 10^{-4}$	$6.9 * 10^{-5}$	$2.2 * 10^{-5}$
7	3.22	0.29	0.07	0.021	$6.3 * 10^{-3}$	$1.9 * 10^{-3}$	$6.1 * 10^{-4}$	$1.9 * 10^{-4}$	$5.8 * 10^{-5}$	$1.8 * 10^{-5}$
8	3.23	0.26	0.06	0.017	$5.4 * 10^{-3}$	$1.6 * 10^{-3}$	$5.0 * 10^{-4}$	$1.5 * 10^{-4}$	$4.7 * 10^{-5}$	$1.4 * 10^{-5}$

B.2.1. Transmission and transition matrices

We identify a single epidemiological stage **E** describing the possible *states-at-infection*, as many states as the number of chosen age classes (n). These are all the possible states of epidemiological newborns. The sub-populations included in the model corresponding to *states-of-infectiousness* are four, $I_{pre}, I_{asym}, I_{mild}, I_{sev}$, defining $4 \cdot n$ states. These states describe the evolution of the infection within an infected host, limited to those stages in which the infected can produce epidemiological offspring. From the viewpoint of transmission, we will expect that just the first n rows are nonnull, corresponding to the n different age stages in which exposed individuals can be, once infected. The resulting *transmission matrix* T is indeed a 5×5 block matrix where each block is a n-dimensional square matrix, whose components are all identically zero except for the blocks composing the first n rows, *i.e.*

$$\mathbf{T}[1:n,] = \left[\, \mathbf{0} \;\; \mathbf{T}_{sym} \; \mathbf{T}_{sym} \; \mathbf{T}_{asym} \; \mathbf{T}_{asym} \,\right] \tag{B.7}$$

where the used notation $[1:n,]$ indicates the first n rows of $\mathbf{T}$. Here $\mathbf{0}$ is a $n \times n$ matrix with entries all equal to zero, and the other blocks are defined as

$$\mathbf{T}_{sym} = diag(\mathbf{N}) \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,n} \\ \vdots & \ddots & \vdots \\ \beta_{n,1} & \cdots & \beta_{n,n} \end{bmatrix}_{sym} \quad \text{and} \quad \mathbf{T}_{asym} = diag(\mathbf{N}) \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,n} \\ \vdots & \ddots & \vdots \\ \beta_{n,1} & \cdots & \beta_{n,n} \end{bmatrix}_{asym} \tag{B.8}$$

being the generic transmission term defined by $\beta^{\gamma}_{i,j} = \tilde{q}^{\gamma} a_i c^{\gamma}_{ij} h_j$, with $\gamma = asym, sym$. To derive the formulation of the *transition matrix* $\boldsymbol{\Sigma}$ is a straightforward step following the definition, and it reads:

$$\boldsymbol{\Sigma} = \begin{bmatrix} -\gamma\mathbb{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \gamma\mathbb{I} & -\theta\mathbb{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \theta\, diag(\mathbf{p}) & -\delta_1\mathbb{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \theta\, diag(\mathbf{1}-\mathbf{p}) & \mathbf{0} & diag(\boldsymbol{\Psi}+\boldsymbol{\delta_2}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & diag(\boldsymbol{\Psi}) & -diag(\boldsymbol{\omega}) \end{bmatrix}, \tag{B.9}$$

where each block is a $n \times n$ matrix, $\mathbb{I}$ is the identity matrix and $\mathbf{1}$ is a n-dimensional vector whose components are all equal 1. The vector $\mathbf{1}-\mathbf{p}$ embeds the age-specific probability for a pre-symptomatic infected to develop symptoms, while all the other epidemiological parameters describe the transition rates between infected states and are presented in Table 1 of the main text. Using properties of block matrices, we can invert $\boldsymbol{\Sigma}$ and get

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} -\gamma^{-1}\mathbb{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\theta^{-1}\mathbb{I} & -\theta^{-1}\mathbb{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\delta_1^{-1} diag(\mathbf{p}) & -\delta_1^{-1} diag(\mathbf{p}) & -\delta_1^{-1}\mathbb{I} & \mathbf{0} & \mathbf{0} \\ -diag(\frac{\mathbf{1}-\mathbf{p}}{\boldsymbol{\Psi}+\boldsymbol{\delta_2}}) & -diag(\frac{\mathbf{1}-\mathbf{p}}{\boldsymbol{\Psi}+\boldsymbol{\delta_2}}) & \mathbf{0} & -diag(\frac{\mathbf{1}}{\boldsymbol{\Psi}+\boldsymbol{\delta_2}}) & \mathbf{0} \\ -diag(\frac{\boldsymbol{\Psi}(\mathbf{1}-\mathbf{p})}{\boldsymbol{\omega}(\boldsymbol{\Psi}+\boldsymbol{\delta_2})}) & -diag(\frac{\boldsymbol{\Psi}(\mathbf{1}-\mathbf{p})}{\boldsymbol{\omega}(\boldsymbol{\Psi}+\boldsymbol{\delta_2})}) & \mathbf{0} & -diag(\frac{\boldsymbol{\Psi}}{\boldsymbol{\omega}(\boldsymbol{\Psi}+\boldsymbol{\delta_2})}) & -diag(\boldsymbol{\omega}^{-1}) \end{bmatrix}. \tag{B.10}$$

Please note that multiplications and fractions of the n-dimensional vectors, in bold, are to be intended in the sense of Hadamard, *i.e.* element-wise.

B.2.2. Matrices $\mathbf{K_L}$ and $\mathbf{K}$

The next generation matrix with large domains will result in a square matrix with null elements except for the first n rows. We can further divide the non-zero rows into five n-dimensional block matrices $\mathbf{K^i_L}$, with $i = 1, \ldots, 5$. We indicate the relevant rows as follows:

$$\mathbf{K_L}[1:n,] = \left[\, \mathbf{K^1_L} \;\; \mathbf{K^2_L} \; \mathbf{K^3_L} \; \mathbf{K^4_L} \; \mathbf{K^5_L} \,\right]. \tag{B.11}$$

The eigenstructure of the operator is fully determined by the block $\mathbf{K^1_L}$, which coincides with the classical next generation matrix, $\mathbf{K}$. This can also be seen by verifying that

$$\mathbf{K} = \mathbf{U}^T \mathbf{K_{LU}} = \mathbf{K^1_L}, \tag{B.12}$$

being $\mathbf{U}$ the matrix whose columns are the vectors of the canonical basis of $\mathbb{R}^{5n}$, corresponding to the rows of $\mathbf{K_L}$ which are not identically zero, i.e.

$$\mathbf{U}^T = [\mathbb{I} \;\; \mathbf{0} \;\; \mathbf{0} \;\; \mathbf{0} \;\; \mathbf{0}].$$

Finally, the next generation operator suffices for our purposes and reads:

$$\mathbf{K} = \mathbf{K^1_L} = \mathbf{T}_{asym}\, diag\left[\frac{\delta_1 + \theta\mathbf{p}}{\delta_1\theta}\right] + \mathbf{T}_{sym}\, diag\left[\frac{(\boldsymbol{\Psi}+\boldsymbol{\omega})(\mathbf{1}-\mathbf{p})}{\boldsymbol{\omega}(\boldsymbol{\Psi}+\boldsymbol{\delta_2})}\right]. \tag{B.13}$$

