Age-specific transmission dynamics of SARS-CoV-2 during the first two years of the pandemic

Otilia Boldea, PhD[1], Amir Alipoor, MSc[1], Sen Pei, PhD[2], Jeffrey Shaman, PhD[2,3], and Ganna Rozhnova, PhD[*4,5,6]

[1]Department of Econometrics and OR and CentER, Tilburg School of Economics and Management, Tilburg University, Tilburg, The Netherlands

[2]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, USA

[3]Columbia Climate School, Columbia University, New York, USA

[4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[5]BioISI—Biosystems & Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

[6]Center for Complex Systems Studies (CCSS), Utrecht University, Utrecht, The Netherlands

December 6, 2022

*Corresponding author:

Dr. Otilia Boldea

Department of Econometrics and OR, and CentER

Tilburg School of Economics and Management, Tilburg University

P.O. Box 90153, Tilburg, The Netherlands

E-mail: o.boldea@tilburguniversity.edu, phone: +31 134663219

Dr. Ganna Rozhnova

Julius Center for Health Sciences and Primary Care

University Medical Center Utrecht

P.O. Box 85500 Utrecht, The Netherlands

Email: g.rozhnova@umcutrecht.nl, phone: +31 631117965

Abstract

During its first two years, the SARS-CoV-2 pandemic manifested as multiple waves shaped by complex interactions between variants of concern, non-pharmaceutical interventions, and the immunological landscape of the population. Understanding how the age-specific epidemiology of SARS-CoV-2 has evolved throughout the pandemic is crucial for informing policy decisions. We developed an inference-based modelling approach to reconstruct the burden of true infections and hospital admissions in children, adolescents and adults over the seven waves of four variants (wild-type, Alpha, Delta, Omicron BA.1) during the first two years of the pandemic, using the Netherlands as the motivating example. We find that reported cases are a considerable underestimate and a generally poor predictor of true infection burden, especially because case reporting differs by age. The contribution of children and adolescents to total infection and hospitalization burden increased with successive variants and was largest during the Omicron BA.1 period. Before the Delta period, almost all infections were primary infections occurring in naive individuals. During the Delta and Omicron BA.1 periods, primary infections were common in children but relatively rare in adults who experienced either re-infections or breakthrough infections. Our approach can be used to understand age-specific epidemiology through successive waves in other countries where random community surveys uncovering true SARS-CoV-2 dynamics are absent but basic surveillance and statistics data are available.

Introduction

During the pandemic, the dynamics of SARS-CoV-2 demonstrated a complex spatio-temporal pattern with multiple waves [1] and pronounced differences in the age-specific burden of confirmed cases and hospitalizations [2, 3]. A notable example is the age distribution of reported cases with much lower number of cases among younger individuals in European countries, including the Netherlands, reported in the first wave of the wild-type variant than in the spike of infections caused by the Omicron BA.1 variant of concern (VoC) [4].

Understanding how the age-specific epidemiology of SARS-CoV-2 has changed during the pandemic [5] is crucial for informing public health policy. For instance, information about how infection burden varies by age and time and the contribution of different age groups to transmission [6, 7] may inform non-pharmaceutical interventions [8–12] like school- and non-school-based measures [13], while understanding of the age-specific hospitalization burden [14] underpins prioritization of vaccination [15, 16]. Some of this information can be provided by surveillance and serological surveys. However, full reconstruction of the age-specific burden of infections and hospitalizations in a country is complicated by several factors. Firstly, under-reporting of cases is age-specific and time-varying due to peculiarities of surveillance systems and testing policies, which varied across ages and time as the pandemic progressed. This fact coupled with the evidence of asymptomatic infection [17] undermines the ability of surveillance to capture the true burden of infections among different age groups. Secondly, nationally representative serological surveys provide information on which subpopulations carry antibodies to SARS-CoV-2 and thus could help to characterize prior infection burden [18,19]. However, due to the waning of immunity after vaccination and infection,

the immunological landscape of the population has become increasingly complex [20]. Implementing representative serosurveys that estimate population immunity among different age groups characterized by varying numbers of prior infections before and after vaccination is difficult, costly and time-consuming. Thirdly, additional factors including VoCs [21], non-pharmaceutical interventions [8–12], changes in population immunity after vaccination or infection, and seasonality in transmission [22–24] complicate estimation of the age-specific burden.

Modelling studies have provided important insights into temporal changes in the epidemiology of SARS-CoV-2 in specific countries [13, 19, 23, 25–29]. Some of these were hypothesis-generating and not rigorously validated against all available evidence [23, 27, 28]. Other studies did not reconstruct age-specific epidemiology [26] or were limited to specific periods of the pandemic such as the first wave [13, 30] or periods of dominance for the Alpha [19, 25] and Delta variants [29, 31]. As data accumulate, formal evaluations based on mathematical models calibrated to different types of observational data [13, 26, 29] are crucial for reconstructing the burden of infections and hospitalizations over long periods of time.

Here we reconstruct the epidemiology of SARS-CoV-2 in the Netherlands over the first 23 months of the pandemic, a period including wild-type, Alpha, Delta and Omicron BA.1 waves, using an age- and regionally stratified transmission model fitted to various data sources (see Supplementary material). Our fitting approach is a Bayesian evidence synthesis [13, 25, 32–34] based on an ensemble adjustment Kalman filter [26, 35, 36] combined with surveillance and national statistics data typically available from individual countries (hospital admissions, serological surveys, PCR testing data, genetic VoC data, vaccination coverage data, social contact matrices, demographic data, regional train and Google mobility data) with the Netherlands used as the motivational example.

To account for control measures targeted at elementary and secondary schoolchildren versus the rest of the population [13] and for age-dependent transmission effects [3, 14, 17, 37–40], the population is stratified into young children (0 to 9 years old), adolescents (10 to 19 years old) and adults (above 19 years old) in twelve Dutch provinces. The model structure is rooted in current knowledge of SARS-CoV-2, which suggests waning of immunity after infection or vaccination [20], potential changes in susceptibility, infectivity and severity of re-infections and breakthrough infections [27, 28], and seasonality in transmission [22–24]. The regional stratification is augmented with real-world mobility across the provinces under the assumption that infected individuals with undocumented infection may travel to and infect susceptible individuals in other provinces whereas infected individuals with confirmed infection stay in their province of origin. Previous studies indicate that epidemic models with transmission dynamics coupled across locations can improve the identifiability of epidemiological parameters [26, 41].

