Inferring age-specific differences in susceptibility to and infectiousness upon SARS-CoV-2 infection based on Belgian social contact data

Nicolas Franco [1,2*], Pietro Coletti[1], Lander Willem[3], Leonardo Angeli[1], Adrien Lajot[4], Steven Abrams[1,5], Philippe Beutels[3], Christel Faes[1], Niel Hens[1,3]

1 *Data Science Institute, Interuniversity Institute of Biostatistics and statistical Bioinformatics (I-BioStat), UHasselt, Hasselt, Belgium*

2 *Namur Institute for Complex Systems (naXys) and Department of Mathematics, University of Namur, Namur, Belgium*

3 *Centre for Health Economic Research and Modelling Infectious Diseases (CHERMID), Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium*

4 *Department of Epidemiology and public health, Sciensano, Brussels, Belgium*

5 *Global Health Institute (GHI), Family Medicine and Population Health, University of Antwerp, Antwerp, Belgium*

* Corresponding author: nicolas.franco@uhasselt.be

Abstract

Several important aspects related to SARS-CoV-2 transmission are not well known due to a lack of appropriate data. However, mathematical and computational tools can be used to extract part of this information from the available data, like some hidden age-related characteristics. In this paper, we investigate age-specific differences in susceptibility to and infectiousness upon contracting SARS-CoV-2 infection. More specifically, we use panel-based social contact data from diary-based surveys conducted in Belgium combined with the next generation principle to infer the relative incidence and we compare this to real-life incidence data. Comparing these two allows for the estimation of age-specific transmission parameters. Our analysis implies the susceptibility in children to be around half of the susceptibility in adults, and even lower for very young children (preschooler). However, the probability of adults and the elderly to contract the infection is decreasing throughout the vaccination campaign, thereby modifying the picture over time.

Introduction

Since the start of the COVID-19 pandemic, a new respiratory disease caused by the SARS-CoV-2 coronavirus, many mathematical and statistical approaches have been considered to identify transmission dynamics and characteristics of the virus. Some of those characteristics are still not completely known due to the lack of appropriate data. However, these characteristics are necessary in order to correctly inform public health policies as well as to develop more advanced scientific tools like mathematical and computational models. Concerning COVID-19, as for most infectious diseases, it quickly became apparent that some of the disease characteristics are strongly age-dependent [1]. In particular, the susceptibility to SARS-CoV-2 infection as well as the infectiousness upon infection may be lower for children than for adults and the elderly. Knowledge of such a difference could have an important impact on public health strategies in terms of prioritization of vaccination or the choice of targeted non-pharmaceutical interventions.

We propose a method to estimate those heterogeneous differences concerning SARS-CoV-2 transmission, focusing on relative susceptibility and infectiousness. Our approach is based on the method used by [2] to estimate susceptibility profiles for influenza A/H1N1. However, we have refined it to include a larger number of age categories and applied the methodology to SARS-CoV-2 transmission using a numerical approach. The method relies on the analysis of social contact data in order to derive an estimate of the relative incidence which is compared to real incidence data.

Social contact surveys [3] coupled with the next generation principle [4, 5] have been used for years to estimate key epidemiological parameters such as the basic (and effective) reproduction number (i.e., the average number of new infections caused by a *typical* infected individual during their entire infectious period in a (fully) susceptible population), relative incidence or differences in susceptibility [2]. The first large-scale social contact study, POLYMOD [6], collected social contact patterns for eight European countries between May 2005 and September 2006. In 2020-2021, social contact data has been collected in the so-called CoMix survey [7–10], initially in the United Kingdom, The Netherlands and Belgium and afterwards extended to other European countries. Comix collected timely social contact information during the COVID-19 pandemic.

Social contact data can be used as a proxy to model SARS-CoV-2 transmission using the so-called *social contact hypothesis* [11], which implies that the age-specific number of infectious contacts is proportional to the self-reported age-specific number of social contacts by a proportionality factor. This proportionality factor, often denoted by q, assumes that the probability of transmission is homogeneous across the different age classes. In the current paper, we aim to disentangle and quantify the heterogeneous components of this proportionality factor further [12], elucidating information on relative age-specific susceptibility and infectiousness. These estimates could serve to inform heterogeneous COVID-19 mathematical models relying on social contact data, such as e.g. mechanistic models [13–16]. Social contact data are also used in [17, 18] to derive heterogeneous contributions to SARS-CoV-2 transmission using an approach based on the reproduction number. We go one step further using an approach based on the relative incidence derived from the next generation principle.

More specifically, we use the CoMix social contact data combined with daily incidence data on the number of new confirmed COVID-19 cases in Belgium over the period December 2020 to May 2021 to estimate the proportionality factor and its heterogeneous unmeasured components. We disentangle potential sources of heterogeneity in the acquisition of SARS-CoV-2 infection especially focusing on the comparison between children (infant, primary and secondary school) and adults. We also estimate the time evolution of the transmission parameters for different adult age classes throughout the vaccination campaign as carried out in Belgium showing an evolution of the proportionality factors over time. Then we present an illustration of the utility of heterogeneous proportionality factors by comparing the reproduction number estimated from the CoMix social contact data to the ones estimated from incidence of cases and hospitalisations, respectively.

Results

Estimation of susceptibility and infectiousness through proportionality factors

The proportionality factor is assumed to be age-specific and denoted by q_{ij} where i and j belong to some age classes. q_{ij} could be further split into heterogeneous components:

$$q_{ij} = \tilde{q}\, a_i\, h_j,$$

where:

- The vector (a_i) represents age-specific differences in factors influencing transmission which are specifically related to the susceptibility of individuals, including, but not limited to, direct (immunological) susceptibility to infection upon exposure (e.g., due to age-specific heterogeneous risk behaviour and/or compliance to non-pharmaceutical interventions not already captured by contact frequency). In order to distinguish it from direct susceptibility to infection, this vector will be referred to as *q-susceptibility*.
- The vector (h_j) describes age-specific differences in factors influencing transmission which are specifically related to the infectiousness of individuals, including, but not limited to, infectiousness after acquiring infection (e.g., due to differences in viral load upon exposure). In order to distinguish it from infectiousness upon infection, this vector will be referred to as *q-infectiousness*.
- The remaining global proportionality factor $\tilde{q}$ captures any remaining residual effect and is of no relevance when considering relative q-susceptibility or q-infectivity.

In the remainder of this paper, we will talk about susceptibility or infectiousness when considering immunological aspects of disease transmission, while we will add the prefix q whenever quantities can carry additional effects related to susceptible and infectious individuals in order to avoid any ambiguity.

If we denote by $w = (w_j)$ a vector representing the relative incidence within age class j and by $\mathbf{M}^T$ a matrix containing the social contact data, then we have the following system:

$$\tilde{q}\ \mathrm{diag}(a_i)\, \mathbf{M}^T\, \mathrm{diag}(h_j)\, w = R_t\, w,$$

where R_t represents the reproduction number. The core matrix of this system, $\mathbf{K} = \tilde{q}\ \mathrm{diag}(a_i)\, \mathbf{M}^T\, \mathrm{diag}(h_j)$ is called the *next generation matrix* and gives the number of new infections in a successive generation. Details concerning the construction of this matrix and the next generation principle can be found in Section Materials and methods.

Using our method, we compare the social contact matrix $\mathbf{M}^T$ extracted from the CoMix social contact survey in Belgium [7] and its derived relative incidence w to the incidence obtained from real-life data in Belgium coming from PCR positive tests [19]. This method allows for an estimation of either the relative q-susceptibility (a_i) or relative q-infectiousness (h_j) by age class, while assuming that the other set of parameters is known from the literature (i.e., holding one of the two vectors fixed). The chosen age groups are $[0, 6)$ years, $[6, 12)$ years, $[12, 18)$ years, $[18, 30)$ years and subsequent 10-year age classes up to 80+ years in order to account for the Belgian educational system. The period of observation goes from 22 December 2020 to 26 May 2021 (and to 15 June 2021 for confirmed cases data). A detailed description of considered data, literature assumptions and fitting procedure is presented in Section Materials and methods.

The estimated relative q-susceptibility for the whole period is presented in Fig 1 and implies that very young children in age group $[0, 6)$ are about 0.182 (95% percentile bootstrap-based CI: 0.146-0.230) times as susceptible compared to the first adult age class $[18, 30)$ with relative susceptibility equal to 1 (95% CI: 0.829-1.252). Primary school students aged $[6, 12)$ have a relative susceptibility of 0.550 (95% CI: 0.427-0.629) and secondary school students aged $[12, 18)$ a susceptibility of 0.603 (95% CI: 0.536-0.700). This shows an increasing q-susceptibility by increasing age up to $[18, 30)$ after which the relative q-susceptibility tends to decrease slightly. Note, however, that this q-susceptibility captures not only differences in clinical susceptibility to infection and the rather low relative q-susceptibility in the $[12, 18)$ age class could therefore be influenced by (compliance to) non-pharmaceutical interventions, over and above the age-specific contact frequencies. Due to the method, these numbers only have a relative interpretation and the proposed representation in Fig 1 is under the assumption of a mean susceptibility of 1 for the first adult class $[18, 30)$.

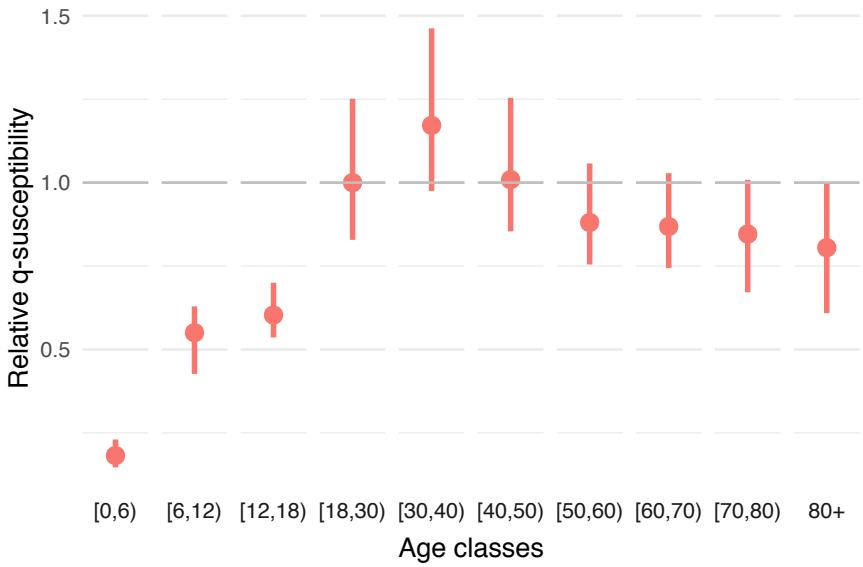

Fig 1. Estimated relative q-susceptibility. The estimation of q-susceptibility is performed by age class using the next generation principle under an assumption on age-varying infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$. The calibration is performed on CoMix waves 12 to 23 (observation period: 22 December 2020 to 15 June 2021). Dots represent means and bars represent 95% percentile (nonparametric) bootstrap-based confidence intervals.

The comparison of the relative incidence as estimated based on the positive PCR test data and the CoMix social contact data is presented in Fig 2. The social contact data are presented by waves starting with wave 12 on 22 December 2020 and an inter-survey wave interval of two weeks for subsequent waves (cf. details in Table A of S1 Appendix). The nationally collected data are represented in blue and the estimates coming from social contact data in two colors: in green, the initial estimate with a homogeneous proportionality factor (i.e., with $a_i = 1$ and $h_j = 1$ for all i, j) and in red, the estimate using heterogeneous q-susceptibility and infectiousness as presented in Fig 1. We clearly observe that estimates of the relative incidence under the homogeneous proportionality factor assumption (green) are very different from the empirical estimates (blue), especially for the young age groups. The relative incidence among adult age classes is estimated relatively well up to a constant, but the relative incidence for children coming from the homogeneous social contacts approach is clearly overestimated, except perhaps during times of school closure (see, e.g., wave 19). This finding provides a clear indication that SARS-CoV-2 transmission is less important in children as compared to adults.

The result of estimating the q-infectiousness for the whole period, is depicted in Fig 3. The estimates also show a potential important heterogeneity concerning the proportionality factor on the infectiousness side. However, this reverse exercise provides less accurate results, with very large confidence intervals and some bootstrap estimates reaching zero, both being problematic when dealing with relative values. Those effects are the result of a lack of constraints for q-infectiousness. Indeed, while it is impossible to reach zero susceptibility for a specific age class when having at the same time non-zero incidence, it is technically allowed for age-specific infectiousness to be zero as the observed incidence could result from transmission from other age classes.

Exact estimates of the components of the q-susceptibility (a_i) and q-infectiousness (h_i) are provided in Tables D and H of S1 Appendix. Additional estimates under the assumption of homogeneity regarding infectiousness or susceptibility (i.e., estimating (a_i) under $h_j = 1$, $\forall j$ or estimating (h_j) under $a_i = 1$, $\forall i$) are also presented in Figs B and J of S1 Appendix together with estimated values and the effect on the relative incidence. These additional estimates provide qualitatively similar results.