Appendix C. Convergence analysis & interpretation

C.1. Age-specific infective values

The $n \times n$ next generation matrix $\mathbf{K}$ linked to our system (4) is strictly positive and thus primitive, plus its formal derivation eliminates the redundant information rendering the assumption of a singular matrix as a natural one. Hence, via the Perron–Frobenius theorem, $\mathbf{K}$'s dominant eigenvalue λ_1 is real, algebraically simple and is a maximum for its spectrum, i.e. it coincides with its spectral radius $\rho(\mathbf{K})$. Furthermore, the magnitude of the remaining eigenvalues is strictly less than λ_1, and the corresponding right ($\mathbf{w}$) and left ($\mathbf{v}$) eigenvectors are real and strictly positive. Suppose we indicate the age-specific number of newly infected individuals at time t by $\mathbf{I}_{(t)}$ and assume that our matrix is diagonalisable. In that case, the set of its right eigenvectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ constitutes a basis for $\mathbb{R}^n$. Thus we can express the initial condition $\mathbf{I}_{(0)}$, corresponding to the number of new infections at the beginning of our observation, as:

$$\mathbf{I}_{(0)} = \sum_{i=1}^{n} c_i \mathbf{w}_i, \tag{C.1}$$

for some coefficients c_i. Now, applying iteratively that $\mathbf{K}\mathbf{w}_i = \lambda_i \mathbf{w}_i$, we see that $\mathbf{I}_{(1)} = \mathbf{K}\mathbf{I}_{(0)} = \sum_{i=1}^{n} c_i \mathbf{K}\mathbf{w}_i = \sum_{i=1}^{n} c_i \lambda_i \mathbf{w}_i$ and thus

$$\mathbf{I}_{(t)} = \sum_{i=1}^{n} c_i \lambda_i^t \mathbf{w}_i. \tag{C.2}$$

The above arguments are sufficient to apply the *strong ergodic theorem* (Cohen et al., 1979), namely:

$$\lim_{t\to\infty} \frac{\mathbf{I}_{(t)}}{\lambda_1^t} = c_1 \mathbf{w_1}. \tag{C.3}$$

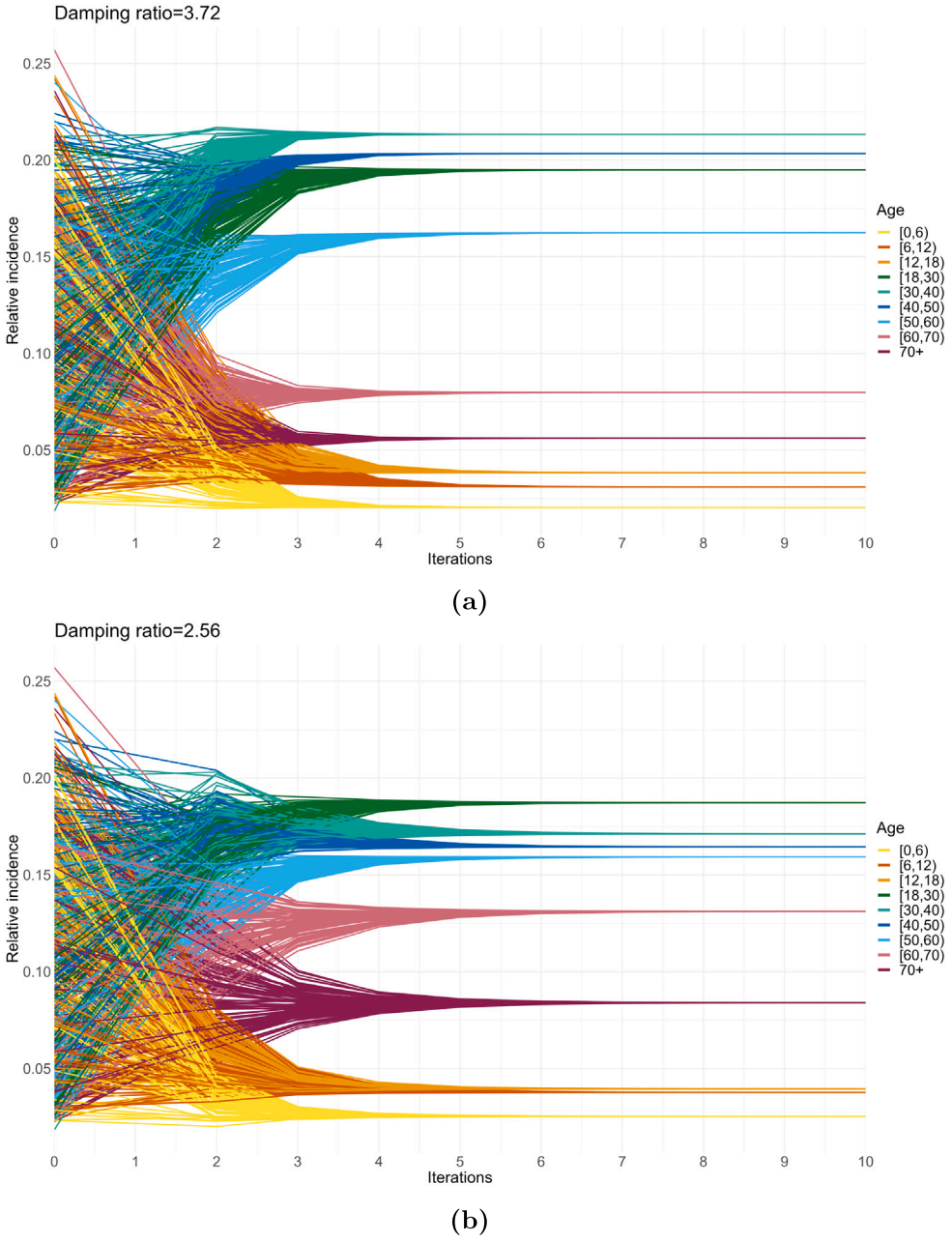


Fig. B.5. Qualitative convergence of the age-specific relative incidence to the distribution offered by the dominant eigenvalue $\mathbf{w}$ of the observed next generation matrix. We compare the beginning of the pandemic – Panel (a) – and the NGM deriving from the reported social contact structure in April 2020 – Panel (b) – during the first lockdown. The latter corresponds to Wave 1 of the CoMix survey in Belgium, with the lowest damping ratio.

In the following we indicate $\mathbf{w_1}$ and $\mathbf{v_1}$ as $\mathbf{w}$ and $\mathbf{v}$ to simplify the notation. The dominant right eigenvector $\mathbf{w}$ expresses the *stable age structure of the infected population* or simply *stable age distribution* — the term stable to be intended as asymptotic incidence in the different age classes, whenever convergence is guaranteed. Eigenvectors are uniquely defined up to a constant and rescaling $\mathbf{w}$ such that $\sum w_i = 1$ enables interpreting its components w_i as the age-specific incidence of the disease. However, due to the arbitrariness of the rescaling, the most appropriate way to translate $\mathbf{w}$'s components is considering the relative quantities w_i/w_j as relative incidence, as done in Franco et al. (2022).