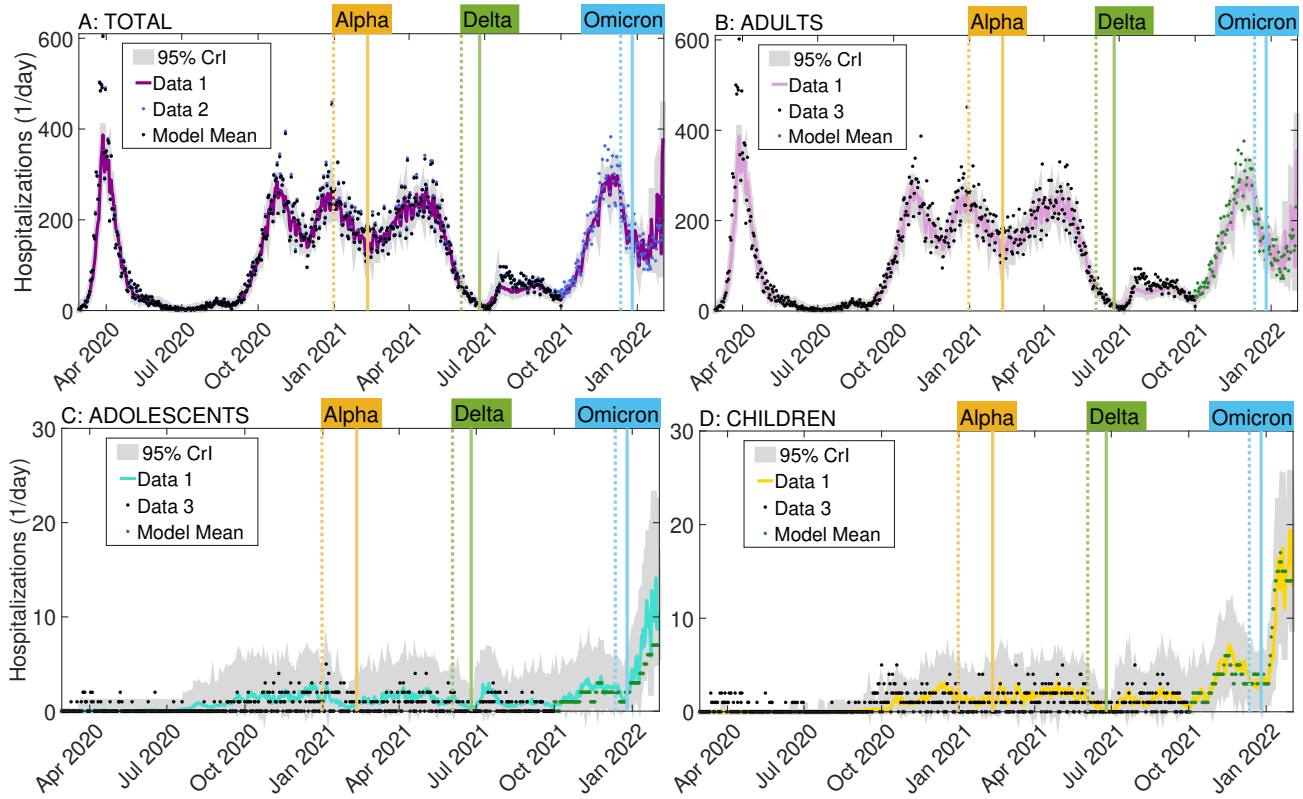


Figure 1. Estimated hospital admissions. Total (**A**) and age-specific national daily hospital admissions in adults (**B**), adolescents (**C**) and children (**D**). The colored lines represent the estimated posterior means for adults (pink), adolescents (turquoise), children (yellow), and all ages (purple). The gray-shaded regions correspond to posterior 95% credible intervals defined as the 2.5% and 97.5% quantiles from 300 posterior ensemble values. The dots are daily hospital admission data used for fitting the model. Data 1 refers to the National Intensive Care Evaluation data obtained from the National Institute for Public Health and the Environment (RIVM); Data 2 and 3 refer to the RIVM Dashboard data (see Table 1 and Supplementary Material, Section 1). The dashed and solid vertical lines indicate when each VoC corresponded to 5% and 50% of samples in the genetic variant data, respectively.

Results

Time-dependent burden of hospital admissions

Our parsimonious model allows for robust estimation of key time-varying epidemiological parameters compatible with national and regional epidemiology of SARS-CoV-2 in pre- and post-vaccination periods (Supplementary Tables A1—A6 and Figures A1—A4; see also Section "Parameter identifiability and sensitivity analyses"). The model reproduces well national and regional hospital admissions in young children, adolescents and adults during all waves that occurred from the first official case on February 27, 2020 until January 31, 2022 (Figure 1, Supplementary Figures A5—A7). At the national level, the total estimated mean number of hospital admissions is 84,674 (95% CrI 82,729—86,151) (Supplementary Table A7) spread over seven waves in the wild-type, Alpha, Delta and Omicron BA.1 periods, with the last Omicron wave starting but not yet ending at the end of the study period, January 31,

2022 (Figure 1 **A**). Of these, two and four waves occurred before and, respectively, after the start of vaccination, and one wave peaked in January 2021 around the onset of the vaccination program. As expected, the pattern of hospital admissions in adults largely mirrors that of total national admissions due to much higher probability of clinical disease and hospitalization in this subpopulation compared to adolescents and children [3,13,14,25] (Figure 1 **B**). In our analyses, we quantified the burden per period of VoCs, defined as the time when a VoC reached a frequency of 5% in the genetic variant data described in Table 1, until another VoC reaches the same frequency. The burden of hospital admissions in adults progressively decreased over the periods of successive VoCs, in line with the expansion of the primary vaccination series and of the first booster campaign (Supplementary Table A7). The estimated cumulative mean number of hospital admissions for adults was thus largest for the wild-type (30,099, 95% CrI 29,126—30,784) and smallest for the incomplete Omicron BA.1 period until January 31, 2022 (8,352, 95% CrI 7,895—8,816) (Supplementary Table A7). The picture is different for adolescents and children for whom the hospitalization burden stayed lower than 5 hospital admissions per day for either group and did not demonstrate a pronounced pattern until October 2021 (Figure 1 **C** and **D**). Unlike in adults, the estimated cumulative mean number of hospital admissions in these two subpopulations was largest in the Omicron BA.1 and Delta periods (Supplementary Table A7; for adolescents, 292, 95% CrI 239—346 and 266, 95% CrI 173—453; for children, 433, 95% CrI 333—586 and 388, 95% CrI 258—693).

Time-dependent burden of confirmed cases and seroprevalence

The model further reproduces the age-specific seroprevalence and confirmed cases nationally (Figure 2 **B**-**D**) and regionally (Figure 2, Supplementary Figures A8—A16). The seroprevalence captures the fraction of the population who have antibodies due to (any) infection. The level of seroprevalence is thus determined by how fast antibodies decay and how fast new infections happen. Confirmed cases refer to the part of infections captured by PCR testing surveillance. The estimated national seroprevalence steadily increased in all age groups reaching 84% (95% CrI 82%—86%) in adults, 90% (95% CrI 87%—93%) in adolescents and 64% (95% CrI 60%—68%) in children by 31 January 2022 (Figure 2 **A** and Supplementary Table A8). In contrast with the magnitude of the first wave of hospital admissions in spring 2020, the case reporting in this period was overall low, even more so among children and adolescents (insets Figure 2 **B**, **C** and **D**). Throughout most of the study period that covers the next five waves, less than 1 case per 1,000 adults was reported daily, but this sharply increased during the Omicron BA.1 wave to 4 cases per 1,000 adults by the end of January 2022 (Figure 2 **B**). The pattern of reported cases was similar in adolescents with less than 1 case per 1,000 adolescents daily during the first five waves; however, substantially more Omicron BA.1 cases were reported in this age group compared to adults during the sixth and seventh waves, skyrocketing from 1 to 12 daily cases per 1,000 adolescents in January 2022 (Figure 2 **C**). More cases were documented in children than in adults in the Omicron BA.1 wave too, i.e., about 6 versus 4 cases per 1,000 individuals per day at the peak, while case reporting in children was much lower than in adults during the rest of