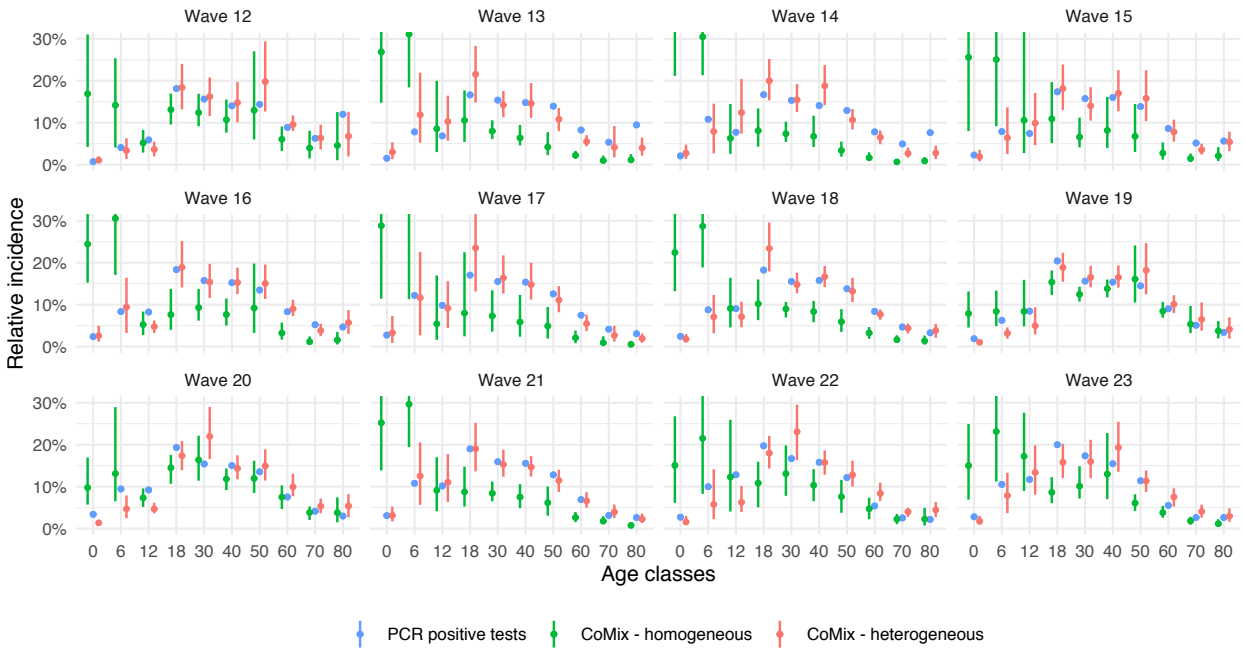

Fig 2. Estimated relative incidence. In blue: relative incidence based on Belgian PCR data. In green: estimated relative incidence based on the next generation principle under the assumption of a homogeneous proportionality factor. In red: estimated relative incidence based on the next generation principle with an estimated age-varying q-susceptibility under the assumption on age-varying infectiousness as in Fig 1. Dots represent means and bars represent 95% percentile (nonparametric) bootstrap-based confidence intervals.

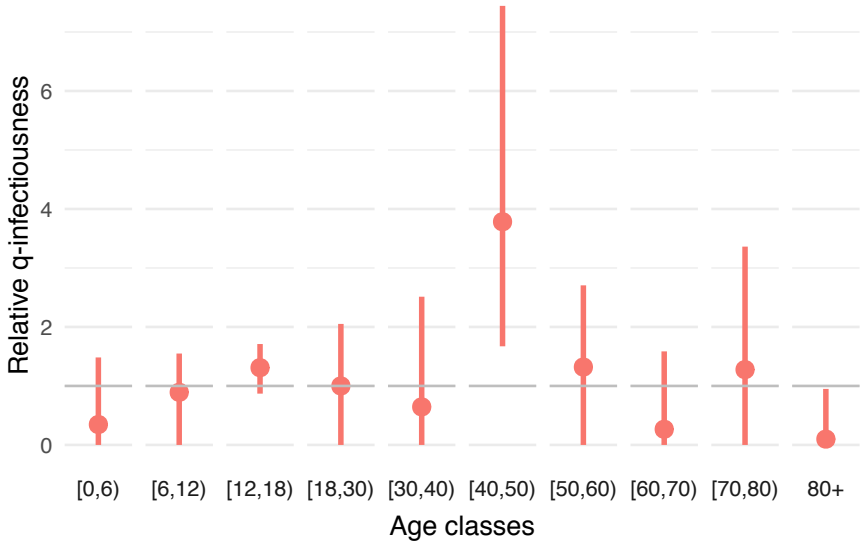

Fig 3. Estimated relative q-infectiousness. The estimation of q-infectiousness is performed by age class using the next generation principle under an assumption on age-varying susceptibility $(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$. The calibration is performed on CoMix waves 12 to 23. Dots represent means and bars represent 95% percentile (nonparametric) bootstrap-based confidence intervals.

Time evolution of proportionality factors

Since proportionality factors capture several effects, they also capture time-dependent effects such as the reduction in susceptibility and infectiousness as a result of the vaccination campaign. In order to account for such a time evolution, we also performed the previous analysis using groups of two consecutive CoMix waves instead of the full period. The decision to consider two CoMix waves (28 days) together is motivated by the fact that a sufficiently long non-holidays period is required as social contacts in children are of importance and the heterogeneity of the transmission concerning adult classes is partially constraint by infection reported by children. Note that the gradual introduction of the alpha variant of concern might also interfere.

Estimates of the time-dependent q-susceptibility relying on the same (time-invariant) assumption with regard to the infectiousness vector are presented in Fig 4. A normalisation of the relative values was performed over the different waves such that the average of the estimated factors for the age classes $[0, 6)$, $[6, 12)$ and $[12, 18)$ is assumed constant. This choice is motivated by the fact that the vaccination campaign was not including children during the entire study period, hence proportionality factors regarding susceptibility can be expected to be stable for these age classes. Thus, the results provide an estimate of the evolution of adults' proportionality factors under an on average constant assumption for children $[0, 18)$.

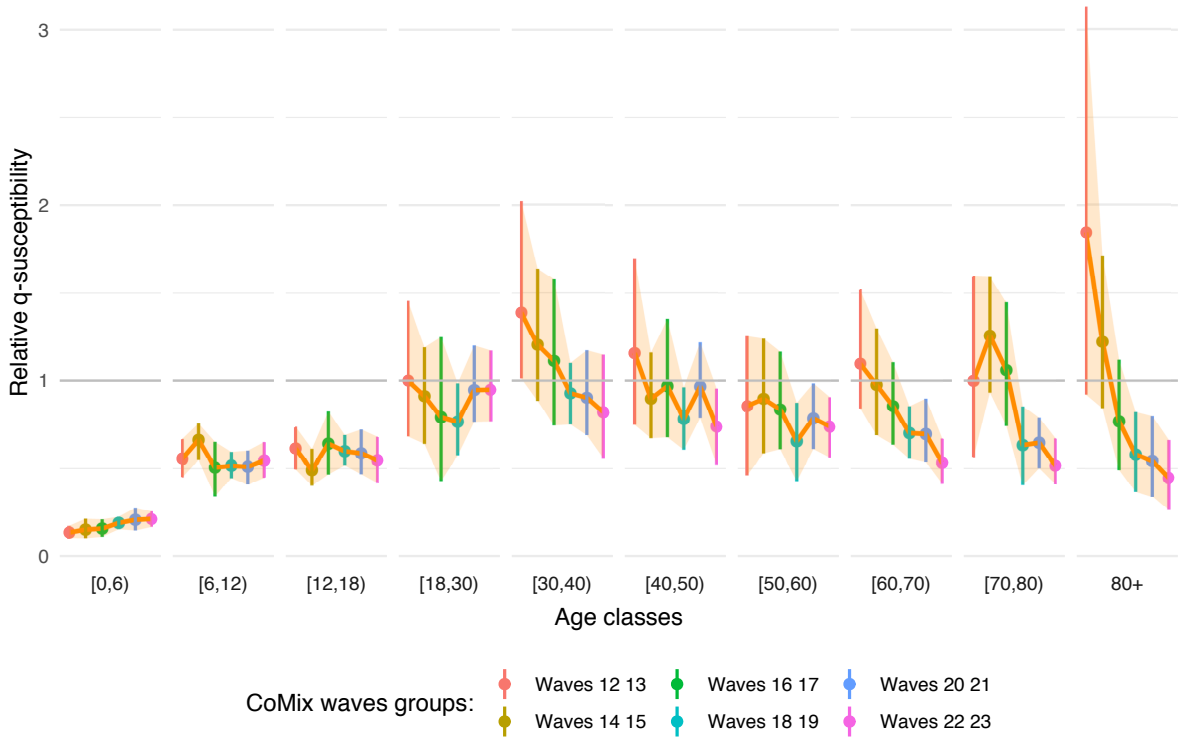

Fig 4. Time-varying estimated relative q-susceptibility. The estimation of q-susceptibility is performed by age class and over time on groups of two consecutive CoMix waves, corresponding to a period of 4 weeks, under an assumption on age-varying infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$. Dots represent means and bars represent 95% percentile (nonparametric) bootstrap-based confidence intervals.

As indicated previously concerning the estimation during the complete time-period, a decreasing q-susceptibility is observed through adult age classes (see Fig 1), with the oldest age class (80+) being the least susceptible among all adults aged 18 years or older. This is a priori in contrast with usual assumptions regarding age-specific susceptibility to SARS-CoV-2 infection. However, in Fig 4, we clearly observe the highest relative q-susceptibility in the 80+ age class for the earliest waves as compared to all other age groups,

or at least an equal q-susceptibility by considering the lower side of the confidence interval. Moreover, the q-susceptibility in the oldest age class decreases rapidly over time towards the lowest relative q-susceptibility equal to 0.446 (95% CI: 0.266-0.660) among the adult age groups. In general, the estimated q-susceptibility is almost similar across the different adult classes during the first period with the exception of the oldest class 80+ with an estimated relative q-susceptibility of 1.844 (95% CI: 0.920-3.127). Overall, q-susceptibility estimates of other age classes tend to decrease over time, albeit at a slower pace and to a lesser extent. This is in line with the implementation of the vaccination policy in Belgium, giving vaccination priority to residents of nursing homes (CoMix waves 13-16, see schematic timeline in Fig A of S1 Appendix) and the elderly in the general population (CoMix waves 18-21), while going gradually down from old to young throughout the study period.

Exact values of the estimates in Fig 4 are provided in Table E of S1 Appendix as well as time-dependent q-susceptibility and q-infectiousness under the various constraints mentioned above.

Time evolution of R_t using contact patterns

In order to check the utility and validity of the use of a heterogeneous proportionality factor, we illustrate its application by determining the reproduction number R_t, or more specifically the variation of the reproduction number over time, and comparing this evolution with R_t directly estimated from confirmed cases/hospitalizations data.

In Fig 5, the variation of the reproduction number computed from the CoMix data is compared to the reproduction number computed either from the number of cases [20] (panel a) or from hospitalizations [19] (panel b). Clearly, specific choices of $q-$susceptibility and $q-$infectiousness affect the computation and in Fig 5 we report results for the homogeneous and heterogeneous scenarios of Fig 2. A homogeneous assumption for $q-$infectiousness and $q-$susceptibility leads to a poor agreement with the reproduction number estimated from both confirmed cases and hospitalizations and is also characterized by a larger uncertainty while the use of the estimated heterogeneous reproduction factor agrees more with reality.

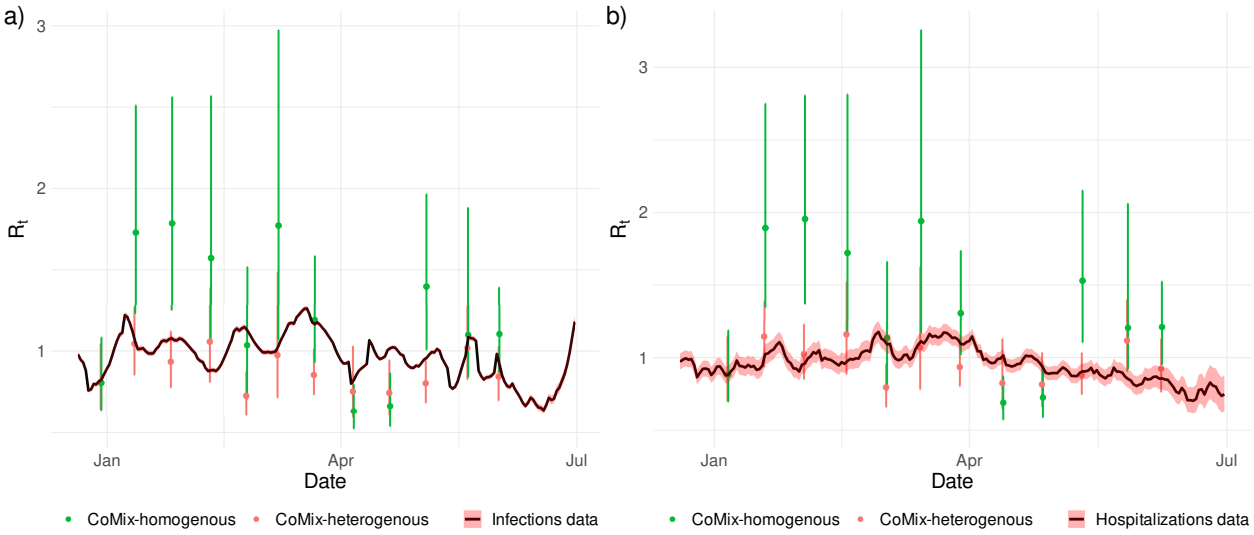

Fig 5. Time varying reproduction number. Reproduction number estimated from the CoMix data using the next generation approach in comparison to the reproduction number estimated from the number of confirmed cases (a) and from the number of hospitalizations (b). In green: estimated R_t under the assumption of a homogeneous proportionality factor. In red: estimated R_t with the estimated age-varying q-susceptibility under the assumption on age-varying infectiousness as in Figs 1 and 2. Dots represent means and bars represent 95% percentile (nonparametric) bootstrap-based confidence intervals.