Moreover, from (C.3), we see how the datum on initially observed infections does not play a role in defining the infected age distribution ($\mathbf{w}$). In contrast, it does affect the size of the infected population. In fact, the constant c_1 in (C.3) is obtained as $c_1 = \langle \mathbf{v}, \mathbf{I}_{(0)} \rangle$, with $\langle \cdot, \cdot \rangle$ indicating the standard scalar product and $\mathbf{v}$ the left dominant eigenvector (proof by Keyfitz and Caswell, 2005). The size of the infected population $\|\mathbf{I}_{(t)}\|$, where $\|\cdot\|$ is the Manhattan norm, after a sufficient number (say t) of $\mathbf{K}$ iterations, can be expressed as $c_1 \lambda_1^t \langle \mathbf{v}, \mathbf{I}_{(0)} \rangle \|\mathbf{w}\|$. If we consider the perturbed initial condition $\mathbf{I}_{(0)} + \mathbf{e}_j$, where $\mathbf{e}_j$ is the jth vector of the canonical basis of $\mathbb{R}^n$, after m steps we will have

$$\|\mathbf{I}_{(t)}\| \approx \lambda_1^t (c_1 \|\mathbf{w}\| + v_j \|\mathbf{w}\|).$$

That is, adding a single infection in age class j (at the initial observation time) produces an increase in the asymptotic infected population equal to $\lambda_1^t v_j \|\mathbf{w}\|$. The above considerations support the interpretation of $\mathbf{v}$ as the vector of the age-specific *infective values*. Again we stress that $\mathbf{v}$ must be interpreted as relative quantity, as it can be scaled by any nonzero constant. In Section 3.3, we study the relative infective value to measure the relative impact on R_t of individuals belonging to two

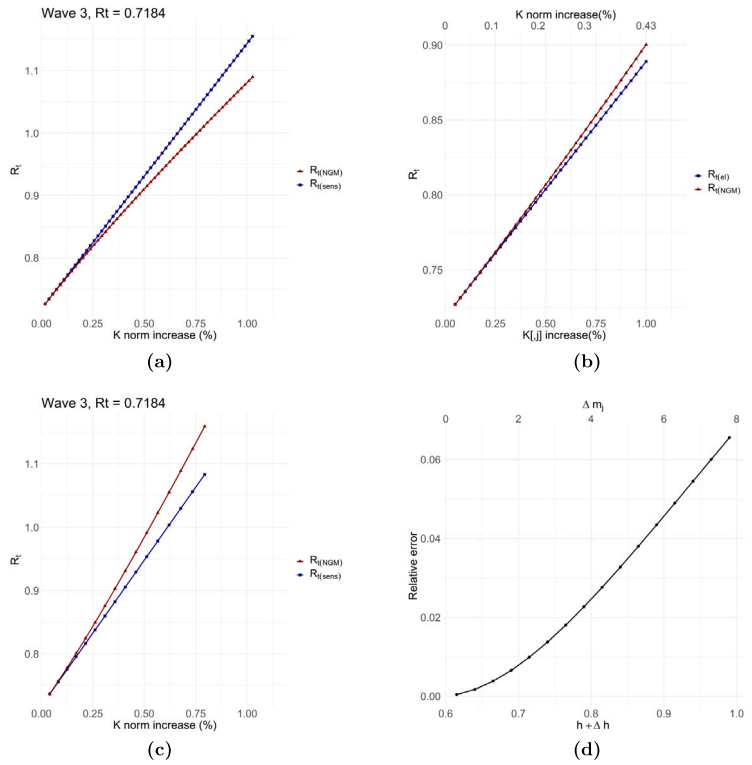


Fig. D.6. First order approximation of the R_t variation using sensitivity measures, illustrated for age group $[18, 30)$. Panel (a) shows the R_t evolution for a uniform increase in NGM column values: blue for approximated R_t and red for the NGM-derived value. Panel (b) presents the proportional increase of these column elements and the approximated value using elasticities. Panel (c) explores the impact of increased q-infectiousness and daily contacts within the age group $[18, 30)$, with the associated relative error displayed in panel (d).

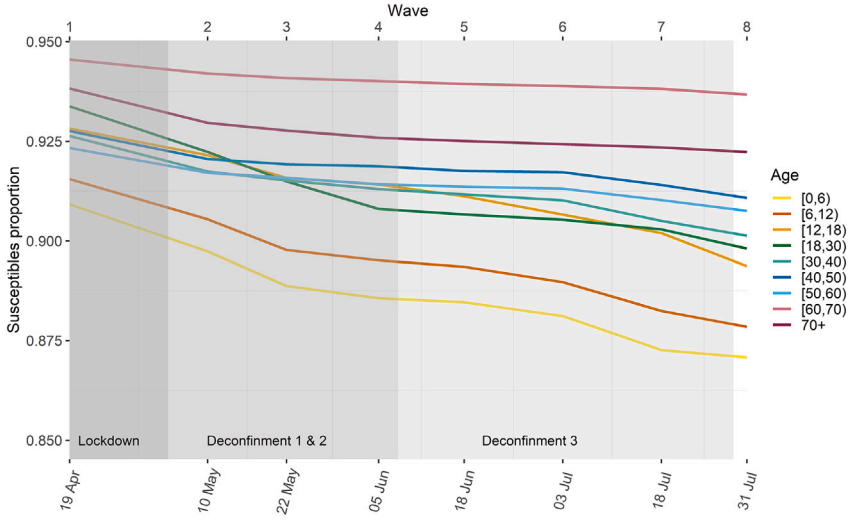


Fig. E.7. Age-specific mean proportion of susceptible individuals for the period April–July 2020, corresponding to the first 8 survey waves of CoMix (Coletti et al., 2021). The mean values are obtained using numerical estimates provided by the stochastic model developed by Abrams et al. (2021).

different age groups when we assume they are equally exposed to the risk of infection.

C.2. Convergence to the stable distribution **w**

We base our arguments on the ergodic properties of the solution to the problem (C.2). In analysing specific moments of the pandemic, one may ask after how many iterations of the next generation matrix the expected incidence of the disease in the various age groups will converge to the eigenvector **w** (rescaled so that $\sum_i w_i = 1$). That defines a time window over which our considerations are more robust. Provided the NGM (**K**) is diagonalisable and that the epidemiological pattern does not undergo significant perturbations, the answer to the above question is related to the accuracy of the limit in (C.3). Since by hypothesis, the set of **K** right eigenvector form a basis for the space of solutions $\mathbb{R}^n$, a popular measure of such accuracy is represented by:

$$d_r = \frac{\lambda_1}{|\lambda_2|}. \tag{C.4}$$

The ratio d_r has been defined as *damping ratio* (Caswell, 2000). The name refers to the speed at which the possible oscillations of the