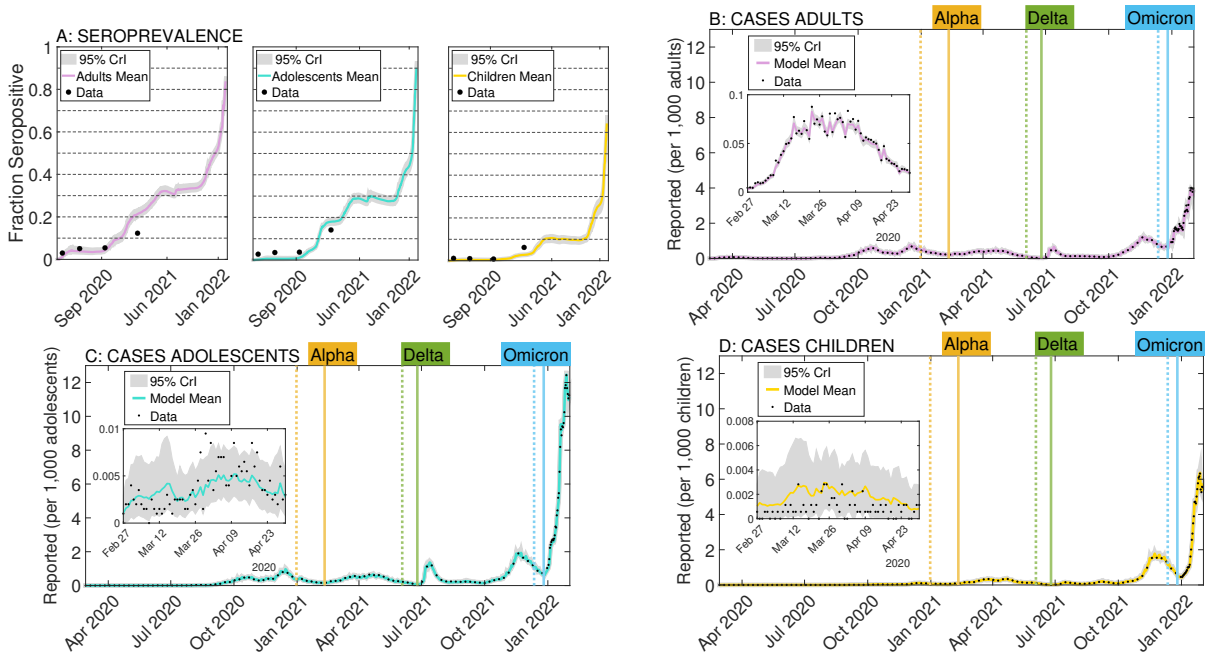


Figure 2. Estimated seroprevalence and confirmed cases. Age-specific national seroprevalence due to infection (**A**) and national confirmed daily cases in adults (**B**), adolescents (**C**) and children (**D**). Inset: the first wave characterized by low case reporting. The colored lines represent the estimated means for adults (pink), adolescents (turquoise) and children (yellow). The gray-shaded regions correspond to posterior 95% credible intervals defined as the 2.5% and 97.5% quantiles from 300 posterior ensemble values. The black dots are seroprevalence (**A**) and daily confirmed cases (**B**, **C**, **D**) data used for fitting the model (see Table 1 and Supplementary Material, Section 1). The dashed and solid vertical lines indicate when each VoC corresponded to 5% and 50% of samples in the data, respectively. For comparison, the scale of the y-axis is the same in **B**, **C** and **D**.

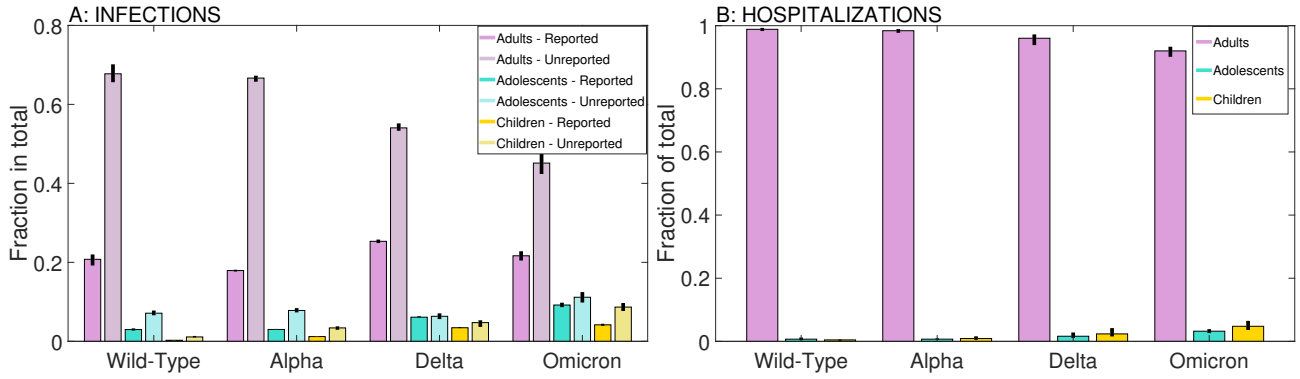


Figure 3. Age distribution of estimated total infections and hospital admissions by VoC. Age-specific fraction of confirmed and unconfirmed cases in the total national infections (**A**) and hospital admissions (**B**) during each VoC period. The bars and the black lines show the means and 95% credible intervals obtained from 300 posterior ensemble values.

the study period (Figure 2 **D**). The observed increase in reported cases in adolescents and children compared to adults may indicate a higher burden of infections in these subpopulations during the Omicron BA.1 period and a change in the age-specific distribution of total national cases which we verify further below.

Age distribution of total infections and hospital admissions per VoC

Our model estimates temporal changes in the age-specific contact rates as a result of implemented control measures (Supplementary Material, Section 2.4). These changes combined with varying mobility patterns of reported and unreported cases across provinces allow estimation of the age-specific distribution of reported and unreported cases in the total national infections during each VoC period (Figure 3 **A** and Supplementary Tables A9—A10). The fraction of total infections in children among total estimated national infections (light and dark yellow bars) steadily increased from 1.4% in the wild-type period to 12.9% in the Omicron BA.1 period , of which about 66% were unreported in total. The same increasing contribution to total infections per successive VoC periods is estimated for adolescents (light and dark turquoise bars) who comprised between 10.1% and 20.4% of all national infections for the wild-type and Omicron BA.1 periods, respectively. The majority of all cases in adolescents were not reported to surveillance, with the estimated fraction of unreported infections 60% in total for all VoCs. The contribution of adults to total infections (light and dark pink bars) decreased from 88.5% to 66.8% through the study period while the fraction of reported adult cases increased from approximately 23.5% during the wild-type period to 32.5% during the Omicron BA.1 period. In terms of the age-specific contribution to total hospital admissions per VoC, the adult population suffered the largest burden during all VoC periods with a minor fraction of hospital admissions attributed to children and adolescents during the circulation of Delta and Omicron BA.1 (Figure 3 **B** and Supplementary Table A11).

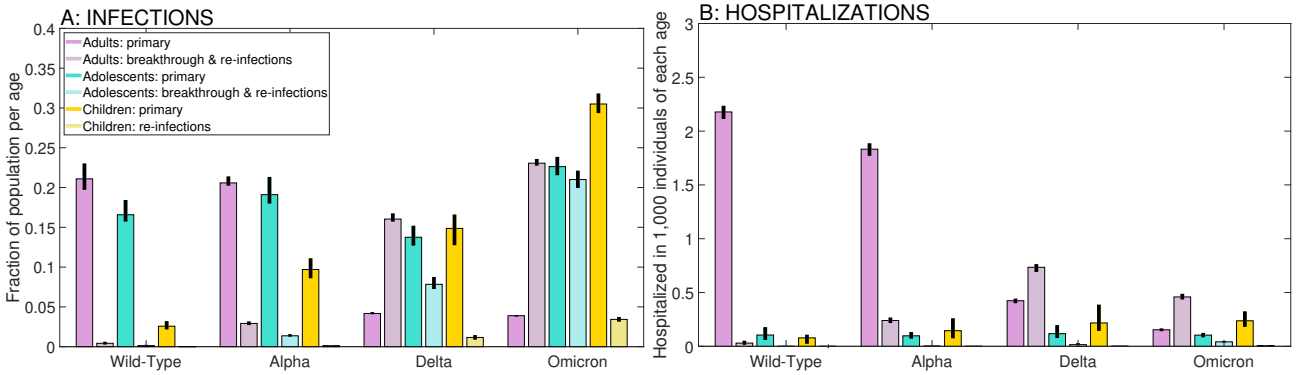


Figure 4. Burden stratification by immune status and VoC. (**A**) Fraction of primary and breakthrough infections or re-infections in the total population of each age group. (**B**) Hospital admissions after primary and breakthrough infections or re-infections per 1,000 individuals in each age group. The bars and the black lines show the means and 95% credible intervals obtained from 300 posterior ensemble values.