Discussion

We have demonstrated in this paper that social contact data can be used to inform transmission parameters and to estimate age-specific characteristics of SARS-CoV-2 transmission. More specifically, the next generation approach enables us to disentangle age-specific differences in transmission rates while relying on temporal changes in social contact behaviour measured using consecutive waves of a social contact panel study. Clearly, SARS-CoV-2 transmission is partly influenced by age-specific differences in contact behaviour, but importantly, additional age-specific factors related to susceptibility and infectiousness, in a broad sense, are quintessential to account for. We have shown that such factors imply a smaller susceptibility for children as compared to adults, with the estimated susceptibility in children being around half of the susceptibility in adults, and even less for very young children (Fig 1). This result is in accordance with a very recent result obtained using CoMix social contact data in England but using a calibration on the reproduction number instead of the next generation approach [17]. With respect to that, we assessed the impact of assuming homogeneous transmission parameters on the reproduction number, showing how (age-)heterogeneous parameters are necessary to correctly align the reproduction number from the CoMix data and the reproduction number estimated from infections or hospitalizations. Moreover, our method is able to estimate time-varying transmission parameters and it shows a gradual decrease in susceptibility of adults in line with the progression of the Belgian vaccination campaign (Fig 4). This decrease implies a progressive change in the dynamics of the epidemic with largely unvaccinated childhood age groups gradually becoming more important drivers of SARS-CoV-2 transmission than predominantly vaccinated adult age groups.

However, our method suffers from several limitations. A potential bias which needs to be acknowledged is the use of PCR data which correspond to the observed relative incidence and do not necessarily correspond to the true relative incidence as each age class is not necessarily tested in the same way, even if we discard periods of strong variation in testing policy. Indeed, even in the absence of a change in testing policy, age-specific differences in symptomatology, disease severity and the probability of developing symptoms upon infection lead to different shares of symptomatic and asymptomatic cases to be detected. Despite the fact that we can infer q-susceptibility and q-infectiousness from the observed PCR test data, we cannot further disentangle both components by estimating the aforementioned quantities simultaneously. By comparing the two separate approaches, the estimation of the relative q-susceptibility seems most informative, since proportionality factors are better constrained by the data. More specifically, the estimated q-susceptibility was identifiable when fitting to real incidence data while q-infectiousness estimates were estimated to be zero for certain age classes, which seems an artefact of the methodology. Another limitation of the proposed method is that a further decomposition of q-susceptibility (or q-infectiousness) in clinical susceptibility (infectiousness) and other external factors relevant for transmission between susceptible and infectious persons is difficult, at least without availability of relevant additional data thereon. Nonetheless, an assessment and quantification of the (relative) q-susceptibility, q-infectiousness and the corresponding relative incidence provides useful insights into heterogeneous SARS-CoV-2 transmission dynamics.

Materials and methods

Social contact data

Our study is based on Belgian social contact data collected within the CoMix survey [7,8] during the epidemic period between December 2020 and May 2021. These data are processed by means of the online Socrates tool [8,21,22]. Participants were asked to fill in a contact dairy including all contacts made during a specific day, reporting the type of contact, location, and age of the contacted person, with a contact defined as an in-person conversation of at least a few words, or a skin-to-skin contact. The CoMix survey was repeatedly performed in different waves and different survey periods. More specifically, an initial survey period containing 8 waves was carried out between 4 March 2020 and 27 July 2020 targeting adults only. A second survey period, still ongoing in 2021, began on 11 November 2020 targeting participants of all ages. The waves are conducted with an interval of two weeks (14 days). For more detailed information on the CoMix survey and the stratification process, the reader is referred to [7,8]. A detailed timetable of the CoMix waves and survey periods is presented in Table A of S1 Appendix. A schematic timeline of CoMix waves according to the evolution of the alpha variant of concern and vaccination campaign in Belgium is presented in Fig A of S1 Appendix.

We use the following notation. N_i denotes the number of individuals in the Belgian population of age i according to Belgian demographic data [23] and integrated into the Socrates tool [22]. In general, we use subscripts i as an index for the participant's age, and j as an index for the contacted person's age. The following observable quantities (dependent on the wave chosen) can be extracted from the survey:

- m_{ij} represents the average daily number of individuals of age j who are contacted by a participant of age i. The elements m_{ij} constitute a matrix $\mathbf{M}$ called *social contact matrix*.
- c_{ij} is the per capita contact rate per day for participants of age i with persons of age j in the population. The elements c_{ij} constitute a matrix $\mathbf{C}$ called the *contact rate matrix*. This matrix is related to the social contact matrix by the relation $c_{ij} = m_{ij}/N_j$.

In theory, due to the reciprocal nature of contacts, the total number of contacts between members of two age classes, as reported by participants in each of the age groups, must be equal, hence $N_i m_{ij} = N_i c_{ij} N_j = N_j c_{ji} N_i = N_j m_{ji}$, which is equivalent to the condition that the contact rate matrix should be symmetric, i.e., $c_{ij} = c_{ji}$, $\forall i, j$. The social contact matrix $\mathbf{M}$ respects the relation $N_i m_{ij} = N_j m_{ji}$, but is in general not symmetric due to differences in N_i and N_j. In practice, the observed total number of contacts $N_i m_{ij}^{\text{raw}}$ and $N_j m_{ji}^{\text{raw}}$ are not necessarily equal due to sampling bias, hence, we calculate the reciprocal social contact matrix by:

$$m_{ij} = \frac{m_{ij}^{\text{raw}} N_i + m_{ji}^{\text{raw}} N_j}{2N_i}.$$

All these notations and definitions are similar to those described in detail in [24], except that the subscripts i and j and order of indices are inverted here such that the definition of the social contact matrix $\mathbf{M}$ corresponds to the natural output of the Socrates tool [22].

Next generation principle

The *social contact hypothesis* [11] implies that the age-specific number of infectious contacts is proportional to the self-reported age-specific number of social contacts. There are two ways to interpret empirical social contact survey data in light of this hypothesis: either survey participants can be infected by their infectious contacts or participants can infect their susceptible contacts. Here, we consider the first interpretation as initial definition – since the CoMix survey did not specifically target infected persons, and symptomatic participants may have been less likely to participate in the survey at the height of their symptoms. However, we will show that the two interpretations lead to the same mathematical result under the assumption of reciprocity of social contacts.

If we denote by w_j the incidence within age class j over a short observation interval (e.g. corresponding to a wave period), then $v_j = w_j/N_j$ is the risk of being infected during the observation interval for that age class (incidence rate or force of infection). The new generation of infected people is given by:

$$w'_i = \sum_j q\, N_i\, m_{ij}\, v_j = \sum_j q\, m_{ij} \frac{N_i}{N_j}\, w_j,$$

where q is a general proportionality factor completely defining the relationship between infection and contact events. The q-factor accommodates several effects such as susceptibility to infection, infectiousness upon infection, duration of the infectious period, type and effectiveness of contacts, seasonality, pre-existing natural and vaccine-induced immunity, etc.

The elements $k_{ij} = q\, m_{ij} \frac{N_i}{N_j} = qN_i c_{ij}$ define a matrix $\mathbf{K}$ called the *next generation matrix* (or *reproduction matrix*) since k_{ij} represents the mean number of individuals of age i that are infected through a single individual of age j during their entire infectious period (for which the time between consecutive generations of infected individuals is chosen to be equal to the average duration of infectiousness).

Note that under the reciprocity assumption leading to a symmetric matrix $\mathbf{C}$, the relation $N_i m_{ij} = N_j m_{ji}$ provides:

$$k_{ij} = q\, m_{ij} \frac{N_i}{N_j} = q\, m_{ji} \qquad \text{or} \qquad \mathbf{K} = q\, \mathbf{M}^T,$$

corresponding to the second interpretation that survey participants (on the right side of the transpose contact matrix $\mathbf{M}^T$) can directly infect their contacts (now on the left side) modulo the proportionality factor. This expression relying on the transpose of the social contact matrix obtained as a direct output of the Socrates tool, $\mathbf{M}$, is chosen because of its better numerical stability.

The recurrence relation of the next generation matrix $\mathbf{K}$:

$$w'' = \mathbf{K}\, w' = \mathbf{K}^2\, w, \qquad w''' = \mathbf{K}^3\, w, \qquad \ldots$$

tends to a stable distribution due to the Perron–Frobenius theorem [25], i.e.,

$$R_t\, w^* = \mathbf{K}\, w^* = q\, \mathbf{M}^T\, w^*,$$

with R_t corresponding to the reproduction number of SARS-CoV-2 [26] which is defined as the leading eigenvalue of the next generation matrix $\mathbf{K}$. More specifically, estimation of the reproduction number R_t and the relative incidence w can be done by computing the leading eigenvalue and corresponding right-eigenvector of $\mathbf{K}$. However, R_t depends on the proportionality factor q, which might be unknown, but the relative incidence w is independent of q and can therefore be directly extracted from the social contact data $\mathbf{M}^T$. The reproduction number is initially the basic reproduction number R_0, but switches to the effective reproduction number as long as social contact data evolve and the proportionality factor q captures the depletion of susceptible. We emphasize here that the eigenvector w is only recovered up to a global constant and therefore individual components w_j have no meaning. What can be interpreted are relative ratios such as w_i/w_j, providing an estimate of the relative incidence in age class i as compared to the incidence in age class j. This vector is usually normalized such that $\sum_i w_i = 1$. In the same way, the incidence rate v_i can be recovered, in relative sense, as the leading left-eigenvector of $\mathbf{M}^T$.

The switch from a homogeneous proportionality factor q to a heterogeneous q_{ij} is performed by assuming:

$$q_{ij} = \tilde{q}\, a_i\, h_j$$

where the vector (a_i) acts on the susceptible side, the vector (h_j) acts on the infectiousness side, and $\tilde{q}$ is a remaining global proportionality factor captures any residual effect. This remaining factor has no influence on the computation of the relative incidence w. However, due to the presence of $\tilde{q}$, the vectors (a_i) and (h_j) only have a relative interpretation.

The heterogeneous next generation matrix is defined as:

$$k_{ij} = \tilde{q}\, a_i\, m_{ij} \frac{N_i}{N_j}\, h_j \qquad \text{or} \qquad \mathbf{K} = \tilde{q}\ \mathrm{diag}(a_i)\, \mathbf{M}^T\, \mathrm{diag}(h_j).$$

We note that we are working here with a next generation matrix with *small domain*. There also exists a next generation matrix with *large domain* taking explicitly into account the different states of the disease and their duration for each age class [4]. However, the small domain approach is appropriate here since we do not work with a dynamical system and heterogeneity in disease duration is part of the effects captured by the proportionality factors.

Estimating relative q-susceptibility (a_i) and q-infectiousness (h_j) from COVID-19 age-structured indicators

The vectors (a_i) and (h_j) have an important impact on the determination of the leading right eigenvector in the system:

$$\mathbf{K}\, w^* = \tilde{q}\ \mathrm{diag}(a_i)\, \mathbf{M}^T\, \mathrm{diag}(h_j)\, w^* = R_t\, w^*.$$

The obtained relative incidence w^* can be compared with the normalized relative incidence $\tilde{w}$ estimated from the observed incidence data in Belgium. Using this approach, we are able to determine q-susceptibility and q-infectiousness corresponding to SARS-CoV-2 transmission in Belgium. However, (a_i) and (h_j) vectors cannot be estimated simultaneously in a unique way from this process since there remains an indeterminacy [2, 12]. Nevertheless, the identifiability problem can be solved by imposing a constraint on one of the two vectors.

For this study, we choose each time a heterogeneous constraint coming from the literature as well as a homogeneous constraint (whose results are only presented in S1 Appendix). The heterogeneous constraints are defined from the following assumptions:

- For the assumption on infectiousness h_j (estimation of q-susceptibility parameters a_i): We consider the probability of an asymptomatic COVID-19 infection in case of SARS-CoV-2 exposure in the Belgian population to be

  $$p = (0.94, 0.92, 0.90, 0.84, 0.61, 0.49, 0.21, 0.02, 0.02, 0.02)$$

  as assumed in [14] using data from [27]. Assuming that the relative infectiousness of asymptomatic versus symptomatic individuals is 0.51 [14], we obtain the following constraint:

  $$(h_j) = 0.51p + 1(1-p) = (0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99).$$

- For the assumption on susceptibility a_i (estimation of q-infectiousness parameters h_j): The assumption is taken from [1]:

  $$(a_i) = (0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74).$$

Data and fitting procedure

We use Belgian data on daily incidence of COVID-19 confirmed by means of a positive PCR test, as provided by the Belgian Institute for Public Health, Sciensano [19]. In order to reduce testing biases, the period of study is restricted to a period with almost constant testing policy (with testing allowed for both symptomatic and asymptomatic people and before biases induced by the introduction of the EU Digital COVID Certificate). Since there is a delay between a change in social contact behaviour and its effect on the relative incidence, we consider PCR test results for the period starting 7 days after the onset of a specific CoMix wave and lasting for 14 days thereafter.

Concerning social contact data, the initial CoMix survey waves (1 to 8) are discarded due to a variable testing policy and lack of information regarding child-child contacts. The three subsequent waves (9 to 11) are also discarded since, despite the introduction of measuring child-child contacts, the information was collected

using a different survey formulation. CoMix waves 12 to 23 correspond to a period with constant testing policy, an identical survey design as well as without vaccination in children, which implies that the results with regard to age classes $[0-6)$, $[6-12)$ and $[12-18)$ years are expected to be more stable. The start of wave 12 corresponds to 22 December 2020 when the vaccination campaign in adults has not been started and the last wave considered corresponds to 26 May 2021, with PCR tests considered up to 15 June 2021 (thus when vaccination of the oldest individuals in the Belgian population was nearly completed).