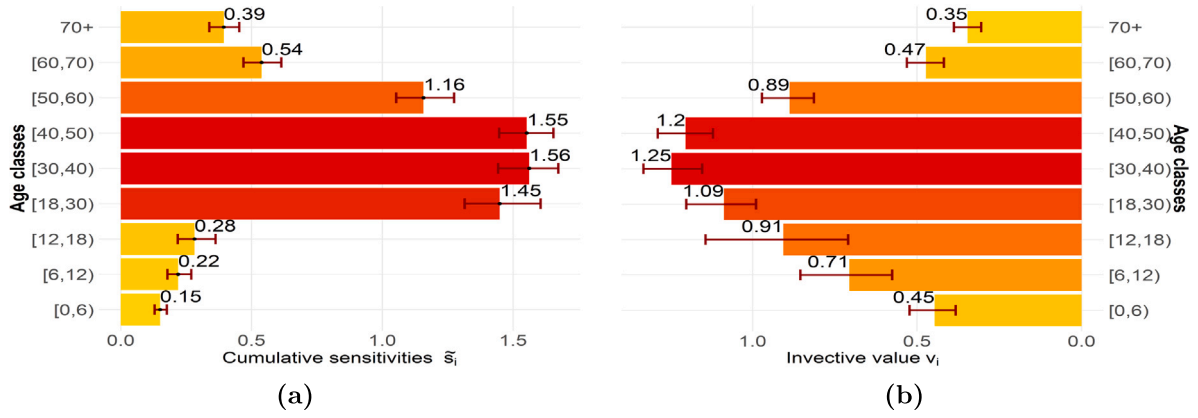


Fig. F.8. Panel(a): age-specific cumulative sensitivities $\tilde{s}_i$. Panel(b): infective value of a single infection in each age group. We show the 95% percentile interval obtained with 1000 bootstraps of the social contact data sample.

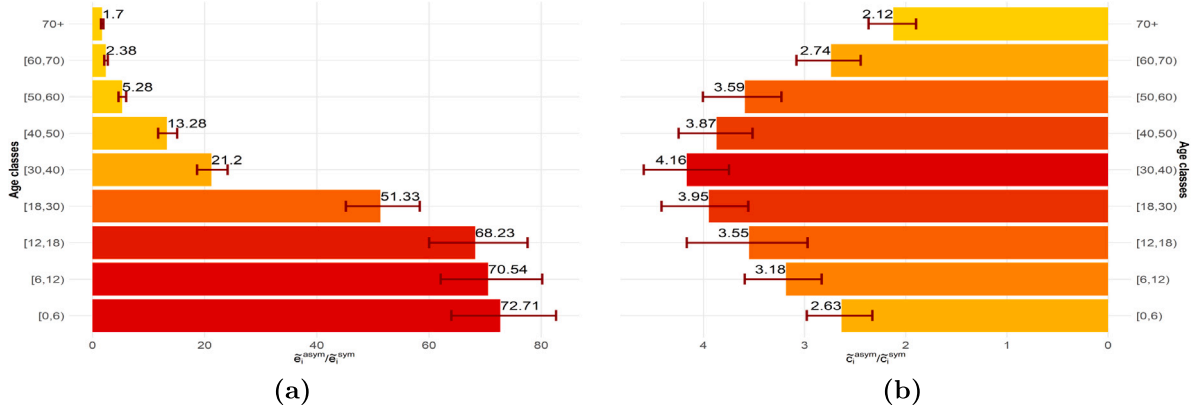


Fig. F.9. Panel(a): the age-specific ratio between elasticity of the basic reproduction number to the transmission term components, respectively corresponding to pre-symptomatic and asymptomatic and symptomatic interactions. Panel(b): the age-specific ratio between the expected number of daily contacts made by pre-symptomatic and asymptomatic individuals and those made by symptomatic ones (panel b). We show the 95% percentile interval obtained with 1000 bootstraps of the social contact data sample.

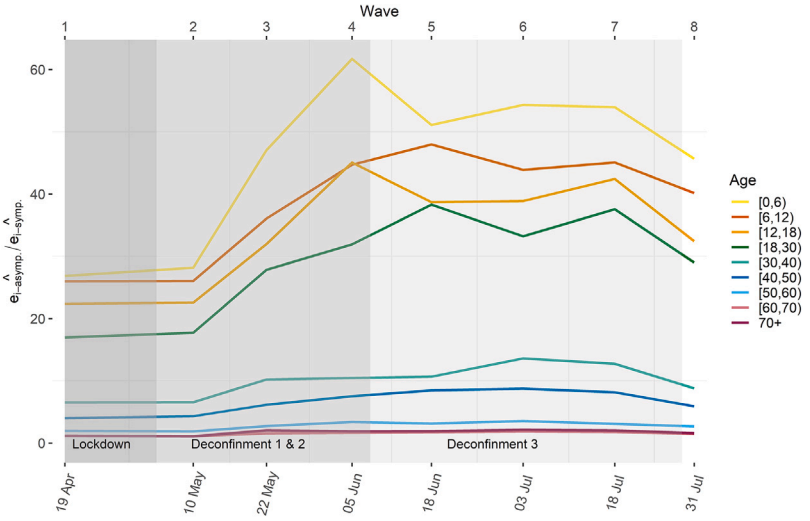


Fig. F.10. Age-specific ratio between elasticity of the effective reproduction number to transmission term components variation (e_l^* defined in Eq. (15), Section 2.6), corresponding to pre-symptomatic and asymptomatic and symptomatic interactions, respectively. Period April–July 2020.

solution are damped while converging to $\mathbf{w}$. The magnitude of such oscillations depends mainly on the eigenvalue with the second largest modulus in (λ_2). Neglecting the contribution of the full spectrum of $\mathbf{K}$, from (C.2) we have

$$\lim_{t\to\infty}\left(\frac{\mathbf{I}(t)}{\lambda_1^t} - c_1\mathbf{w} - c_2\mathbf{w_2}d_r^{-t}\right), \tag{C.5}$$