Age-specific burden stratification by immune status and VoC

We further analyzed the age-specific burden of total infections and hospital admissions stratified by immune status and period of VoC (Figure 4 and Supplementary Tables A12—A13). For this, we distinguished infections in fully susceptible (naive) individuals (primary infections) from infections in individuals who were vaccinated (breakthrough infections) or lost immunity after primary infection (re-infections) (Figure 4 **A**). Hospital admission burden was stratified into hospital admissions after primary infection versus after re-infection or breakthrough infection despite vaccination (Figure 4 **B**). An estimated 21.1% (95% CrI 19.7%—23.0%) and 20.6% (95% CrI 20.2%—21.4%) of adults (pink bars) had primary infection during the wild-type and Alpha periods, respectively, while re-infections or breakthrough infections were experienced in only 3.3% of adults for both periods combined. This pattern reversed during the Delta and Omicron periods as re-infections or breakthrough infections occurred in 16.0% (95% CrI 15.7%—16.7%) and 23.1% (95% CrI 22.7%—23.6%) of adults, respectively, while primary infections were experienced by only about 4% of adults during each period. Similar dynamics are estimated for the burden of hospital admissions in adults; namely, there were more hospital admissions due to Delta and Omicron BA.1 breakthrough infections/re-infections than due to primary infections by these variants (Figure 4 **B**). We stress that larger hospitalization burden in adults associated with breakthrough infections/re-infections than with primary infections during the Delta and Omicron BA.1 periods is a consequence of high vaccination coverage in adults while the hospitalization probability after vaccination or prior infection is very low compared to that in adults without any immunity. In the extreme case of a fully vaccinated adult population, all hospitalization burden would be attributed to breakthrough infections, and this would be much smaller than the burden the totally naive population experienced at the beginning of the pandemic. The burden of primary infections in adolescents (turquoise bars) was only slightly lower than that of adults for the wild-type (16.6% of the age group, 95% CrI 15.7%—18.4%) and Alpha (19.1% of the age group, 95% CrI 18.0%—21.3%) VoC. However, unlike for adults, primary infections in adolescents were more common than breakthrough infections or re-infections during the Delta and Omicron BA.1 VoCs, too (13.7% and 22.6% of the age group, respectively), owing possibly to the delayed vaccination schedule for adolescents. The burden in adolescents was especially high during the Omicron BA.1 wave when an estimated 43.6% of all adolescents were infected. In part because young children were the last to start vaccination, the burden in this group increased with time and was mainly due to primary infections which, in turn, led to a slight increase of hospitalizations (yellow bars, Figure 4 **A** and **B**).

Parameter identifiability and sensitivity analyses

The posterior distribution of most parameters tighten after the first wave (Supplementary Material, Section 6.1), indicating that the data are informative in recovering model parameters. To further verify system identifiability in the first wave, we generated one synthetic outbreak and verified that the generated infection and hospitalization data follow similar trends to the original data. To assess parameter identifiability in the first wave, we considered

higher, lower and time-varying case detection rates. For each parameter combination, 100 synthetic outbreaks were generated, each of which was used as data to re-estimate the model parameters. Across the three model configurations, the "true" parameters were either within the range of the posterior mean densities, or the differences were not too large (Supplementary Material, Section 6.2).

Sensitivity analyses were conducted for the model parameters that were not estimated but rather calibrated (see Methods and Supplementary Material, Section 7). The results were validated against Dutch seroprevalence estimates, some of which were not used in the estimation. The selected values for calibrated parameters yielded results that are closer to the Dutch seroprevalence estimates than alternative calibrations.

Discussion

We developed an inference-based modelling approach to reconstruct how the age-specific epidemiology of SARS-CoV-2 changed through time and was shaped by the complex interaction between VoCs, non-pharmaceutical interventions, and the immunological landscape of the population. Our study provides several insights into changes in the age-specific burden of infections and hospitalizations over seven waves with different VoCs during the first two years of the pandemic in the Netherlands. Firstly, cases reported to surveillance were a considerable underestimate of total infections, as the majority of infections in all age groups remained unidentified despite overall improvement of case reporting through time. Reported cases are a generally poor predictor of true incidence of infection, especially because reporting differs substantially by age. Secondly, the contribution of children and adolescents to total infection and hospitalization burden increased with successive VoCs and was largest during the Omicron BA.1 period. This increase demonstrates that despite being recognized as less virulent than previous VoCs [42], variants such as Omicron BA.1 can impose substantial burdens on adolescents and children and additional pressure on healthcare systems. Thirdly, we observed a shift in the pattern of infections and hospitalizations by immune status and age. Before Delta, almost all infections were primary infections occurring in naive individuals. However, during the Delta and Omicron BA.1 periods primary infections were common in children who were infected less frequently early on and for whom vaccination roll-out lagged behind that of other age groups, but were relatively rare in adults who experienced either re-infections or breakthrough infections.

A similar shift in the dynamics was reported in a study based on identified case data from the UK [43]. However, unlike in [43] where re-infections dramatically increased during the Omicron wave, in our study the shift in primary infections versus re-infections and breakthrough infections is estimated to have begun during the Delta period. The discrepancy between the two studies could be attributed to the population stratification, used in this study. For computational efficiency, our model distinguishes fully naive individuals from those with immunity induced by either vaccine or natural infection, therefore re-infections and breakthrough infections are grouped.

To our knowledge, this is the first and most comprehensive application of a computationally efficient inference-based

modeling approach that provides estimates of infection and hospitalization burden by age, region, VoC and immune status. The system is applied over seven waves during the first two pandemic years in the Netherlands [13,30,44–49]. While we focus on the Dutch case, our framework can be applied to other countries given the common surveillance and national statistics data typically available for most countries. Compared to other studies [23, 26, 27, 50], we separately model contacts in elementary schools, secondary schools and the rest of the population to enable to more reliably disentanglement of the roles of children, adolescents and adults in transmission. We estimate contact changes due to non-pharmaceutical interventions outside the school environment based on proxy mobility measures [51] — Google and train mobility at transit stations. From this proxy, we observe that most measures are taken up by the population ahead of official implementation. We also estimate the speed of behavioral changes and the number of age-specific contacts after each intervention [13,25,52]. This is in contrast to other inference-based studies (e.g. [29]) that assume that age-related mixing patterns remain constant over time.

Our model is fitted to regional reported cases, regional hospital admissions, and national seroprevalence. We find that the first two data sources, combined with modelling movement of unreported infections across regions using mobility data, are crucial for identifying unreported infections with our inference method. The national seroprevalence data play a secondary role in reconstructing transmission dynamics due to the low frequency of serosurveys (i.e., four surveys during 23 months) and the fact that our underlying transmission model is stochastic, unlike in other studies [13,25]. Our approach could therefore be extended to countries where seroprevalence estimates are not available at a high enough frequency.

Our model has limitations. Firstly, not testing children and adolescents in the early stages of the pandemic makes the identification of their age-specific parameters (i.e., susceptibility) weaker in that initial period. Later on, as testing capacity was expanded to younger individuals, we observe good identification of all age-specific parameters. Secondly, we did not model older age categories separately. It is technically straightforward to stratify the adult population into smaller age categories relevant for estimating the burden in older ages. However, older individuals travel less which means that our metapopulation model with mobility may not capture their case detection rates, and other data might be needed to estimate their age-specific parameters. Thirdly, our choice of stratification into primary versus breakthrough infections and re-infections is justified for the time period when a large proportion of the population did not yet have any immunity to SARS-CoV-2. As SARS-CoV-2 transitions from pandemicity to endemicity, primary infections will be experienced only by very young children born into the population [28, 53]. For later periods, our model could be extended to differentiate between several immunity classes such as individuals with primary vaccination series and various boosters, prior infections, and hybrid immunity [54].