The calibration is performed using the statistical software R with a best fit search using the Hellinger distance (which is suitable for distributions) on relative incidences w^* and $\tilde{w}$. The optimization is done by means of a random walk, until no change in the distance during 100 consecutive iterations is observed, and with an initial homogeneous prior ($a_i = 1$ or $h_i = 1,\ \forall i = 1 \dots 10$) and steps of length $\mathcal{N}\big(0, 0.005^2\big)^{10}$. The sensitivity analysis is performed by repeating the process over 200 nonparametric bootstraps using the previous posterior as new prior. Uncertainty is quantified by means and 95% percentile confidence intervals (i.e., 2.5% and 97.5% quantiles of all bootstrap-based estimates).

Since q-susceptibility and q-infectiousness vectors represent relative values, a normalization process should be chosen for the representation of the results. We choose the following normalization:

- The average $\frac{a_1+a_2+a_3}{3}$ or $\frac{h_1+h_2+h_3}{3}$ of the estimated factors for the age classes [0,6), [6,12) and [12,18) (children, no subject to vaccination during the complete period) is assumed to be constant across all bootstrap runs and wave groups (i.e. all combinations of two CoMix waves) if applicable.

- The mean $\overline{a_4}$ or $\overline{h_4}$ (across the bootstrap runs) for the first adult age class [18,30) is set to one. This constraint is chosen because we mainly want to compare susceptibility and infectiousness of children versus adults while the age class $[18, 30)$ is one of the most stable ones in the bootstrapping process.

R_t evolution from contact patterns

Via the next generation approach, the ratio of the eigenvalues of two next generation matrices can be used to evaluate the relative reduction in the reproduction number. This can be done to compare the time-varying reproduction number derived from the CoMix survey with independent evaluations of the reproduction number. We use as comparison the R_t computed from the daily number of cases [20] and the daily number of hospitalizations [19]. In order to account for the time delays associated with infections and hospitalizations (e.g. time to develop symptoms, time to hospitalisations, etc.), the reproduction number computed from the CoMix social contact data was shifted forward in time. A time shift of 7 (14) days is considered when comparing R_t estimates with the reproduction number computed from the number of confirmed cases (respectively hospitalizations). As the reproduction number is known up to the overall constant $\tilde{q}$, we fix the reproduction number for CoMix wave 12 to be equal to the reproduction number computed from infections or hospitalizations. Uncertainty due to sampling variability is estimated via 10000 nonparametric bootstraps.

Supporting information

S1 Appendix. Supplementary Material. Presentation of the complete results with all figures and tables containing values. Table A: Timetable of CoMix starting dates. Fig A. Schematic timeline of CoMix waves. Figs. B to Q: Estimates of relative proportionality factors and corresponding relative incidence under different assumptions. Table B to I: Exact estimates of the components of the proportionality factors.

Acknowledgments

We thank several researchers from the SIMID COVID-19 consortium (interuniversity collaboration between University of Antwerp (CHERMID) and UHasselt (DSI, CenStat) as well as other researchers from the Interuniversity Institute of Biostatistics and statistical Bioinformatics (I-BioStat) (KU Leuven and UHasselt) for numerous constructive discussions and meetings. The authors thank the EpiPose consortium partners for useful discussions and for help in setting up the CoMix survey as part of EpiPose. The authors are also very grateful for access to the data from the Belgian Scientific Institute for Public Health, Sciensano.

Author contributions

Conceptualisation: NF, SA, FB, NH; Literature research: NF, LA, AL; Data preparation and/or collection: PC, LW; Data analysis code: NF, PC; Coordination: NH; All authors interpreted the findings, contributed to writing the manuscript and approved the final version of the manuscript for publication.

Financial support

This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program – Project TransMID (NF, PC and NH, Grant Agreement 682540). This project was also funded by the European Union's Horizon 2020 Research and Innovations Programme – Project EpiPose, Epidemic intelligence to Minimize COVID-19's Public Health, Societal and Economical Impact (NH, grant number 101003688). LW gratefully acknowledges support from the Research Foundation Flanders (FWO) (postdoctoral fellowships 1234620N). NH acknowledges funding from the Antwerp Study Center for Infectious Diseases (ASCID) and the Methusalem-Centre of Excellence consortium VAX–IDEA. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests

None.

Data availability statement

R codes and all necessary data to run the codes (potentially aggregate) are available on GitHub at `https://github.com/nicolas-franco-unamur/Next-gen`. CoMix data and age-class re-aggregate PCR tests data are provided with the code. CoMix social contact data are also available via `http://www.socialcontactdata.org`. PCR tests data are owned by Sciensano, publicly available in aggregate form at `https://epistat.wiv-isp.be/covid/` while non-aggregate data can be requested via the online form `https://epistat.wiv-isp.be/datarequest`.

References

[1] Davies NG, Klepac P, Liu Y, Prem K, Jit M, Pearson CAB, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nature Medicine. 2020;26(8):1205–1211. Available from: https://doi.org/10.1038/s41591-020-0962-9.

[2] Flasche S, Hens N, Boëlle PY, Mossong J, van Ballegooijen WM, Nunes B, et al. Different transmission patterns in the early stages of the influenza A(H1N1)v pandemic: A comparative analysis of 12 European countries. Epidemics. 2011;3(2):125–133. Available from: https://www.sciencedirect.com/science/article/pii/S1755436511000235.

[3] Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, et al. A Systematic Review of Social Contact Surveys to Inform Transmission Models of Close-contact Infections. Epidemiology. 2019;30(5). Available from: https://journals.lww.com/epidem/Fulltext/2019/09000/A_Systematic_Review_of_Social_Contact_Surveys_to.15.aspx.

[4] Diekmann O, Heesterbeek JAP. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. vol. 5. John Wiley & Sons; 2000.

[5] Santermans E, Goeyvaerts N, Melegaro A, Edmunds WJ, Faes C, Aerts M, et al. The social contact hypothesis under the assumption of endemic equilibrium: Elucidating the transmission potential of VZV in Europe. Epidemics. 2015;11:14–23. Available from: https://www.sciencedirect.com/science/article/pii/S175543651500002X.

[6] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLOS Medicine. 2008 03;5(3):1–1. Available from: https://doi.org/10.1371/journal.pmed.0050074.

[7] Coletti P, Wambua J, Gimma A, Willem L, Vercruysse S, Vanhoutte B, et al. CoMix: comparing mixing patterns in the Belgian population during and after lockdown. Scientific Reports. 2020 Dec;10(1):21885. Available from: https://doi.org/10.1038/s41598-020-78540-7.

[8] Verelst F, Hermans L, Vercruysse S, Gimma A, Coletti P, Backer JA, et al. 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 Medicine. 2021;19(1):254. Available from: https://doi.org/10.1186/s12916-021-02133-y.

[9] Jarvis CI, Van Zandvoort K, Gimma A, Prem K, Auzenbergs M, O'Reilly K, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Medicine. 2020;18(1):124. Available from: https://doi.org/10.1186/s12916-020-01597-8.

[10] Gimma A, Munday JD, Wong KL, Coletti P, van Zandvoort K, Prem K, et al. CoMix: Changes in social contacts as measured by the contact survey during the COVID-19 pandemic in England between March 2020 and March 2021. medRxiv. 2021;Available from: https://www.medrxiv.org/content/early/2021/05/30/2021.05.28.21257973.

[11] Wallinga J, Teunis P, Kretzschmar M. Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. American Journal of Epidemiology. 2006 09;164(10):936–944. Available from: https://doi.org/10.1093/aje/kwj317.

[12] Goeyvaerts N, Hens N, Ogunjimi B, Aerts M, Shkedy Z, Damme PV, et al. Estimating infectious disease parameters from data on social contacts and serological status. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2010;59(2):255–277. Available from: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2009.00693.x.

[13] Willem L, Abrams S, Libin PJK, Coletti P, Kuylen E, Petrof O, et al. The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19. Nature Communications. 2021;12(1):1524. Available from: https://doi.org/10.1038/s41467-021-21747-7.

[14] Abrams S, Wambua J, Santermans E, Willem L, Kuylen E, Coletti P, et al. Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. Epidemics. 2021;35:100449. Available from: https://www.sciencedirect.com/science/article/pii/S1755436521000116.

[15] Franco N. COVID-19 Belgium: Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts. Epidemics. 2021;37:100490. Available from: https://www.sciencedirect.com/science/article/pii/S1755436521000414.

[16] Coletti P, Libin P, Petrof O, Willem L, Abrams S, Herzog SA, et al. A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies. BMC Infectious Diseases. 2021 May;21(1):503. Available from: https://doi.org/10.1186/s12879-021-06092-w.

[17] Munday JD, Jarvis CI, Gimma A, Wong KLM, van Zandvoort K, Liu Y, et al. Estimating the impact of reopening schools on the reproduction number of SARS-CoV-2 in England, using weekly contact survey data. BMC Medicine. 2021;19(1):233. Available from: https://doi.org/10.1186/s12916-021-02107-0.

[18] Chin T, Feehan DM, Buckee CO, Mahmud AS. Contact surveys reveal heterogeneities in age-group contributions to SARS-CoV-2 dynamics in the United States. medRxiv. 2021;Available from: https://www.medrxiv.org/content/early/2021/09/27/2021.09.25.21264082.

[19] Sciensano, the Belgian public health institute; 2021. https://epistat.wiv-isp.be/covid/.

[20] Data Science Institute UHasselt, COVID-19 Data Dashboard; 2021. https://gjbex.github.io/DSI_UHasselt_covid_dashboard/.

[21] Willem L, Van Hoang T, Funk S, Coletti P, Beutels P, Hens N. SOCRATES: an online tool leveraging a social contact data sharing initiative to assess mitigation strategies for COVID-19. BMC Research Notes. 2020;13(1):293. Available from: https://doi.org/10.1186/s13104-020-05136-9.

[22] Social Contact Rates (SOCRATES) Data Tool: as part of the socialcontactdata.org initiative; 2021. http://www.socialcontactdata.org/socrates-comix/.

[23] StatBel, the Belgian statistical office; 2020. https://statbel.fgov.be/en.

[24] Held L, Hens N, O'Neill P, Jacco W. Handbook of Infectious Disease Data Analysis. Boca Raton: Chapman and Hall/CRC; 2020.

[25] Meyer CD. Matrix Analysis and Applied Linear Algebra. Other Titles in Applied Mathematics. SIAM; 2000. Available from: https://books.google.be/books?id=-7JeAwAAQBAJ.

[26] Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology. 1990;28(4):365–382. Available from: https://doi.org/10.1007/BF00178324.

[27] Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine. 2020;26(4):506–510. Available from: https://doi.org/10.1038/s41591-020-0822-7.

S1 Appendix: Supplementary Material

In this Supplementary Material, we present the complete results obtained on the estimation of heterogeneous proportionality factors using the next generation matrix. Social contact data $\mathbf{M}$ are taken from CoMix waves 12 to 23 as described in the timetable given in Table A and timeline in Fig A. Relative incidences are computed as the leading eigenvector of the next generation matrix $\mathrm{diag}(a_i)\,\mathbf{M}^T\,\mathrm{diag}(h_j)$ and compared with the normalized relative incidence estimated from PCR positive tests. The q-susceptibility (a_i) and q-infectiousness (h_j) vectors are either assumed or estimated. We present 4 different sets of results:

- Estimation of q-susceptibility (a_i) using homogeneous infectiousness assumption $h_j = 1\ \forall j$ (Figs B to E and Tables B and C)
- Estimation of q-susceptibility (a_i) using heterogeneous infectiousness assumption (h_j) coming from the literature (Figs F to I and Tables D and E)
- Estimation of q-infectiousness (h_j) using homogeneous susceptibility assumption $a_i = 1\ \forall i$ (Figs J to M and Tables F and G)
- Estimation of q-infectiousness (h_j) using heterogeneous susceptibility assumption (a_i) coming from the literature (Figs N to Q and Tables H and I)

Each time, we present the estimate of the proportionality factors for the complete period (waves 12 to 23) or estimated by groups of two waves, with the exact values presented in the subsequent tables. The normalization method is described at the beginning of each subsection. We associate to the results the estimate of the relative incidence, with the real data presented in blue, the estimate without any proportionality factor $a_i = h_i = 1\ \forall i$ in green and the estimate with the assumed and estimated proportionality factors in red. Dots represent means and bars represent 95% nonparametric bootstrap confidence intervals (2.5% and 97.5% quantiles).

Survey 1		Survey 2a		Survey 2b	
wave 1	20 April 2020	wave 9	11 November 2020	wave 12	22 December 2020
wave 2	5 May 2020	wave 10	25 November 2020	wave 13	6 January 2021
wave 3	18 May 2020	wave 11	9 December 2020	wave 14	20 January 2021
wave 4	1 June 2020			wave 15	3 February 2021
wave 5	15 June 2020			wave 16	17 February 2021
wave 6	29 June 2020			wave 17	3 March 2021
wave 7	13 July 2020			wave 18	17 March 2021
wave 8	27 July 2020			wave 19	31 March 2021
				wave 20	14 April 2021
				wave 21	28 April 2021
				wave 22	12 May 2021
				wave 23	26 May 2021

Table A. Timetable of CoMix with starting dates. The first survey (waves 1-8) did not include children. The second survey included children with a preliminary version of the survey (waves 9-11) and a modified survey (waves 12-...).

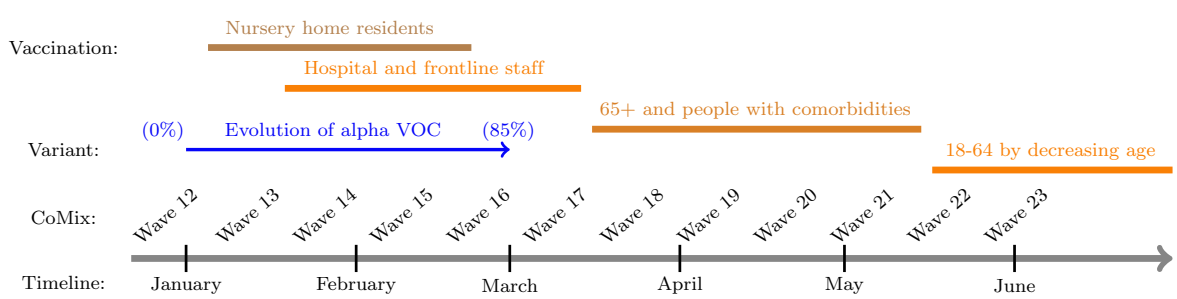

Fig A. Schematic timeline of CoMix waves, vaccination campaign in Belgium and evolution of alpha VOC.