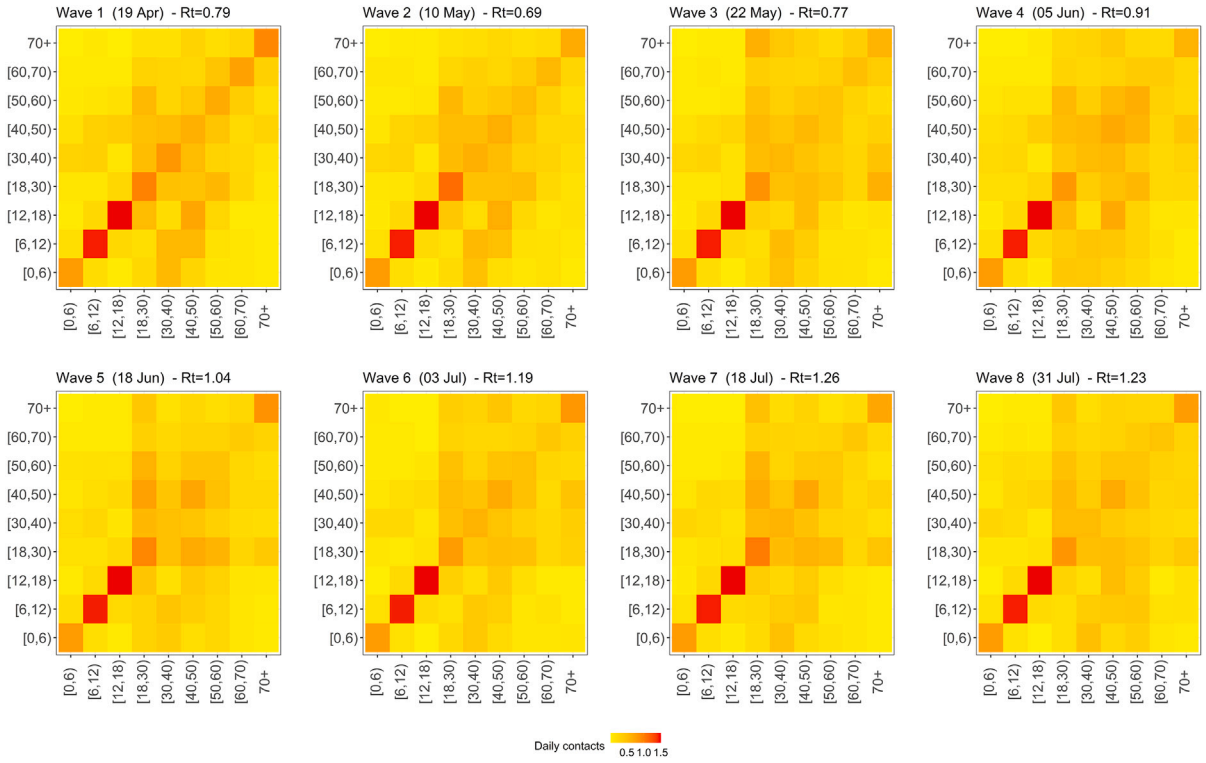


Fig. F.11. Social contact matrices: CoMix survey waves 1 to 8 (Belgium April–July 2020).

implying there exists a positive constant M such that

$$\| \frac{\mathbf{I}(t)}{\lambda_1^t} - c_1 \mathbf{w} \| \leq M d_r^{-t} = M e^{-t \, log(d_r)}. \tag{C.6}$$

This means that the demographic distribution of the infected population converges to $\mathbf{w}$ with an error exponentially decaying at least as fast as $e^{-t \ log(d_r)}$. After $\bar{t}$ iteration of $\mathbf{K}$, the contribution of λ_1 to the solution will be x time that of λ_2, namely

$$\left(\frac{\lambda_1}{|\lambda_2|} \right)^{\bar{t}} = x \quad \text{or equivalently} \quad \bar{t} = \frac{log x}{log d_r}. \tag{C.7}$$

A damping ratio of 2 means that it is sufficient just one iteration for the contribution to the asymptotic population of infected in the direction of $\mathbf{w}$ to double that in the direction of $\mathbf{w}_2$.

In Fig. B.5 we plotted the evolution of the proportion of infections $\frac{I_i^t}{\|\mathbf{I}_{(t)}\|}$ ($\| \cdot \|$ being the Manhattan norm) in the different age groups as obtained by iteratively applying the NGM to a hundred randomly generated initial conditions $\mathbf{I}_{(0)}$. We can appreciate how, within ten iterations, corresponding in our case to ten days, the magnitude of the damping ratio d_r affects the convergence speed to the stable age distribution $\mathbf{w}$. In Table C.5, we provide quantitative information about the impact of the damping ratio on convergence. We report the distance between the normalised vector $\tilde{\mathbf{I}_{(t)}} = \frac{\mathbf{I}_{(t)}}{\|\mathbf{I}_{(t)}\|}$ and $\mathbf{w}$ (also normalised such that $\sum_i w_i = 1$) for the first eight iterations including the initial condition, i.e. $t = 0, 1, \ldots, 8$. The employed distance is a scaled Manhattan distance

$$d(\tilde{\mathbf{I}_{(t)}}, \mathbf{w}) := Err_t = \frac{1}{2} \sum_{i=1}^{n} |\tilde{I_{(t)}^i} - w_i| \tag{C.8}$$

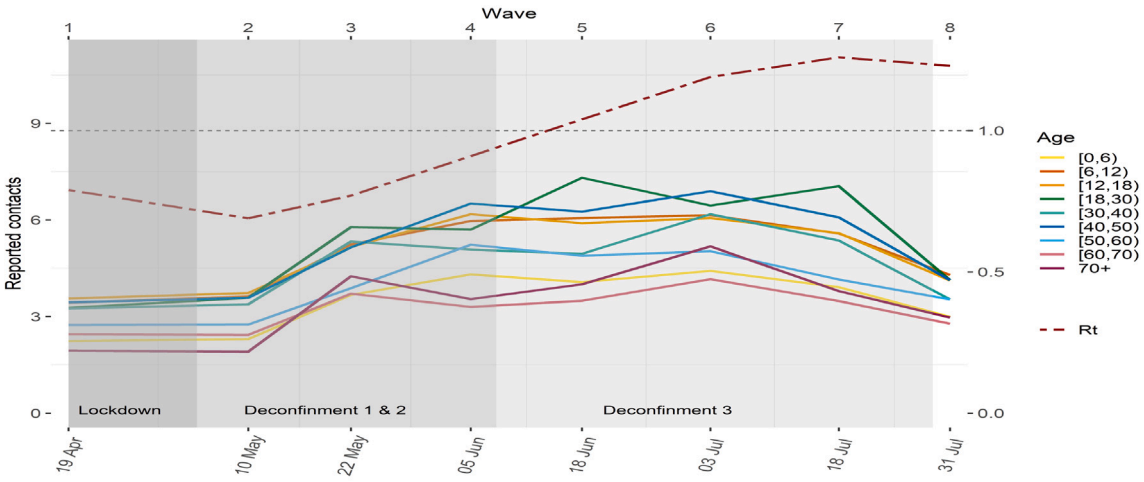
which assumes values in $[0, 1]$.

Considering the NGMs at different points in time, the convergence speed difference is neglectable with a decreasing d_r. After three iterations, the distance from the stable distribution is of the order of 10^{-3}. $\mathbf{w}$

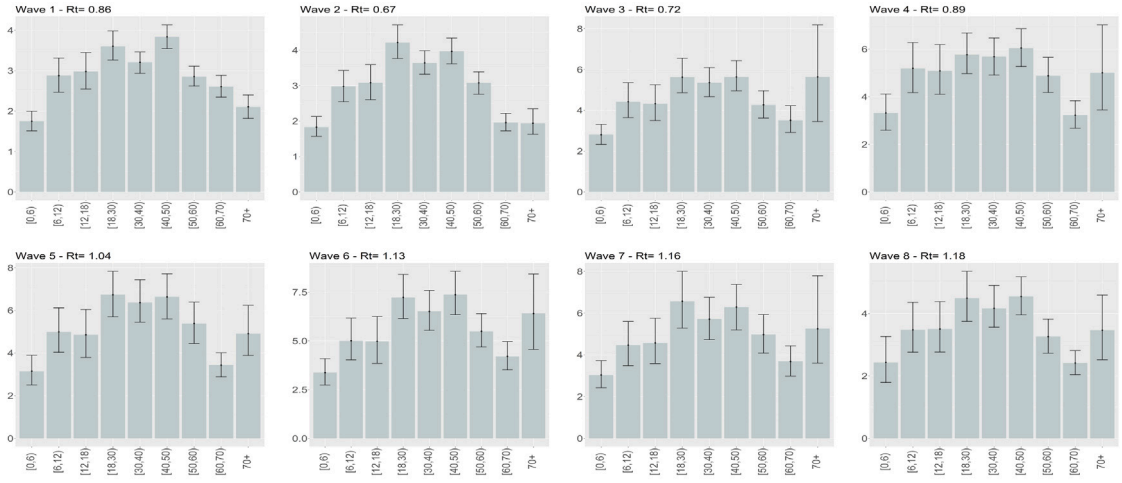
Appendix D. R_t first-order approximation

We analyse the performance of sensitivity measures in approximating R_t variations according to the formulas (12) and (13). The sensitivity measure is derived from the NGM corresponding to social contact data from wave 3 of the CoMix survey (5 May 2022).