In conclusion, we developed an inference-based transmission model that estimates how the age-specific epidemiology of SARS-CoV-2 changes over time. This approach is relevant for countries in which random community surveys uncovering true SARS-CoV-2 dynamics are absent but basic surveillance and statistics data are available. The findings of our study on the burden of infections and hospitalizations in children, adolescents and adults are

important for informing public health policy on non-pharmaceutical interventions and vaccination.

Methods

Overview

The transmission model was calibrated using surveillance and national statistics data (PCR testing data, hospital admissions, serological surveys, demographic data, regional train and Google mobility data, vaccination coverage data, genetic VoC data, social contact matrices) for the Netherlands. Parameter estimates were obtained from the model fit to (i) age- and province-stratified SARS-CoV-2 case notification data in the period from February 27, 2020 until January 31, 2022; (ii) age- and province-stratified COVID-19 hospital admission data in the same period; (iii) cross-sectional age-stratified national seroprevalence data from four serosurveys assessed on April 3, June 4 and September 20, 2020, and February 11, 2021 (median dates). Additional data for the model input were: (iv) population by age and province on January 1, 2020; (v) daily commuters across twelve provinces computed from the Dutch national train data in the period from February 1, 2020 until September 30, 2022, and Google mobility data from February 5, 2020 until January 31, 2022; (vi) daily full vaccinations per province and age category in the period from January 31, 2021 until January 31, 2022; (vii) weekly boosters administered in the period from November 21, 2021 until January 31, 2022; (viii) weekly genetic VoC data in the period from December 1, 2020 until January 31, 2022; (ix) school and non-school contact matrices from three surveys: before the pandemic, in April 2020 and June 2020.

Data

Table 1 gives an overview of the data, notation and sources. More details on how each dataset was constructed to be used in the model fitting are given in the Supplementary Material, Section 1.

Transmission model

We developed a stochastic compartmental metapopulation model describing SARS-CoV-2 transmission in the population of the Netherlands stratified by province, disease status and age. The schematic of the model diagram outlining the disease and mobility dynamics, as well as an overview of the main model parameters used in the diagram are shown in Figure 5.

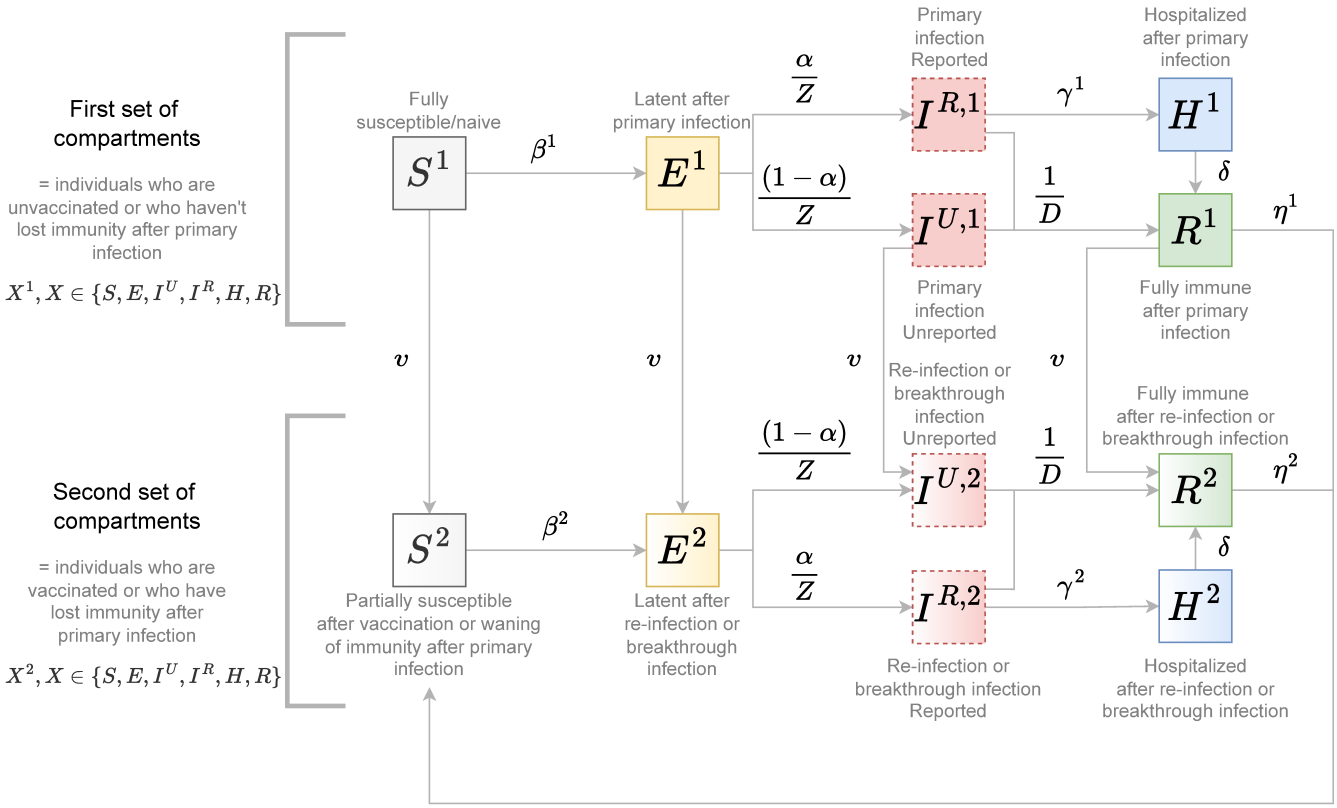
Disease dynamics The fully naive susceptible individuals (S^1) in each age group can become latently infected (E^1) with the age-specific force of infection β^1. After an average latent period Z, the latently infected individuals become infectious (primary infection, I^1). Out of the total daily new infectious cases, an age-specific fraction α is reported to surveillance ($I^{R,1}$), and the $(1-\alpha)$ fraction is unreported ($I^{U,1}$). Both reported and unreported infectious individuals may recover without hospitalization (R^1) after an average infectious period D, but only

Table 1. Overview of the data used in the model fitting.