Estimation of q-susceptibility using homogeneous infectiousness

Method: Estimation of the (a_i) relative q-susceptibility vector.
Assumption: homogeneous infectiousness $(h_j) = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.
Normalization method: Mean q-susceptibility among children age classes [0,6), [6,12) and [12,18) is assumed constant among bootstraps and wave groups (if applicable). The mean of the first adult age class [18,30) is set to 1 for the first period.

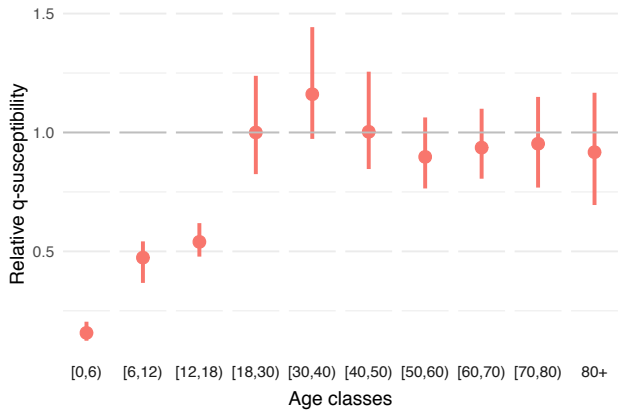

Fig B. Estimation of relative q-susceptibility using assumption on infectiousness $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

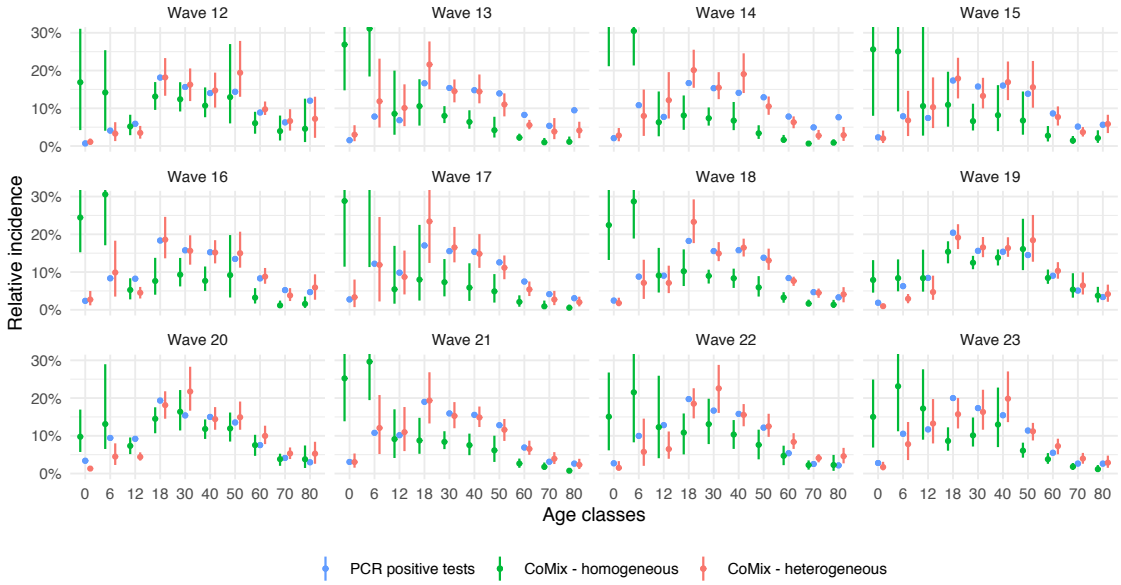

Fig C. Relative incidence using estimated q-susceptibility and assumption on infectiousness $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

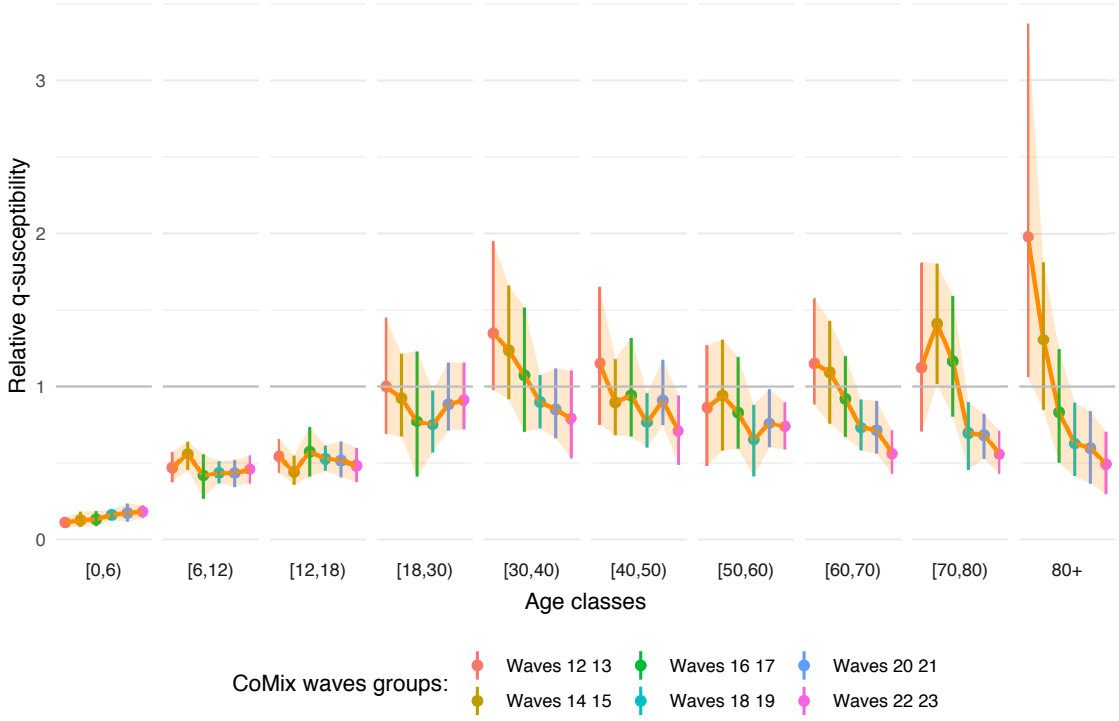

Fig D. Estimation of relative q-susceptibility with time evolution using assumption on infectiousness $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

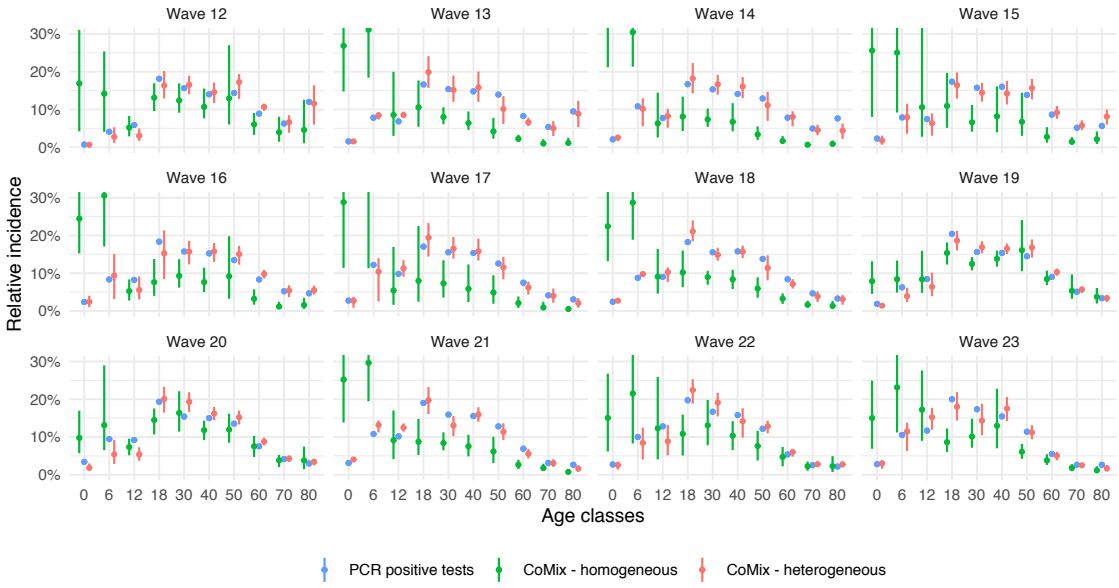

Fig E. Relative incidence using estimated q-susceptibility with time evolution and assumption on infectiousness $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

ageclass	mean	median	sd	lower	upper
[0,6)	0.157	0.155	0.020	0.124	0.204
[6,12)	0.474	0.481	0.045	0.367	0.542
[12,18)	0.540	0.535	0.039	0.478	0.619
[18,30)	1.000	0.993	0.103	0.825	1.238
[30,40)	1.161	1.149	0.122	0.973	1.443
[40,50)	1.003	0.992	0.100	0.846	1.256
[50,60)	0.897	0.889	0.083	0.765	1.063
[60,70)	0.936	0.935	0.073	0.806	1.100
[70,80)	0.953	0.951	0.095	0.768	1.150
80+	0.917	0.914	0.115	0.695	1.167

Table B. Relative q-susceptibility using assumption on infectiousness $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ corresponding to Figure B.

ageclass	mean	median	sd	lower	upper
[0,6)	0.112	0.112	0.015	0.081	0.143
[6,12)	0.469	0.470	0.053	0.374	0.571
[12,18)	0.543	0.543	0.057	0.438	0.659
[18,30)	1.000	0.966	0.201	0.691	1.449
[30,40)	1.347	1.314	0.225	0.978	1.949
[40,50)	1.150	1.114	0.240	0.755	1.649
[50,60)	0.864	0.862	0.211	0.481	1.265
[60,70)	1.149	1.114	0.178	0.885	1.573
[70,80)	1.123	1.113	0.258	0.706	1.806
80+	1.979	1.885	0.583	1.062	3.367

(a) Waves 12 13

ageclass	mean	median	sd	lower	upper
[0,6)	0.128	0.125	0.024	0.084	0.183
[6,12)	0.555	0.561	0.048	0.455	0.636
[12,18)	0.441	0.437	0.051	0.360	0.548
[18,30)	0.926	0.912	0.137	0.675	1.210
[30,40)	1.232	1.219	0.174	0.918	1.657
[40,50)	0.899	0.890	0.124	0.683	1.176
[50,60)	0.942	0.949	0.183	0.579	1.300
[60,70)	1.091	1.087	0.171	0.762	1.427
[70,80)	1.409	1.393	0.199	1.016	1.797
80+	1.303	1.286	0.238	0.850	1.806

(b) Waves 14 15

ageclass	mean	median	sd	lower	upper
[0,6)	0.133	0.131	0.025	0.089	0.186
[6,12)	0.419	0.424	0.073	0.270	0.556
[12,18)	0.572	0.568	0.084	0.409	0.735
[18,30)	0.775	0.761	0.218	0.410	1.225
[30,40)	1.072	1.053	0.217	0.704	1.516
[40,50)	0.944	0.926	0.169	0.675	1.314
[50,60)	0.833	0.826	0.153	0.591	1.189
[60,70)	0.921	0.917	0.133	0.671	1.195
[70,80)	1.162	1.149	0.202	0.807	1.592
80+	0.835	0.826	0.194	0.501	1.240

(c) Waves 16 17

ageclass	mean	median	sd	lower	upper
[0,6)	0.162	0.161	0.018	0.126	0.194
[6,12)	0.435	0.437	0.035	0.368	0.509
[12,18)	0.527	0.526	0.041	0.451	0.612
[18,30)	0.753	0.756	0.111	0.568	0.974
[30,40)	0.901	0.898	0.098	0.726	1.074
[40,50)	0.769	0.761	0.095	0.599	0.958
[50,60)	0.652	0.643	0.129	0.415	0.883
[60,70)	0.734	0.731	0.087	0.580	0.918
[70,80)	0.697	0.694	0.125	0.456	0.900
80+	0.626	0.613	0.129	0.413	0.897

(d) Waves 18 19

ageclass	mean	median	sd	lower	upper
[0,6)	0.173	0.172	0.030	0.117	0.235
[6,12)	0.434	0.434	0.049	0.339	0.515
[12,18)	0.517	0.517	0.063	0.406	0.638
[18,30)	0.886	0.869	0.120	0.712	1.154
[30,40)	0.853	0.839	0.121	0.663	1.117
[40,50)	0.911	0.904	0.112	0.747	1.172
[50,60)	0.762	0.754	0.104	0.601	0.983
[60,70)	0.715	0.712	0.091	0.560	0.909
[70,80)	0.683	0.683	0.078	0.525	0.826
80+	0.595	0.601	0.126	0.366	0.843

(e) Waves 20 21

ageclass	mean	median	sd	lower	upper
[0,6)	0.183	0.182	0.021	0.141	0.224
[6,12)	0.460	0.460	0.047	0.367	0.550
[12,18)	0.482	0.482	0.058	0.376	0.597
[18,30)	0.912	0.904	0.112	0.720	1.154
[30,40)	0.794	0.784	0.153	0.530	1.103
[40,50)	0.710	0.710	0.125	0.490	0.943
[50,60)	0.740	0.745	0.088	0.586	0.899
[60,70)	0.561	0.557	0.070	0.433	0.715
[70,80)	0.557	0.553	0.074	0.432	0.712
80+	0.492	0.486	0.101	0.297	0.704

(f) Waves 22 23

Table C. Relative q-susceptibility with time evolution using assumption on infectiousness $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ corresponding to Figure D.