Panel (a) focuses on the age group $[18, 30)$, which exhibited peak sensitivity $\tilde{s}_j$. We imposed a steady and uniform increase in the associated NGM column, reflecting a uniform growth in infections from this age group. The blue line indicates the approximated R_t; the red shows the R_t derived directly from the updated $\mathbf{K}$. Each k_{ij} element (with $j = [18, 30)$) recursively increases by $m\delta$ (with $\delta = 0.005$ and m the number of the iteration) until the perturbed $\mathbf{K}$'s Frobenius norm is twice its original value. For panel (b), we applied a consistent proportional increase of 2.5% to the same $\mathbf{K}$ column, doubling its values. The resultant increase in the Frobenius norm of the perturbed matrix is shown in the top x-axis. In panel (c), we combined increased q-infectiousness (e.g. simulating increasing transmissibility due to emerging virus variants) and daily contacts number (e.g. reflecting deconfinement strategies) for the $[18, 30)$ age group. Panel (d) compares the approximated and updated R_t, plotting the associated relative error against an increased q-infectiousness ($h_{[18,30)]}$) and social contact number($\sum_i m_{ij}$). The first-order approximation performs well, especially for norm increases below 50%. A 100% rise in $\mathbf{K}$'s column entries is associated with a 43% increase in its Frobenius norm. The relative R_t approximation error in panel (d) suggests errors around 7% for up to 8 additional daily contacts and q-infectiousness at 1 (meaning 100% chance to transmit the virus during a contact). Concerning observed variations in the NGM ($\mathbf{K}$) and the social contact matrix ($\mathbf{M}$), we noted the most significant changes during lockdown and deconfinement phases in Belgium. Specifically, the maximal mean absolute variation of k_{ij} was -0.0191 during early May 2020 (lockdown, wave 2) and 0.0189 in early June 2020 (deconfinement, wave 4). These changes correspond to a maximal proportional variation in $\mathbf{K}$'s Frobenius norm of 24% and 26%, respectively. During deconfinement at the end of May 2020



(a)



(b)

Fig. F.12. (a) Mean age-specific cumulative reported contacts during CoMix survey waves from 1 (19 April 2020) to 8 (31 July 2020); the dashed red line represents the R_t evolution as estimated from PCR positive tests by Sciensano (2021). In (b), we display the evolution of the age-specific average daily contacts with the 95% percentile interval bars obtained with 1000 bootstraps of the social contact data sample for each survey wave.

(wave 3), we recorded the most significant average contact variation in daily contacts, 1.6. The greatest absolute change in number of contacts occurred within the $[18, 30)$ age group, which reported an average of 2.9 fewer contacts per day by the end of July 2020 (wave 8). See also Fig. F.12. These findings suggest that, as long as epidemiological conditions remain relatively stable, our sensitivity ($\tilde{s}_j$) and elasticity ($\tilde{e}_j$) measures provide valuable insight into the age-specific impacts on transmission. The above result suggests that the impact of the underlying nonlinearity of the R_t as a function of its columns, given significant parameter perturbations, remains relatively modest. Consequently, as long as the epidemiological framework remains relatively stable without drastic shifts in the NGM structure, sensitivity ($\tilde{s}_j$) and elasticity ($\tilde{e}_j$) measures offer valuable insights into age-specific impacts on transmission at the observation time.

Appendix E. Age-specific susceptibility evolution.

Estimating the evolution of age-specific susceptibility, particularly for the age group $[0, 12)$, poses several challenges. In the present work, relying on the stochastic reference model developed by Abrams et al. (2021), we used the mean of 400 values obtained by 10 stochastic realisations based on 40 different parameters set estimated via Markov Chain Monte Carlo (MCMC) approach, Fig. E.7. The initial model calibration relied heavily on early serological data (April 2020, Li et al., 2020a), where the sample size for children under 12 was notably limited, contributing to high uncertainty, as highlighted in Abrams et al. (2021). Consequently, the model initially inferred a significant initial infection rate among young children below 10 years, many of whom might not have developed symptomatic disease. This resulted in a lower starting susceptible proportion of susceptible for group $[0, 12)$ as compared to other age groups (Fig. E.7). Subsequently, the model was calibrated mainly on hospitalisation and deaths by age group (Sciensano, 2021). Given the marginal contribution of children to hospitalisations and deaths, it is inherently more challenging to capture the nuances of age groups like the $[0, 12)$ years cohort. We observe a more marked decline in susceptible proportion for the $[0, 12)$ age group. This might seem to contradict other studies highlighting a lower seroprevalence of IgG antibodies in minors compared to adults (Herzog et al., 2022; Rajmil, 2020). However, seroprevalence does not necessarily equate directly to susceptibility in the longer term, especially