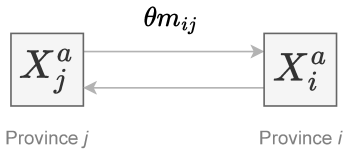
Data	Notation*	Source
Population of age group k in province i	N_{ik}	CBS, January 1, 2020 https://opendata.cbs.nl/statline/portal.html?_la=en&_catalog=CBS&tableId=37259eng&_theme=1135
Daily reported cases	$Inew_{ik}^{obs}(t)$	RIVM Dashboard https://data.rivm.nl/covid-19/COVID-19_casus_landelijk.csv
Daily hospital admissions		
Dataset 1	$Hnew_{ik,(1)}^{obs}(t)$	RIVM data, from February 27, 2020 until September 30, 2021
Dataset 2	$Hnew_{i,(2)}^{obs}(t)$	RIVM Dashboard, from October 1, 2021 onward https://data.rivm.nl/covid-19/COVID-19_ziekenhuisopnames.csv https://data.rivm.nl/covid-19/COVID-19_ziekenhuisopnames_tm_03102021.csv
Dataset 3	$Hnew_{k,(3)}^{obs}(t)$	RIVM Dashboard, from October 1, 2021 onward https://data.rivm.nl/covid-19/COVID-19_ziekenhuis_ic_opnames_per_leeftijdsgroep.csv https://data.rivm.nl/covid-19/COVID-19_ziekenhuis_ic_opnames_per_leeftijdsgroep_tm_03102021.csv
Seroprevalence data		obtained from RIVM PIENTER Corona Study
Serosurvey rounds	$h = 1, 2, 3, 4$	https://www.rivm.nl/en/pienter-corona-study/results
Median inclusion time	t_h^{ser}	
Fraction seropositive	ser_{hk}	
Sample size	n_{hk}	
Train mobility between provinces i and j	$M_{ij}(t)$	computed from National Railway Company (NS) data, from February 27, 2020 until September 30, 2021; extrapolated until January 31, 2022 using Google mobility data
Fraction commuters from province j to i	w_{ij}	CBS, year 2019 https://www.cbs.nl/nl-nl/cijfers/detail/83628NED
Google mobility	y_g	https://www.google.com/covid19/mobility/ from February 7, 2020 onward
Daily vaccinations	$V_{ik}(t)$	constructed from RIVM Dashboard https://data.rivm.nl/covid-19/COVID-19_vaccinatiegraad_per_gemeente_per_week_leeftijd.csv and RIVM data https://www.rivm.nl/en/covid-19-vaccination/archive-covid-19-vaccination-figures-2021
Booster transition function	$g_{B,k}(t), k = 1, 2$	fitted to RIVM data https://www.rivm.nl/en/covid-19-vaccination/archive-covid-19-vaccination-figures-2022
Variants transition function	$g_\ell(t), \ell \in \{\alpha, \delta, o\}$	fitted to RIVM Dashboard data https://data.rivm.nl/covid-19/COVID-19_varianten.csv
Contact matrices		computed from [38, 55]
Elementary school contacts	$c_{ikk^*,ES}(t)$	and additional information on school closures and vacations
Secondary school contacts	$c_{ikk^*,SS}(t)$	and additional information on school closures and vacations
All-setting contacts	$c_{kk^*}^{(r)}$, $r = 1, 2, 3$	calculated from [55], r refers to survey rounds
Non-school contacts	$c_{kk^*,NS}^{(r)}$, $r = 1, 2, 3$	calculated from [38, 55], r refers to survey rounds

*The indices $i, j = 1, \ldots, 12$ denote the province of the Netherlands. The indices k, k^* denote the age group, namely $k, k^* = 1$ — adults (> 19 years old), $k, k^* = 2$ — adolescents ($10 - 19$ years old), and $k, k^* = 3$ — children ($0 - 9$ years old). The index $h = 1, 2, 3, 4$ denotes the serosurvey round. The index $l = \alpha, \delta, o$ denotes Alpha, Delta, and Omicron VoCs.

A Disease dynamics



B Mobility dynamics



$$X \in \{S, E, I^U, R\},\ a \in \{1, 2\},\ j \neq i$$

C Overview of the parameters

Age-specific parameters

β^1	- force of infection for fully susceptible
β^2	- force of infection for partially susceptible
α	- case detection rate
γ^1	- hospitalization rate after primary infection
γ^2	- hospitalization rate after re-infection or breakthrough infection
v	- vaccination rate
$1/\delta$	- hospitalization period
θ	- mobility reporting error
m_{ij}	- mobility between province j and province i

Constant parameters

Z	- latent period
D	- infectious period
$1/\eta^1$	- duration of immunity after primary infection
$1/\eta^2$	- duration of immunity after re-infection or breakthrough infection

Figure 5. Schematic of the metapopulation transmission model. (**A**) Disease dynamics. (**B**) Mobility dynamics. (**C**) Overview of the parameters used in the model diagram. For simplicity of presentation, the age-specific indices are not shown in the schematic.

reported infectious individuals may be hospitalized (H^1) with the age-specific rate γ^1. The hospitalized individuals are discharged after an average age-specific hospitalization period $1/\delta$. After a primary infection, individuals lose immunity and become partially susceptible (S^2) with the rate η^1. The disease progression for the partially susceptible individuals is similar to that for the fully susceptible individuals. However, they get reinfected with the age-specific force of infection β^2, individuals reported with re-infection are hospitalized with the age-specific rate γ^2, and the immunity after re-infection is lost with the rate η^2. Upon losing immunity after re-infection, the individuals return to the same partially susceptible compartment (S^2). Vaccination of individuals with the age-specific rate ν occurs in all disease stages except for those with primary reported infection or hospitalization. We used a simplified approach where susceptible individuals after vaccination and individuals recovered after primary infection whose immunity has waned are grouped together in one class of partially susceptible individuals (S^2). Therefore, re-infections and breakthrough infections are grouped too. Vaccination or prior infection have three effects: (i) lower susceptibility to re-infection and breakthrough infection that affects β^2, (ii) lower infectivity of re-infection and breakthrough infection (not shown in the diagram), and (iii) lower hospitalization rate after re-infection and breakthrough infection γ^2. The model assumes a constant population size through the study period.

Mobility dynamics The model assumes that unreported infected travel to and infect other individuals in other regions, and that susceptible individuals can travel and get exposed in another region, while reported infected do not travel. Because of heterogeneity in population size and mobility across regions, this assumption equips the model with additional dynamics for unreported infected compared to reported infected. The identification of unreported cases is possible as long as the case detection rate features common parameters across regions, and there are sufficient observations available across regions or time [26, 56, 57]. These two conditions are met in our study. Firstly, the case detection rate is assumed common across regions due to the ability to test at any location available across the country. Secondly, we estimate the model at a daily frequency for all twelve Dutch provinces.

A full description of the model and the model equations are reported in the Supplementary Material, Section 2.

Parameter inference

The model is fitted to infection, hospitalization and seroprevalence data using the ensemble adjustment Kalman filter [35]. This method allows inference on a high-dimensional system of observed and unobserved variables and parameters in a computationally efficient way. First, a large number of ensembles are drawn from priors on all parameters and state variables. The latter then evolve according to a stochastic version of the metapopulation model (Figure 5). In the updates, each ensemble member is individually propagated forward by means of a closed form approximation akin to the normal distribution but with an additional bias correction. The updates are done sequentially based on the data in each province and its neighbours, allowing for a large number of state variables and parameters to be updated. Further details on the estimation and the algorithm can be found in the Supplementary Material, Sections 3.3—3.4.

A small number of parameters that could not be identified were calibrated. Their calibrations are described and motivated in the Supplementary Material, Section 3.1. The rest of the parameters were estimated. Their choice of priors is detailed and motivated in the Supplementary Material, Section 3.2. The parameter posteriors and their time-evolution over variant periods are discussed in the Supplementary Material, Section 4.

Model outcomes

The exact calculations behind Figures 1—4 are explained in the Supplementary Material, Sections 5.1—5.4. The data plotted in Figures 3 and 4 can be found in Supplementary Tables A9 and A11—A13, respectively.

Details on the parameter identifiability and sensitivity analyses

To verify system identifiability in the first wave, we fixed the parameters at their estimated posteriors in the original sample (observations from February 27, 2020 until March 30, 2020, see Supplementary Table A14), generated one synthetic outbreak and re-estimated the model parameters on the synthetic data. Besides showing that the synthetic data are similar to the original data and are fitted well by our method, we also verified that the fitted and model seroprevalence match (Supplementary Material, Section 6.1, Table A15 and Figures A17—A20). To assess parameter identifiability in the first wave, we considered three parameter configurations: (i) same as described above, with case detection rates equal to the original sample posterior means; (ii) larger case detection rates; (iii) case detection rates as in (i) until March 30, 2020, and as in (ii) until April 30, 2020. For each parameter combination, 100 synthetic outbreaks were generated, each of which was used as data to re-estimate the model parameters. Across the three model configurations, the "true" parameters were either within the range of the posterior mean densities or the differences were not too large (Supplementary Material, Section 6.2, Figures A21—A23). In case (iii), the filter also approximated well and instantaneously the increase in case detection rates, even though this increase was not modelled. This suggests that changes in testing capacity or recommendations, and other unmodelled parameter changes can be captured by our model inference technique even in periods of high uncertainty, when the data are less informative (Supplementary Figure A24).