Estimation of q-susceptibility using heterogeneous infectiousness

Method: Estimation of the (a_i) relative q-susceptibility vector.
Assumption: heterogeneous infectiousness $(h_j) = (0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$ using the proportion of asymptomatic cases in the Belgian population with asymptomatic infectiousness assumed at 0.51 as used in [14] using data from [27].
Normalization method: Mean q-susceptibility among children age classes [0,6), [6,12) and [12,18) is assumed constant among bootstraps and wave groups (if applicable). The mean of the first adult age class [18,30) is set to 1 for the first period.

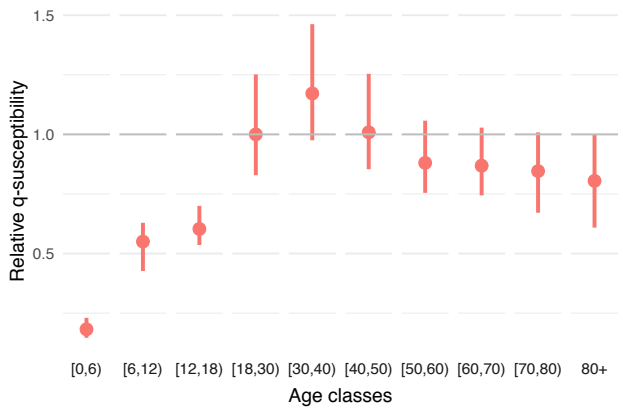

Fig F. Estimation of relative q-susceptibility using assumption on infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$.

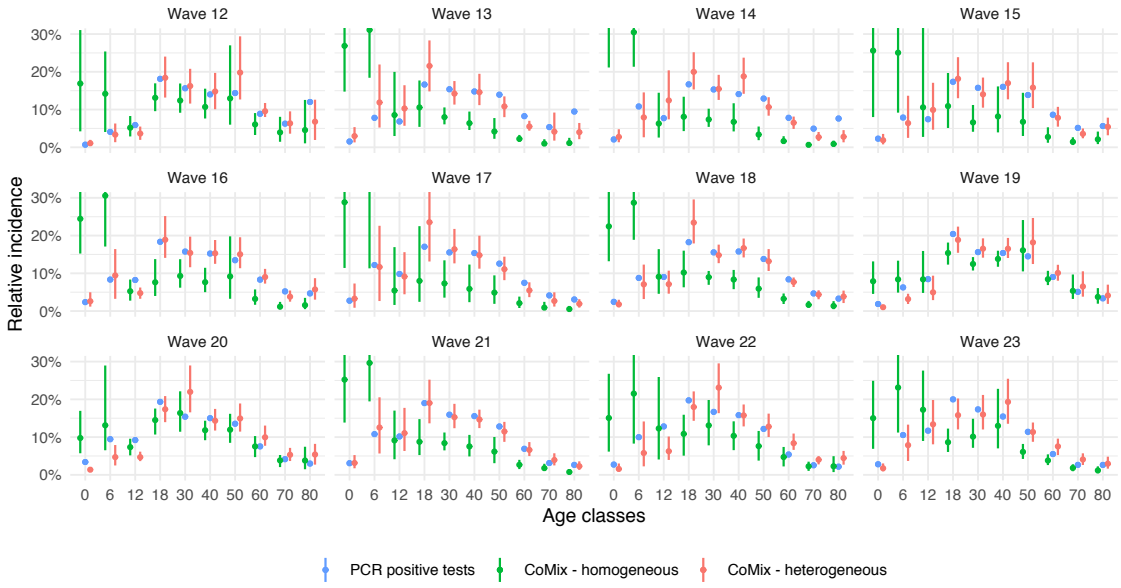

Fig G. Relative incidence using estimated q-susceptibility and assumption on infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$.

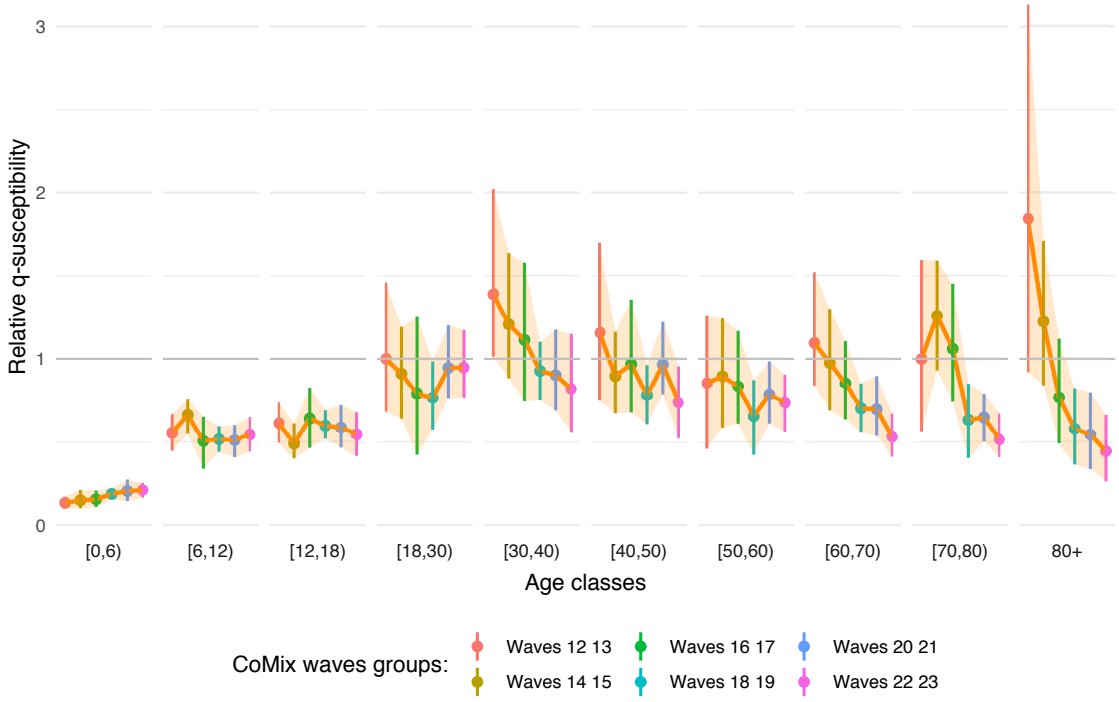

Fig H. Estimation of relative q-susceptibility with time evolution using assumption on infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$.

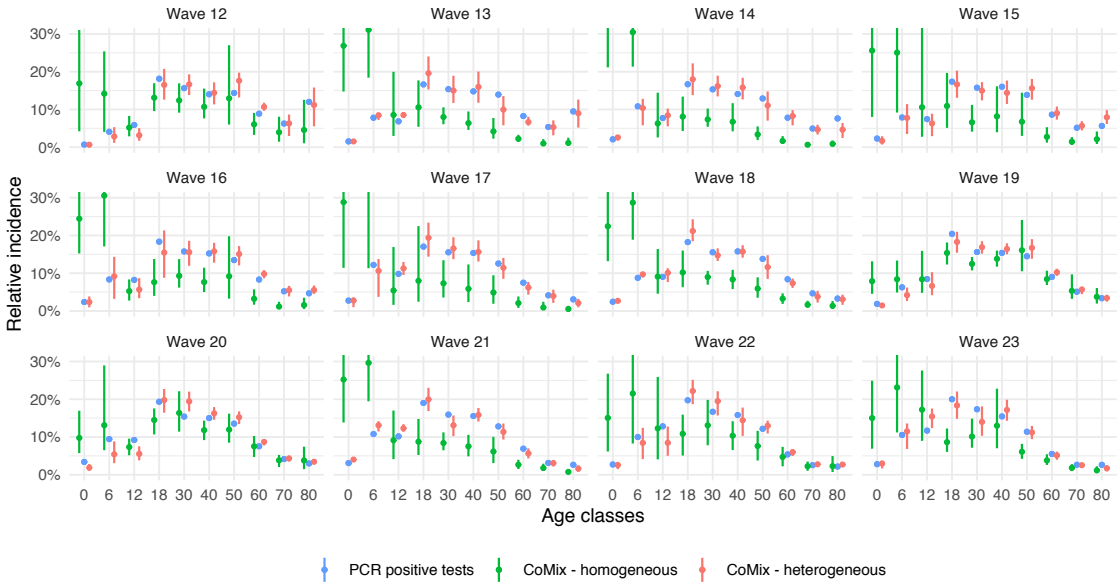

Fig I. Relative incidence using estimated q-susceptibility with time evolution and assumption on infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$.

ageclass	mean	median	sd	lower	upper
[0,6)	0.182	0.180	0.020	0.146	0.230
[6,12)	0.550	0.559	0.051	0.427	0.629
[12,18)	0.603	0.599	0.044	0.536	0.700
[18,30)	1.000	0.993	0.102	0.829	1.252
[30,40)	1.172	1.162	0.120	0.975	1.462
[40,50)	1.009	0.997	0.100	0.854	1.254
[50,60)	0.880	0.875	0.082	0.755	1.057
[60,70)	0.869	0.864	0.070	0.744	1.028
[70,80)	0.846	0.849	0.090	0.671	1.009
80+	0.805	0.798	0.101	0.609	1.000

Table D. Relative q-susceptibility using assumption on infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$ corresponding to Figure F.

ageclass	mean	median	sd	lower	upper
[0,6)	0.134	0.134	0.018	0.103	0.172
[6,12)	0.555	0.556	0.059	0.449	0.666
[12,18)	0.613	0.616	0.064	0.495	0.735
[18,30)	1.000	0.971	0.208	0.680	1.454
[30,40)	1.387	1.346	0.238	1.012	2.019
[40,50)	1.157	1.125	0.249	0.746	1.695
[50,60)	0.853	0.853	0.216	0.461	1.255
[60,70)	1.096	1.066	0.169	0.839	1.519
[70,80)	0.999	0.976	0.249	0.563	1.593
80+	1.844	1.763	0.567	0.920	3.127

(a) Waves 12 13

ageclass	mean	median	sd	lower	upper
[0,6)	0.151	0.148	0.027	0.103	0.212
[6,12)	0.660	0.663	0.056	0.548	0.753
[12,18)	0.491	0.484	0.058	0.402	0.611
[18,30)	0.911	0.912	0.136	0.639	1.191
[30,40)	1.205	1.189	0.178	0.883	1.637
[40,50)	0.896	0.888	0.123	0.670	1.161
[50,60)	0.895	0.893	0.177	0.585	1.241
[60,70)	0.975	0.980	0.155	0.687	1.299
[70,80)	1.258	1.257	0.181	0.931	1.590
80+	1.222	1.198	0.231	0.841	1.715

(b) Waves 14 15

ageclass	mean	median	sd	lower	upper
[0,6)	0.157	0.157	0.026	0.110	0.209
[6,12)	0.506	0.513	0.081	0.341	0.650
[12,18)	0.639	0.630	0.093	0.466	0.822
[18,30)	0.789	0.785	0.221	0.428	1.250
[30,40)	1.113	1.089	0.222	0.742	1.577
[40,50)	0.968	0.957	0.170	0.676	1.353
[50,60)	0.834	0.825	0.148	0.609	1.166
[60,70)	0.853	0.854	0.122	0.634	1.105
[70,80)	1.060	1.041	0.187	0.739	1.446
80+	0.765	0.746	0.159	0.491	1.119

(c) Waves 16 17

ageclass	mean	median	sd	lower	upper
[0,6)	0.189	0.188	0.018	0.153	0.224
[6,12)	0.516	0.520	0.039	0.444	0.593
[12,18)	0.597	0.593	0.044	0.517	0.688
[18,30)	0.763	0.761	0.113	0.574	0.984
[30,40)	0.927	0.925	0.097	0.748	1.101
[40,50)	0.779	0.775	0.093	0.606	0.961
[50,60)	0.654	0.645	0.123	0.428	0.872
[60,70)	0.699	0.701	0.078	0.555	0.852
[70,80)	0.632	0.640	0.122	0.405	0.851
80+	0.578	0.577	0.122	0.365	0.819

(d) Waves 18 19

ageclass	mean	median	sd	lower	upper
[0,6)	0.206	0.204	0.033	0.146	0.274
[6,12)	0.511	0.513	0.055	0.407	0.601
[12,18)	0.585	0.588	0.069	0.467	0.719
[18,30)	0.946	0.936	0.119	0.763	1.201
[30,40)	0.902	0.900	0.129	0.688	1.174
[40,50)	0.965	0.953	0.115	0.785	1.219
[50,60)	0.782	0.776	0.102	0.610	0.984
[60,70)	0.694	0.691	0.091	0.537	0.896
[70,80)	0.645	0.651	0.077	0.502	0.787
80+	0.543	0.530	0.122	0.338	0.796

(e) Waves 20 21

ageclass	mean	median	sd	lower	upper
[0,6)	0.211	0.210	0.023	0.167	0.255
[6,12)	0.545	0.545	0.054	0.447	0.649
[12,18)	0.546	0.546	0.066	0.417	0.678
[18,30)	0.948	0.929	0.111	0.766	1.172
[30,40)	0.819	0.810	0.153	0.559	1.149
[40,50)	0.737	0.735	0.122	0.524	0.954
[50,60)	0.737	0.736	0.088	0.562	0.905
[60,70)	0.532	0.530	0.067	0.414	0.668
[70,80)	0.516	0.507	0.071	0.410	0.669
80+	0.446	0.439	0.099	0.266	0.660

(f) Waves 22 23

Table E. Relative q-susceptibility with time evolution using assumption on infectiousness $(0.54, 0.55, 0.56, 0.59, 0.7, 0.76, 0.9, 0.99, 0.99, 0.99)$ corresponding to Figure H.