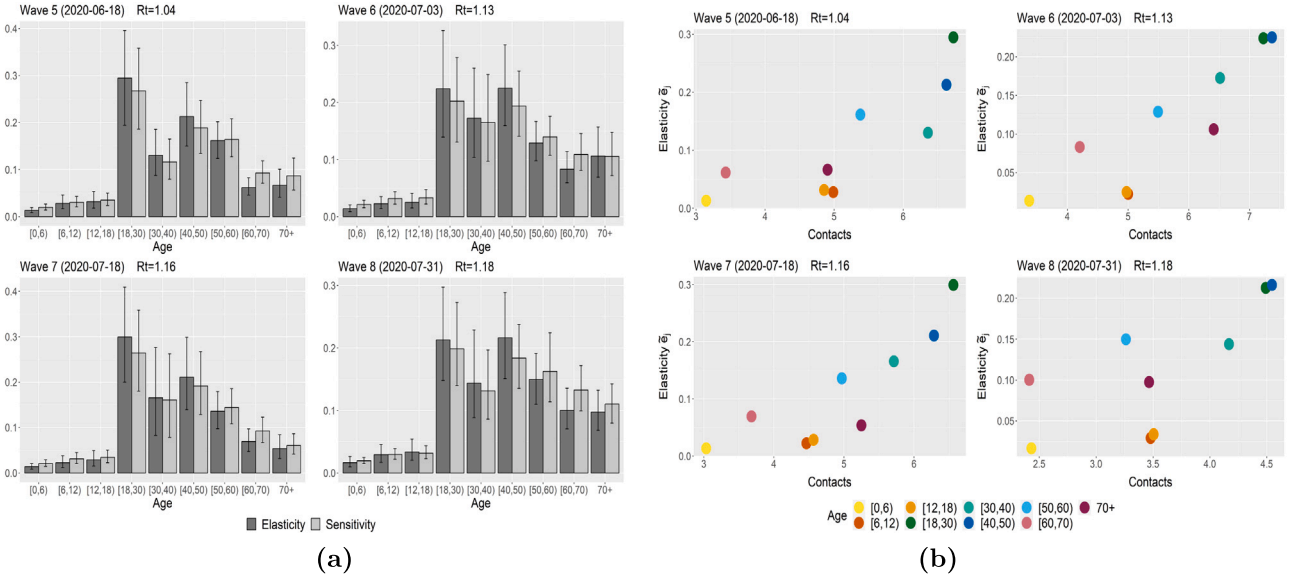


Fig. F.13. Age-specific sensitivity and elasticity evolution June–July 2020. Panel (a) shows the evolution of the cumulative indices $\tilde{s}_j$ $\tilde{e}_j$ for all $i, j = 1, \dots, n$ reported with the error bars as evaluated from 1000 bootstrap samples of the social contact matrices. Panel (b) consists of consecutive plots comparing cumulative contacts (x-axis) and cumulative elasticities (y-axis). The age groups in the top-right portion of the graphs contribute the most to transmission at the observation time. This corresponds to different CoMix survey waves (Coletti et al., 2020), for which we indicate the calendar time and the R_t obtained from Sciensano (2021).

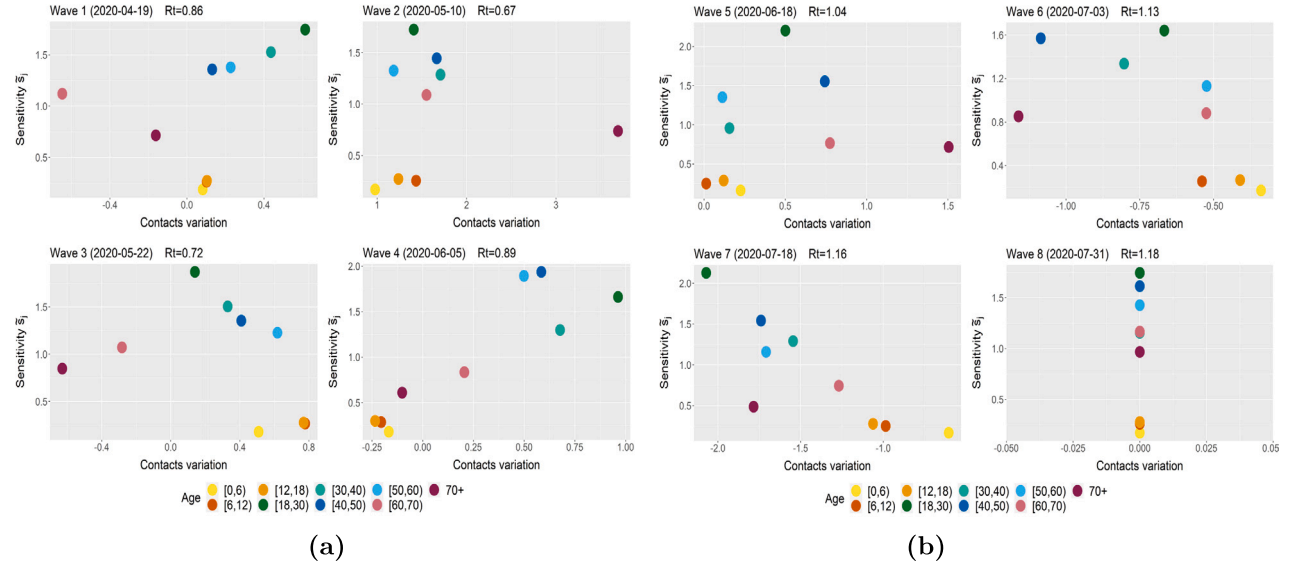


Fig. F.14. Age-specific sensitivity and contact variation, April–July 2020. For each survey wave, we plot the contact difference between the observed and the subsequent wave on the x-axis and the age-specific sensitivity on the y-axis. The presence of dots in the top-right region of the plots might identify undesirable contact evolutions, i.e. the most sensitive groups at time t are the ones that increase contacts the most at time $t + 1$.

when considering the nuances of humoural versus cellular immunity. The exact interplay of these immunities might differ in children, with possibilities like children having a more predominant cellular immune response. Future efforts to collect more robust, longitudinal serological data could significantly enhance our understanding and assessment of age-specific virus dynamics. Given these complexities, we acknowledge that our model's conclusions concerning this age group should be interpreted with caution.

Appendix F. Additional graphs

F.1. Prepandemic sensitivities & age-specific infective value

See Fig. F.8.

F.2. Age-specific relative impact on R_0 - pre-symptomatic & asymptomatic vs symptomatic transmission

See Figs. F.9 and F.10.

F.3. Reported contacts evolution: 19/04/2020–31/07/2020

See Figs. F.11 and F.12.

F.4. Further analysis: 18/06/2020–31/07/2020

See Figs. F.13 and F.14.

References

Abrams, S., Wambua, J., Santermans, E., Willem, L., Kuylen, E., Coletti, P., Libin, P., Faes, C., Petrof, O., Herzog, S., et al., 2021. Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. Epidemics 35, 100449. http://dx.doi.org/10.1016/j.epidem.2021.100449.

Begon, M., Bennett, M., Bowers, R.G., French, N.P., Hazel, S., Turner, J., 2002. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. Epidemiol. Infect. 129 (1), 147–153.

Caswell, H., 2000. Matrix Population Models, Vol. 1. Sinauer Sunderland, MA, USA.

Caswell, H., 2019. Sensitivity Analysis: Matrix Methods in Demography and Ecology. Springer Nature.

Chin, T., Feehan, D.M., Buckee, C.O., Mahmud, A.S., 2021. Contact surveys reveal heterogeneities in age-group contributions to SARS-CoV-2 dynamics in the United States.

Cohen, J.E., et al., 1979. Ergodic theorems in demography. Bull. Amer. Math. Soc. (N.S.) 1 (2), 275–295.

Coletti, P., Libin, P., Petrof, O., Willem, L., Abrams, S., Herzog, S.A., Faes, C., Kuylen, E., Wambua, J., Beutels, P., et al., 2021. A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies. BMC Infect. Dis. 21 (1), 1–12.

Coletti, P., Wambua, J., Gimma, A., Willem, L., Vercruysse, S., Vanhoutte, B., Jarvis, C.I., Van Zandvoort, K., Edmunds, J., Beutels, P., et al., 2020. Comix: comparing mixing patterns in the Belgian population during and after lockdown. Sci. Rep. 10 (1), 1–10.

Davies, N.G., Klepac, P., Liu, Y., Prem, K., Jit, M., Eggo, R.M., 2020. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat. Med. 26 (8), 1205–1211.