In our sensitivity analyses, we re-estimated the model over the entire sample period: (i) without the serosurvey data; (ii) with lower and higher immunity waning rate after primary infection; (iii) with lower and higher immunity waning rates at the start of Omicron BA.1 period after both primary and breakthrough infection/re-infection; (iv) with higher transmissibility of Omicron BA.1; (v) without seasonality. We found that in most cases, except (iii), our calibrations lead to closer matches to available seroprevalence estimates independently of whether these were used in the model estimation or not (Supplementary Material, Section 7 and Tables A16—A19). In cases (iii)-(iv), the results are robust to whether immunity wanes on average after two, three or four months for the Omicron BA.1 variant, and a 40% or 60% higher transmissibility compared to the Delta variant.

References

[1] World Health Organization. Coronavirus (COVID-19) Dashboard; 2022. Available from: https://covid19.who.int/.

[2] Variation in the COVID-19 infection-fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. The Lancet. 2022;399(10334):1469–1488. doi:10.1016/S0140-6736(21)02867-1.

[3] O'Driscoll M, Dos Santos GR, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. Nature. 2021;590(7844):140–145. doi:https://doi.org/10.1038/s41586-020-2918-0.

[4] Coronavirus dashboard; 2020. Available from: https://coronadashboard.government.nl/.

[5] Koelle K, Martin MA, Antia R, Lopman B, Dean NE. The changing epidemiology of SARS-CoV-2. Science. 2022;375(6585):1116–1121. doi:10.1126/science.abm4915.

[6] Monod M, Blenkinsop A, Xi X, Hebert D, Bershan S, Tietze S, et al. Age groups that sustain resurging COVID-19 epidemics in the United States. Science. 2021;371(6536):eabe8372. doi:10.1126/science.abe8372.

[7] 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. doi:10.1038/s41591-020-0962-9.

[8] Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. Physics Reports. 2021;913:1–52. doi:https://doi.org/10.1016/j.physrep.2021.02.001.

[9] Liu Y, Morgenstern C, Kelly J, Lowe R, Group CCW, Jit M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. BMC medicine. 2021;19(1):40–40. doi:10.1186/s12916-020-01872-8.

[10] Sharma M, Mindermann S, Rogers-Smith C, Leech G, Snodin B, Ahuja J, et al. Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. Nature communications. 2021;12(1):5820–5820. doi:10.1038/s41467-021-26013-4.

[11] Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. Science. 2021;371(6531):eabd9338. doi:10.1126/science.abd9338.

[12] Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. The Lancet Infectious Diseases. 2021;21(2):193–202. doi:10.1016/S1473-3099(20)30785-4.

[13] Rozhnova G, van Dorp CH, Bruijning-Verhagen P, Bootsma M, van de Wijgert J, Bonten M, et al. Model-based evaluation of school and non-school-related measures to control the COVID-19 pandemic. Nature Communications. 2021;12(1):1614. doi:10.1038/s41467-021-21899-6.

[14] Flook M, Jackson C, Vasileiou E, Simpson CR, Muckian MD, Agrawal U, et al. Informing the public health response to COVID-19: a systematic review of risk factors for disease, severity, and mortality. BMC infectious diseases. 2021;21(1):342–342. doi:10.1186/s12879-021-05992-1.

[15] Bubar KM, Reinholt K, Kissler SM, Lipsitch M, Cobey S, Grad YH, et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. Science. 2021;371(6532):916–921. doi:10.1126/science.abe6959.

[16] Matrajt L, Eaton J, Leung T, Brown ER. Vaccine optimization for COVID-19: Who to vaccinate first? Science Advances. 2021;7(6):eabf1374. doi:10.1126/sciadv.abf1374.

[17] Sah P, Fitzpatrick MC, Zimmer CF, Abdollahi E, Juden-Kelly L, Moghadas SM, et al. Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. Proceedings of the National Academy of Sciences. 2021;118(34):e2109229118. doi:10.1073/pnas.2109229118.

[18] Shioda K, Lau MSY, Kraay ANM, Nelson KN, Siegler AJ, Sullivan PS, et al. Estimating the Cumulative Incidence of SARS-CoV-2 Infection and the Infection Fatality Ratio in Light of Waning Antibodies. Epidemiology. 2021;32(4).

[19] Hozé N, Paireau J, Lapidus N, Tran Kiem C, Salje H, Severi G, et al. Monitoring the proportion of the population infected by SARS-CoV-2 using age-stratified hospitalisation and serological data: a modelling study. The Lancet Public Health. 2021;6(6):e408–e415. doi:10.1016/S2468-2667(21)00064-5.

[20] Milne G, Hames T, Scotton C, Gent N, Johnsen A, Anderson RM, et al. Does infection with or vaccination against SARS-CoV-2 lead to lasting immunity? The Lancet Respiratory Medicine. 2021;9(12):1450–1466. doi:10.1016/S2213-2600(21)00407-0.

[21] World Health Organization. Tracking SARS-CoV-2 variants; 2022. Available from: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/.

[22] Townsend JP, Lamb AD, Hassler HB, Sah P, Nishio AA, Nguyen C, et al. Projecting the seasonality of endemic COVID-19. medRxiv. 2022;doi:10.1101/2022.01.26.22269905.

[23] Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020;368(6493):860–868. doi:10.1126/science.abb5793.

[24] Neher RA, Dyrdak R, Druelle V, Hodcroft EB, Albert J. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. Swiss Med Wkly. 2020;150. doi:10.4414/smw.2020.20224.

[25] Viana J, van Dorp CH, Nunes A, Gomes MC, van Boven M, Kretzschmar ME, et al. Controlling the pandemic during the SARS-CoV-2 vaccination rollout. Nature Communications. 2021;12(1):3674. doi:10.1038/s41467-021-23938-8.

[26] Pei S, Yamana TK, Kandula S, Galanti M, Shaman J. Burden and characteristics of COVID-19 in the United States during 2020. Nature. 2021;598(7880):338–341. doi:10.1038/s41586-021-03914-4.

[27] Saad-Roy CM, Wagner CE, Baker RE, Morris SE, Farrar J, Graham AL, et al. Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years. Science. 2020;370(6518):811–818. doi:10.1126/science.abd7343.

[28] Lavine JS, Bjornstad ON, Antia R. Immunological characteristics govern the transition of COVID-19 to endemicity. Science. 2021;371(6530):741–745. doi:10.1126/science.abe6522.

[29] Sonabend R, Whittles LK, Imai N, Perez-Guzman PN, Knock ES, Rawson T, et al. Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in England: a mathematical modelling study. The Lancet. 2021;398(10313):1825–1835. doi:10.1016/S0140-6736(21)02276-5.

[30] Dekker MM, Coffeng LE, Pijpers FP, Panja D, de Vlas SJ. Reducing societal impacts of SARS-CoV-2 interventions through subnational implementation. medRxiv. 2022;doi:10.1101/2022.03.31.22273222.

[31] Bosetti P, Tran Kiem C, Andronico A, Colizza V, Yazdanpanah Y, Fontanet A, et al. Epidemiology and control of SARS-CoV-2 epidemics in partially vaccinated populations: a modeling study applied to France. BMC Medicine. 2022;20(1):33. doi:10.1186/s12916-022-02235-1.