Estimation of q-infectiousness using homogeneous susceptibility

Method: Estimation of the (h_j) relative q-infectiousness vector.
Assumption: homogeneous susceptibility $(a_i) = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.
Normalization method: No normalization accross bootstraps and wave groups is applied here since the q-infectiousness among children age classes is estimated at $(0, 0, 0)$ for several bootstraps and this prevents using the same normaliztion method than in other subsections. The mean of the first adult age class [18,30) is set to 1 for the first period.

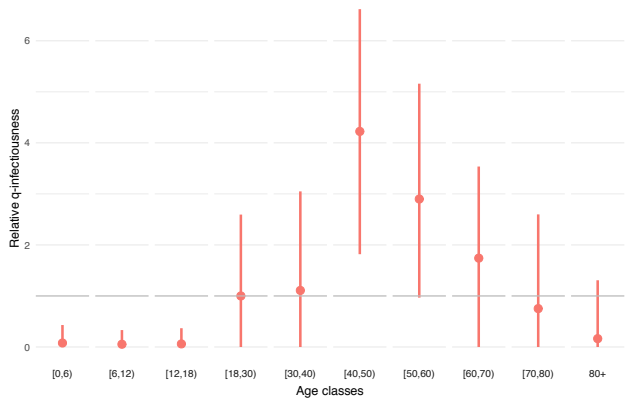

Fig J. Estimation of relative q-infectiousness using assumption on susceptibility $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

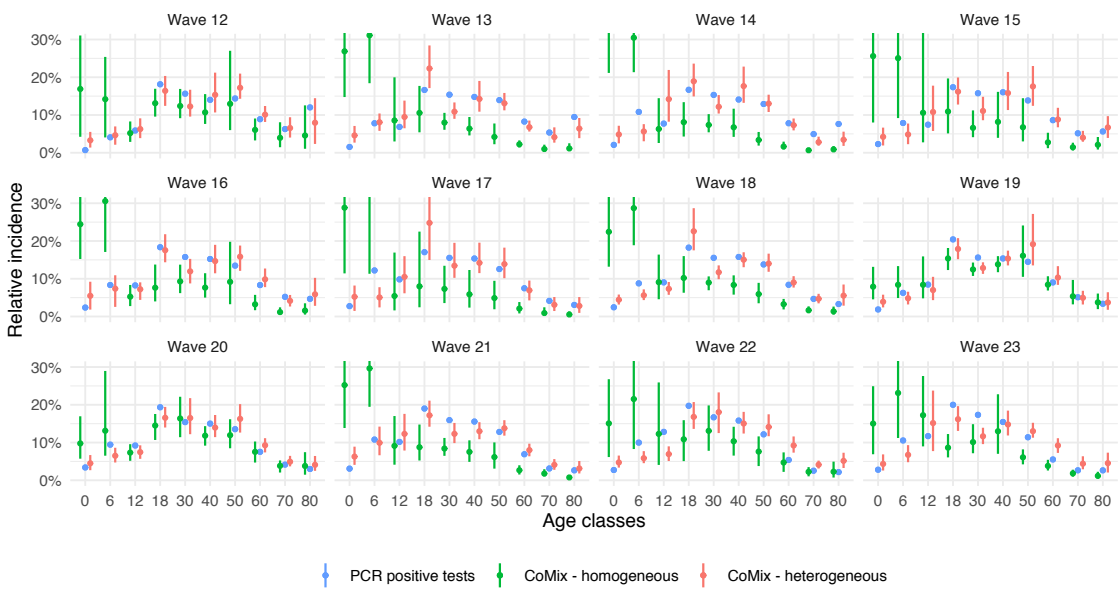

Fig K. Relative incidence using estimated q-infectiousness and assumption on susceptibility $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

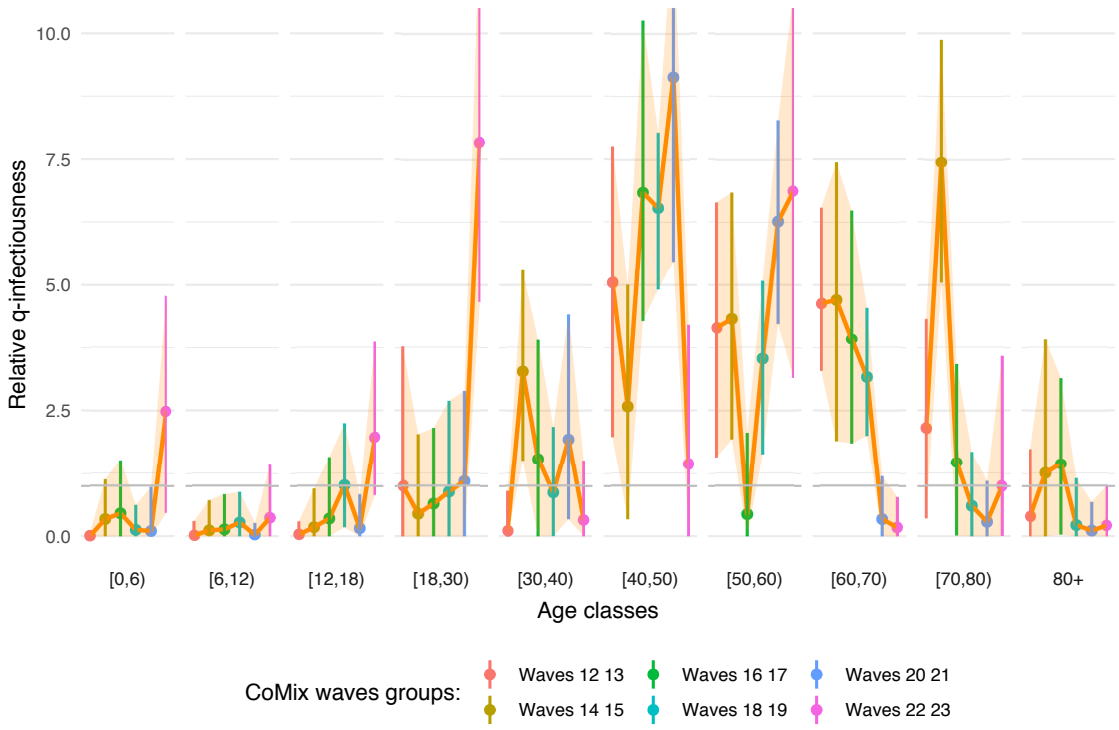

Fig L. Estimation of relative q-infectiousness with time evolution using assumption on susceptibility $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

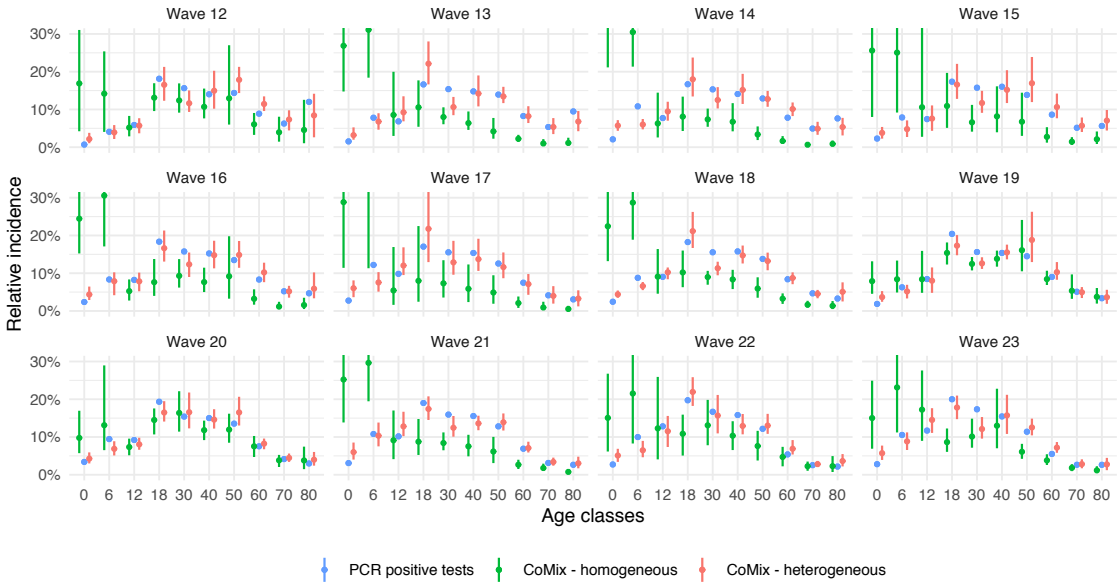

Fig M. Relative incidence using estimated q-infectiousness with time evolution and assumption on susceptibility $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$.

ageclass	mean	median	sd	lower	upper
[0,6)	0.078	0.000	0.129	0.000	0.430
[6,12)	0.054	0.000	0.111	0.000	0.331
[12,18)	0.060	0.000	0.106	0.000	0.368
[18,30)	1.000	0.872	0.733	0.000	2.594
[30,40)	1.108	0.991	0.910	0.000	3.049
[40,50)	4.224	4.220	1.176	1.818	6.619
[50,60)	2.899	2.896	1.144	0.968	5.159
[60,70)	1.741	1.694	0.882	0.000	3.536
[70,80)	0.752	0.513	0.810	0.000	2.598
80+	0.164	0.000	0.372	0.000	1.307

Table F. Relative q-infectiousness using assumption on susceptibility $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ corresponding to Figure J.

ageclass	mean	median	sd	lower	upper
[0,6)	0.009	0.000	0.034	0.000	0.071
[6,12)	0.018	0.000	0.086	0.000	0.300
[12,18)	0.036	0.000	0.103	0.000	0.295
[18,30)	1.000	0.639	1.136	0.000	3.763
[30,40)	0.108	0.012	0.232	0.000	0.897
[40,50)	5.047	5.173	1.413	1.977	7.763
[50,60)	4.143	4.061	1.254	1.574	6.643
[60,70)	4.625	4.608	0.839	3.281	6.540
[70,80)	2.149	2.073	1.013	0.353	4.318
80+	0.394	0.154	0.556	0.000	1.739

(a) Waves 12 13

ageclass	mean	median	sd	lower	upper
[0,6)	0.340	0.267	0.330	0.000	1.124
[6,12)	0.117	0.023	0.214	0.000	0.711
[12,18)	0.180	0.029	0.271	0.000	0.943
[18,30)	0.445	0.147	0.632	0.000	2.034
[30,40)	3.270	3.137	1.012	1.508	5.281
[40,50)	2.581	2.458	1.152	0.335	5.005
[50,60)	4.321	4.263	1.242	1.930	6.838
[60,70)	4.689	4.919	1.575	1.894	7.435
[70,80)	7.442	7.427	1.299	5.041	9.871
80+	1.252	1.047	1.024	0.000	3.903

(b) Waves 14 15

ageclass	mean	median	sd	lower	upper
[0,6)	0.453	0.359	0.474	0.000	1.516
[6,12)	0.137	0.004	0.236	0.000	0.831
[12,18)	0.350	0.172	0.468	0.000	1.580
[18,30)	0.644	0.566	0.581	0.000	2.159
[30,40)	1.541	1.399	1.168	0.000	3.898
[40,50)	6.836	6.716	1.669	4.277	10.259
[50,60)	0.436	0.114	0.613	0.000	2.063
[60,70)	3.914	3.893	1.357	1.850	6.483
[70,80)	1.468	1.447	0.988	0.016	3.418
80+	1.439	1.362	0.810	0.034	3.138

(c) Waves 16 17

ageclass	mean	median	sd	lower	upper
[0,6)	0.128	0.042	0.180	0.000	0.618
[6,12)	0.272	0.227	0.238	0.000	0.876
[12,18)	1.008	0.938	0.523	0.178	2.250
[18,30)	0.880	0.649	0.870	0.000	2.692
[30,40)	0.864	0.774	0.601	0.003	2.178
[40,50)	6.531	6.746	0.835	4.912	8.031
[50,60)	3.524	3.749	0.945	1.635	5.077
[60,70)	3.161	3.088	0.694	1.998	4.534
[70,80)	0.600	0.447	0.542	0.000	1.683
80+	0.221	0.084	0.315	0.000	1.147

(d) Waves 18 19

ageclass	mean	median	sd	lower	upper
[0,6)	0.105	0.001	0.268	0.000	0.971
[6,12)	0.031	0.000	0.112	0.000	0.263
[12,18)	0.158	0.026	0.254	0.000	0.826
[18,30)	1.088	0.833	0.897	0.000	2.888
[30,40)	1.928	1.887	0.924	0.336	4.407
[40,50)	9.128	9.524	1.643	5.465	12.172
[50,60)	6.266	6.296	1.093	4.218	8.278
[60,70)	0.338	0.178	0.386	0.000	1.188
[70,80)	0.285	0.156	0.307	0.000	1.093
80+	0.113	0.029	0.263	0.000	0.675

(e) Waves 20 21

ageclass	mean	median	sd	lower	upper
[0,6)	2.481	2.357	1.230	0.461	4.773
[6,12)	0.365	0.219	0.450	0.000	1.416
[12,18)	1.966	1.842	0.794	0.811	3.857
[18,30)	7.831	7.097	2.411	4.653	12.346
[30,40)	0.322	0.108	0.561	0.000	1.513
[40,50)	1.436	1.264	1.076	0.000	4.204
[50,60)	6.863	7.341	2.154	3.143	10.833
[60,70)	0.174	0.079	0.305	0.000	0.773
[70,80)	0.999	0.672	0.992	0.007	3.578
80+	0.214	0.104	0.274	0.000	1.010

(f) Waves 22 23

Table G. Relative q-infectiousness with time evolution using assumption on susceptibility $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ corresponding to Figure L.