Diekmann, O., Heesterbeek, H., Britton, T., 2012. Mathematical Tools for Understanding Infectious Disease Dynamics, Vol. 7. Princeton University Press.

Diekmann, O., Heesterbeek, J.A.P., Metz, J.A., 1990. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28 (4), 365–382.

Diekmann, O., Heesterbeek, J., Roberts, M.G., 2010. The construction of next generation matrices for compartmental epidemic models. J. R. Soc. Interface 7 (47), 873–885.

Du, Q., Zhang, D., Hu, W., Li, X., Xia, Q., Wen, T., Jia, H., 2021. Nosocomial infection of COVID-19: A new challenge for healthcare professionals. Int. J. Mol. Med. 47 (4), 1.

Franco, N., Coletti, P., Willem, L., Angeli, L., Lajot, A., Abrams, S., Beutels, P., Faes, C., Hens, N., 2022. Inferring age-specific differences in susceptibility to and infectiousness upon SARS-CoV-2 infection based on belgian social contact data. PLoS Comput. Biol. 18 (3), e1009965.

Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., Hens, N., 2020. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Eurosurveillance 25 (17), 2000257.

Government(BE), 2020. Coronavirus: Belgium outlines its exit strategy. https://www.info-coronavirus.be/en/news/archive/nsc-24-04/.

Heesterbeek, J.A.P., Metz, J.A., 1993. The saturating contact rate in marriage-and epidemic models. J. Math. Biol. 31, 529–539.

Held, L., Hens, N., D O'Neill, P., Wallinga, J., 2019. Handbook of Infectious Disease Data Analysis. CRC Press.

Herzog, S.A., De Bie, J., Abrams, S., Wouters, I., Ekinci, E., Patteet, L., Coppens, A., De Spiegeleer, S., Beutels, P., Van Damme, P., et al., 2022. Seroprevalence of IgG antibodies against SARS-CoV-2–a serial prospective cross-sectional nationwide study of residual samples, Belgium, March to October 2020. Eurosurveillance 27 (9), 2100419.

Hoang, T.V., Coletti, P., Kifle, Y.W., Kerckhove, K.V., Vercruysse, S., Willem, L., Beutels, P., Hens, N., 2021. Close contact infection dynamics over time: insights from a second large-scale social contact survey in Flanders, Belgium, in 2010–2011. BMC Infect. Dis. 21 (1), 1–15.

Keyfitz, N., Caswell, H., 2005. Applied Mathematical Demography, Vol. 47. Springer.

Lau, M.S., Grenfell, B., Thomas, M., Bryan, M., Nelson, K., Lopman, B., 2020. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. Proc. Natl. Acad. Sci. 117 (36), 22430–22435.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y., et al., 2020a. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N. Engl. J. Med. 382 (13), 1199–1207. http://dx.doi.org/10.1056/NEJMoa2001316.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020b. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 368 (6490), 489–493.

Liu, Y., Gu, Z., Xia, S., Shi, B., Zhou, X.-N., Shi, Y., Liu, J., 2020. What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization. EClinicalMedicine 22, 100354.

Lovell-Read, F.A., Shen, S., Thompson, R.N., 2022. Estimating local outbreak risks and the effects of non-pharmaceutical interventions in age-structured populations: SARS-CoV-2 as a case study. J. Theoret. Biol. 535, 110983.

McEvoy, D., McAloon, C., Collins, A., Hunt, K., Butler, F., Byrne, A., Casey-Bryars, M., Barber, A., Griffin, J., Lane, E.A., et al., 2021. Relative infectiousness of asymptomatic SARS-CoV-2 infected persons compared with symptomatic individuals: a rapid scoping review. BMJ Open 11 (5), e042354.

Metz, J.A., Diekmann, O., 1986. A gentle introduction to structured population models: three worked examples. In: The Dynamics of Physiologically Structured Populations. Springer, pp. 3–45.

Metz, J.A., Diekmann, O., 2014. The Dynamics of Physiologically Structured Populations, Vol. 68. Springer.

Monod, M., Blenkinsop, A., Xi, X., Hebert, D., Bershan, S., Tietze, S., Baguelin, M., Bradley, V.C., Chen, Y., Coupland, H., et al., 2021. Age groups that sustain resurging COVID-19 epidemics in the United States. Science 371 (6536), eabe8372.

Nabi, K.N., 2020. Forecasting COVID-19 pandemic: A data-driven analysis. Chaos Solitons Fractals 139, 110046.

Rajmil, L., 2020. Role of children in the transmission of the COVID-19 pandemic: a rapid scoping review. BMJ Paediatr. Open 4 (1).

Sciensano, 2021. The belgian public health institute. https://epistat.wiv-isp.be/covid/.

StatBel, 2020. The belgian statistical office. https://statbel.fgov.be/en.

Van Hoang, T., Coletti, P., Kifle, Y.W., Van Kerckhove, K., Vercruysse, S., Willem, L., Beutels, P., Hens, N., 2021. Close contact infection dynamics over time: insights from a second large-scale social contact survey in Flanders, Belgium, in 2010–2011. BMC Infect. Dis. 21 (1), 1–15.

Van Kerckhove, K., Hens, N., Edmunds, W.J., Eames, K.T.D., 2013. The impact of illness on social networks: Implications for transmission and control of influenza. Am. J. Epidemiol. 178 (11), 1655–1662. http://dx.doi.org/10.1093/aje/kwt196.

Verelst, F., Hermans, L., Vercruysse, S., Gimma, A., Coletti, P., Backer, J.A., Wong, K.L., Wambua, J., van Zandvoort, K., Willem, L., et al., 2021. SOCRATES-CoMix: a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries. BMC Med. 19 (1), 1–7.

Viner, R.M., Russell, S.J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., Booy, R., 2020. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. Lancet Child Adolesc. Health 4 (5), 397–404.

Wallinga, J., van Boven, M., Lipsitch, M., 2010. Optimizing infectious disease interventions during an emerging epidemic. Proc. Natl. Acad. Sci. 107 (2), 923–928.

Wallinga, J., Teunis, P., Kretzschmar, M., 2006. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am. J. Epidemiol. 164 (10), 936–944.

WHO, 2020. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. https://covid19.who.int/?gclid=EAIaIQobChMIzPH5wtWe6gIVhagYCh2yvgfsEAAYASAAEgL-G_D_BwE.

Willem, L., Abrams, S., Libin, P.J., Coletti, P., Kuylen, E., Petrof, O., Møgelmose, S., Wambua, J., Herzog, S.A., Faes, C., et al., 2021. The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19. Nat. Commun. 12 (1), 1–9.

Willem, L., Van Hoang, T., Funk, S., Coletti, P., Beutels, P., Hens, N., 2020. SOCRATES: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for COVID-19. BMC Res. Notes 13 (1), 1–8.

Zhu, Y., Bloxham, C.J., Hulme, K.D., Sinclair, J.E., Tong, Z.W.M., Steele, L.E., Noye, E.C., Lu, J., Xia, Y., Chew, K.Y., et al., 2021. A meta-analysis on the role of children in severe acute respiratory syndrome coronavirus 2 in household transmission clusters. Clin. Infect. Dis. 72 (12), e1146–e1153.