[32] Presanis AM, De Angelis D, Goubar A, Gill ON, Ades AE. Bayesian evidence synthesis for a transmission dynamic model for HIV among men who have sex with men. Biostatistics. 2011;12(4):666–681. doi:10.1093/biostatistics/kxr006.

[33] Birrell PJ, Ketsetzis G, Gay NJ, Cooper BS, Presanis AM, Harris RJ, et al. Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. Proceedings of the National Academy of Sciences. 2011;108(45):18238–18243. doi:10.1073/pnas.1103002108.

[34] Rozhnova G, Kretzschmar ME, van der Klis F, van Baarle D, Korndewal M, Vossen AC, et al. Short- and long-term impact of vaccination against cytomegalovirus: a modeling study. BMC Medicine. 2020;18. doi:https://doi.org/10.1186/s12916-020-01629-3.

[35] Anderson JL. An Ensemble Adjustment Kalman Filter for Data Assimilation. Monthly Weather Review. 2001;129(12):2884 – 2903. doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

[36] Pei S, Kandula S, Shaman J. Differential effects of intervention timing on COVID-19 spread in the United States. Science Advances. 2020;6(49):eabd6370. doi:10.1126/sciadv.abd6370.

[37] 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;5(3):1–1. doi:10.1371/journal.pmed.0050074.

[38] Mistry D, Litvinova M, Pastore Y Piontti A, Chinazzi M, Fumanelli L, Gomes MFC, et al. Inferring high-resolution human mixing patterns for disease modeling. Nature Communications. 2021;12(1):323–323. doi:10.1038/s41467-020-20544-y.

[39] Goldstein E, Lipsitch M, Cevik M. On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. The Journal of Infectious Diseases. 2020;doi:10.1093/infdis/jiaa691.

[40] Viner RM, Mytton OT, Bonell C, Melendez-Torres GJ, Ward J, Hudson L, et al. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. JAMA Pediatrics. 2021;175(2):143–156. doi:10.1001/jamapediatrics.2020.4573.

[41] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020;368(6490):489–493. doi:10.1126/science.abb3221.

[42] European Centre for Disease Prevention and Control. SARS-CoV-2 variants of concern as of 14 October 2022.; 2022. Available from: https://www.ecdc.europa.eu/en/covid-19/variants-concern.

[43] Patterns of reported infection and reinfection of SARS-CoV-2 in England. Journal of Theoretical Biology;556:111299.

[44] Miura F, Leung KY, Klinkenberg D, Ainslie KEC, Wallinga J. Optimal vaccine allocation for COVID-19 in the Netherlands: A data-driven prioritization. PLOS Computational Biology. 2021;17(12):1–13. doi:10.1371/journal.pcbi.1009697.

[45] Ainslie KEC, Backer J, de Boer P, van Hoek AJ, Klinkenberg D, Altes HK, et al. The impact of vaccinating adolescents and children on COVID-19 disease outcomes. medRxiv. 2021;doi:10.1101/2021.10.21.21265318.

[46] van Boven M, Hetebrij WA, Swart AM, Nagelkerke E, van der Beek RFHJ, Stouten S, et al. Modelling patterns of SARS-CoV-2 circulation in the Netherlands, August 2020-February 2022, revealed by a nationwide sewage surveillance program. medRxiv. 2022;doi:10.1101/2022.05.25.22275569.

[47] Schoot Uiterkamp MHH, Gösgens M, Heesterbeek H, van der Hofstad R, Litvak N. The role of inter-regional mobility in forecasting SARS-CoV-2 transmission. Journal of The Royal Society Interface. 2022;19(193):20220486. doi:10.1098/rsif.2022.0486.

[48] Gösgens M, Hendriks T, Boon M, Steenbakkers W, Heesterbeek H, van der Hofstad R, et al. Trade-offs between mobility restrictions and transmission of SARS-CoV-2. Journal of The Royal Society Interface. 2021;18(175):20200936. doi:10.1098/rsif.2020.0936.

[49] de Vlas SJ, Coffeng LE. Achieving herd immunity against COVID-19 at the country level by the exit strategy of a phased lift of control. Scientific Reports. 2021;11(1):4445. doi:10.1038/s41598-021-83492-7.

[50] Brand SPC, Ojal J, Aziza R, Were V, Okiro EA, Kombe IK, et al. COVID-19 transmission dynamics underlying epidemic waves in Kenya. Science. 2021;374(6570):989–994. doi:10.1126/science.abk0414.

[51] Kishore N, Taylor AR, Jacob PE, Vembar N, Cohen T, Buckee CO, et al. Evaluating the reliability of mobility metrics from aggregated mobile phone data as proxies for SARS-CoV-2 transmission in the USA: a population-based study. The Lancet Digital Health. 2022;4(1):e27–e36. doi:10.1016/S2589-7500(21)00214-4.

[52] Boldea O, Cornea-Madeira A, Madeira J. Disentangling the effect of measures, variants and vaccines on SARS-CoV-2 Infections in England: A dynamic intensity model. medRxiv. 2022;doi:10.1101/2022.03.09.22272165.

[53] Cohen LE, Spiro DJ, Viboud C. Projecting the SARS-CoV-2 transition from pandemicity to endemicity: Epidemiological and immunological considerations. PLOS Pathogens. 2022;18(6):1–21. doi:10.1371/journal.ppat.1010591.

[54] Bobrovitz N, Ware H, Ma X, Li Z, Hosseini R, Cao C, et al. Protective effectiveness of prior SARS-CoV-2 infection and hybrid immunity against Omicron infection and severe disease: a systematic review and meta-regression. medRxiv. 2022;doi:10.1101/2022.10.02.22280610.

[55] Backer JA, Mollema L, Vos ER, Klinkenberg D, van der Klis FR, de Melker HE, et al. Impact of physical distancing measures against COVID-19 on contacts and mixing patterns: repeated cross-sectional surveys, the Netherlands, 2016-17, April 2020 and June 2020. Eurosurveillance. 2021;26(8). doi:10.2807/1560-7917.ES.2021.26.8.2000994.

[56] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New England Journal of Medicine. 2020;doi:10.1056/NEJMoa2001316.

[57] Hortaçsu A, Liu J, Schwieg T. Estimating the fraction of unreported infections in epidemics with a known epicenter: An application to COVID-19. Journal of Econometrics. 2021;220(1):106–129. doi:10.1016/j.jeconom.2020.07.047.

Acknowledgements

G.R. was supported by the VERDI project (101045989), 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 Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. G.R. and S.P. were supported by the Fundação para a Ciência e a Tecnologia project

2022.01448.PTDC. O.B. and A.A. were supported by the Tilburg University Fund. J.S. and S.P. were supported by US NIAID grant R01AI163023 and CDC contract 75D30122C14289. We thank Professor Marc Bonten (University Medical Center Utrecht), Professor Dennis Huisman (Erasmus University Rotterdam) and Jan Banninga (Nederlandse Spoorwegen) for their substantial help in obtaining the data for this study. We also thank Bettina Siflinger (Tilburg University) for the Tilburg CentER LISS Panel Survey questions added to support our study, and Marino van Zelst (Wageningen University) for useful discussions on data availability and cleaning.

Competing interests

J.S. and Columbia University declare partial ownership of SK Analytics. The other authors declare no competing interests.

Additional information

Supplementary Material contains details of the data, the model, inference, system identifiability, sensitivity analyses, Supplementary figures and tables.

Correspondence

Correspondence and material requests should be addressed to Dr. Otilia Boldea, Department of Econometrics and OR, and CentER, Tilburg School of Economics and Management, Tilburg University, P.O. Box 90153, Tilburg, The Netherlands, email: o.boldea@tilburguniversity.edu; Dr. Ganna Rozhnova, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, P.O. Box 85500 Utrecht, The Netherlands, email: g.rozhnova@umcutrecht.nl.