Estimation of q-infectiousness using heterogeneous susceptibility

Method: Estimation of the (h_j) relative q-infectiousness vector.
Assumption: heterogeneous susceptibility $(a_i) = (0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$ taken from [1].
Normalization method: Mean q-infectiousness among children age classes [0,6), [6,12) and [12,18) is assumed constant among bootstraps and wave groups (if applicable). The mean of the first adult age class [18,30) is set to 1 for the first period.

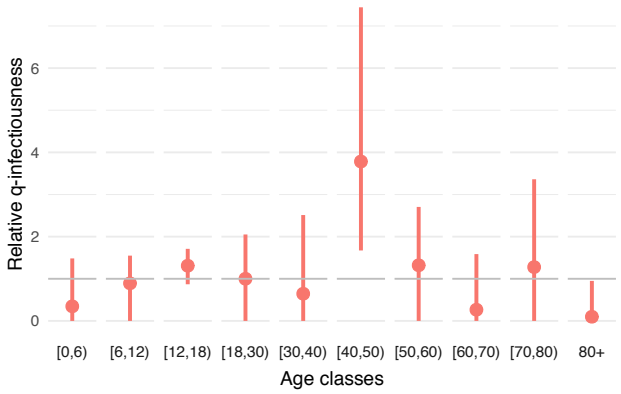

Fig N. Estimation of relative q-infectiousness using assumption on susceptibility $(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$.

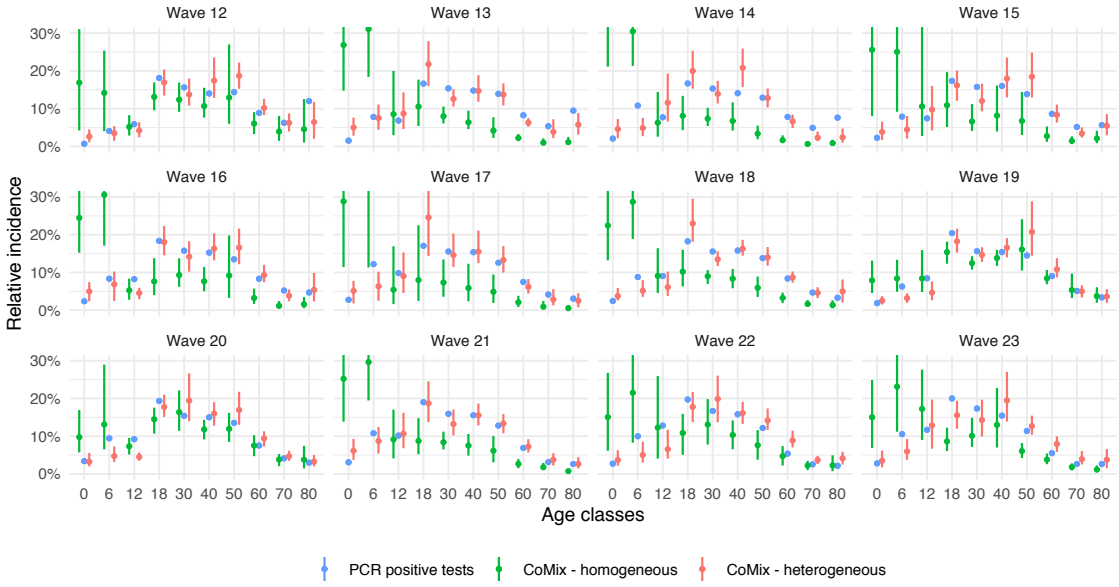

Fig O. Relative incidence using estimated q-infectiousness and assumption on susceptibility $(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$.

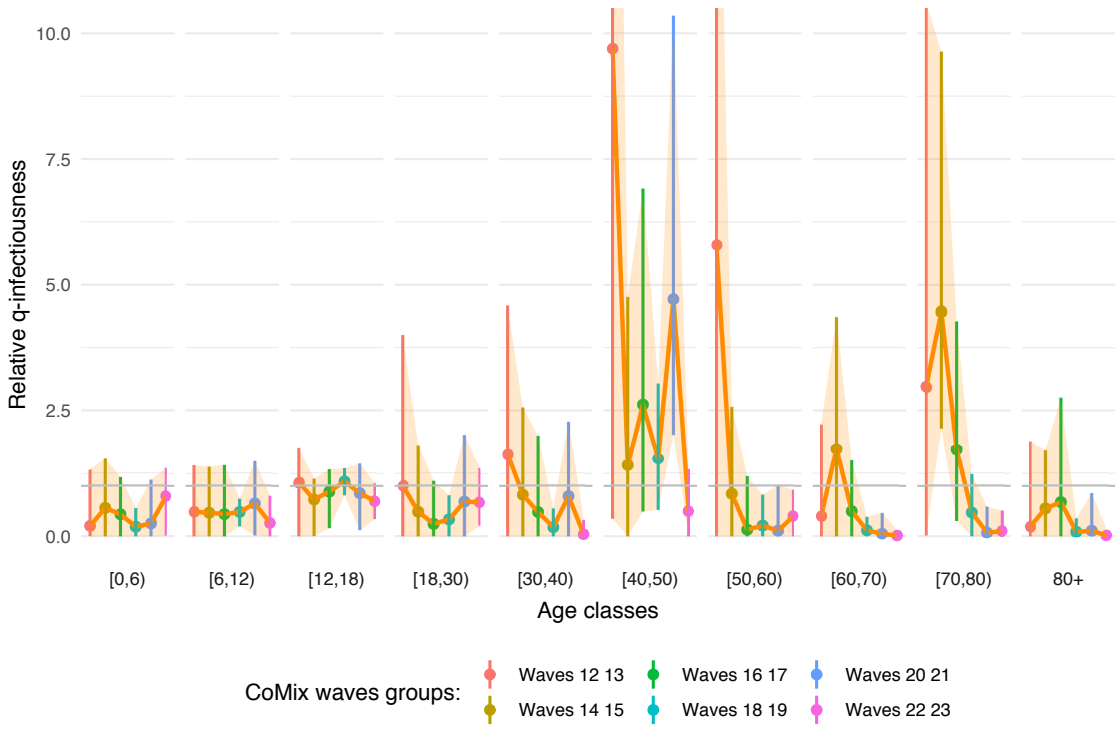

Fig P. Estimation of relative q-infectiousness with time evolution using assumption on susceptibility $(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$.

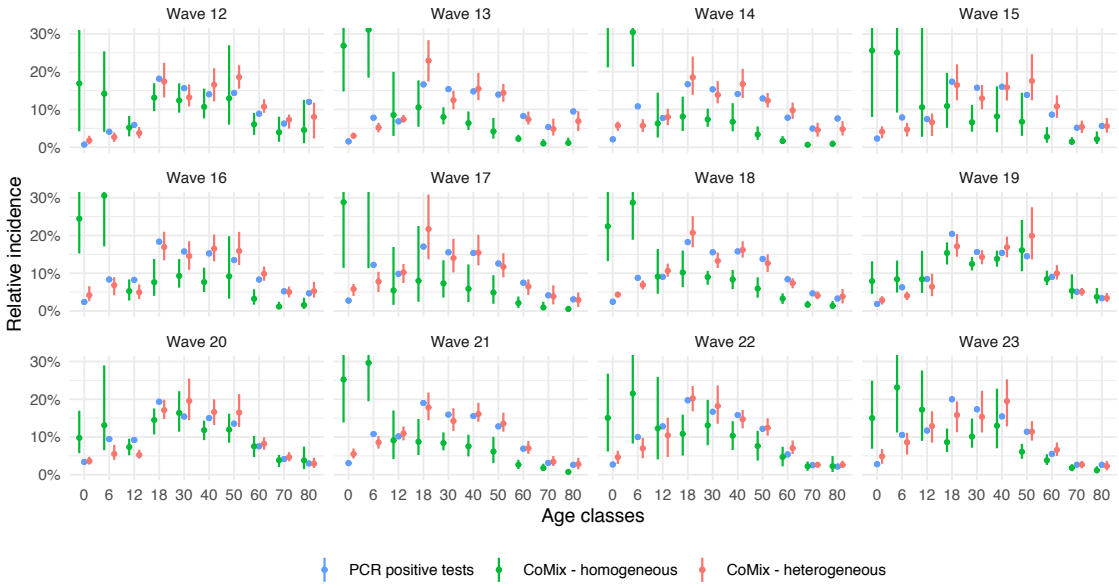

Fig Q. Relative incidence using estimated q-infectiousness with time evolution and assumption on susceptibility $(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$.

ageclass	mean	median	sd	lower	upper
[0,6)	0.346	0.034	0.461	0.000	1.483
[6,12)	0.892	0.957	0.404	0.000	1.549
[12,18)	1.310	1.326	0.231	0.870	1.710
[18,30)	1.000	0.967	0.544	0.000	2.051
[30,40)	0.645	0.375	0.811	0.000	2.512
[40,50)	3.783	3.541	1.586	1.672	7.443
[50,60)	1.320	1.257	0.741	0.000	2.705
[60,70)	0.266	0.000	0.445	0.000	1.586
[70,80)	1.277	1.155	0.903	0.000	3.362
80+	0.099	0.000	0.279	0.000	0.950

Table H. Relative q-infectiousness using assumption on susceptibility $(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$ corresponding to Figure N.

ageclass	mean	median	sd	lower	upper
[0,6)	0.204	0.013	0.363	0.000	1.319
[6,12)	0.483	0.454	0.383	0.000	1.402
[12,18)	1.061	1.070	0.443	0.001	1.748
[18,30)	1.000	0.625	1.259	0.000	3.990
[30,40)	1.625	0.052	12.906	0.000	4.577
[40,50)	9.694	3.895	49.732	0.345	25.974
[50,60)	5.792	2.434	31.815	0.000	17.774
[60,70)	0.398	0.063	0.802	0.000	2.223
[70,80)	2.973	2.087	4.053	0.018	10.597
80+	0.189	0.000	0.505	0.000	1.891

(a) Waves 12 13

ageclass	mean	median	sd	lower	upper
[0,6)	0.560	0.547	0.453	0.000	1.535
[6,12)	0.464	0.404	0.436	0.000	1.370
[12,18)	0.724	0.760	0.278	0.005	1.134
[18,30)	0.484	0.287	0.554	0.000	1.814
[30,40)	0.818	0.700	0.695	0.000	2.559
[40,50)	1.404	1.172	1.127	0.000	4.761
[50,60)	0.838	0.739	0.711	0.000	2.570
[60,70)	1.737	1.647	1.033	0.000	4.353
[70,80)	4.463	4.026	1.995	2.142	9.635
80+	0.549	0.501	0.550	0.000	1.707

(b) Waves 14 15

ageclass	mean	median	sd	lower	upper
[0,6)	0.433	0.425	0.412	0.000	1.167
[6,12)	0.436	0.388	0.398	0.000	1.405
[12,18)	0.879	0.897	0.267	0.161	1.322
[18,30)	0.246	0.153	0.311	0.000	1.093
[30,40)	0.479	0.212	0.661	0.000	2.002
[40,50)	2.615	2.145	1.843	0.488	6.913
[50,60)	0.133	0.000	0.341	0.000	1.184
[60,70)	0.495	0.400	0.432	0.000	1.497
[70,80)	1.725	1.432	1.105	0.303	4.271
80+	0.674	0.501	0.764	0.000	2.751

(c) Waves 16 17

ageclass	mean	median	sd	lower	upper
[0,6)	0.189	0.146	0.177	0.000	0.556
[6,12)	0.480	0.484	0.147	0.196	0.740
[12,18)	1.079	1.090	0.142	0.806	1.342
[18,30)	0.333	0.320	0.230	0.000	0.806
[30,40)	0.181	0.139	0.178	0.000	0.548
[40,50)	1.536	1.418	0.703	0.520	3.044
[50,60)	0.212	0.143	0.235	0.000	0.820
[60,70)	0.120	0.083	0.127	0.000	0.384
[70,80)	0.469	0.432	0.328	0.001	1.227
80+	0.089	0.021	0.115	0.000	0.354

(d) Waves 18 19

ageclass	mean	median	sd	lower	upper
[0,6)	0.255	0.038	0.367	0.000	1.115
[6,12)	0.646	0.662	0.376	0.018	1.484
[12,18)	0.847	0.867	0.335	0.121	1.433
[18,30)	0.682	0.607	0.546	0.000	2.017
[30,40)	0.790	0.709	0.631	0.000	2.279
[40,50)	4.716	4.233	2.591	2.017	10.348
[50,60)	0.112	0.002	0.262	0.000	0.999
[60,70)	0.049	0.003	0.149	0.000	0.461
[70,80)	0.072	0.004	0.196	0.000	0.585
80+	0.111	0.004	0.248	0.000	0.851

(e) Waves 20 21

ageclass	mean	median	sd	lower	upper
[0,6)	0.794	0.823	0.321	0.013	1.352
[6,12)	0.263	0.208	0.247	0.000	0.799
[12,18)	0.692	0.689	0.189	0.339	1.044
[18,30)	0.671	0.644	0.274	0.210	1.346
[30,40)	0.039	0.000	0.109	0.000	0.323
[40,50)	0.496	0.436	0.404	0.000	1.327
[50,60)	0.399	0.391	0.270	0.000	0.918
[60,70)	0.014	0.000	0.042	0.000	0.123
[70,80)	0.104	0.053	0.144	0.000	0.507
80+	0.016	0.004	0.033	0.000	0.105

(f) Waves 22 23

Table I. Relative q-infectiousness with time evolution using assumption on susceptibility $(0.4, 0.39, 0.38, 0.79, 0.86, 0.8, 0.82, 0.88, 0.74, 0.74)$ corresponding to Figure P.