UNIVERSITÀ DEGLI STUDI DI UDINE
hic sunt futura

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

**UNIVERSITY OF UDINE**
in partnership with **BRUNO KESSLER FOUNDATION**

PhD Course in Computer Science, Mathematics and Physics
XXXIV Cycle

---

# Assessing the determinants of COVID-19 burden to address disease-control policy decisions

---


*Author:*
**Margherita Galli**

*Supervisor:*
**Dr. Stefano Merler**

Academic Year 2021 - 2022

# Abstract

Mathematical modeling has been crucial to address fundamental issues related to COVID-19 disease-control policy decisions. This thesis deals with the robust quantification of the disease burden associated with COVID-19 across different socio-demographic settings. The presented work includes the statistical analysis of novel epidemiological records to provide solid estimates describing the clinical course of SARS-CoV-2 infections and the simulation of data-driven models to forecast the potential impact of COVID-19 in rural and urban areas of Ethiopia. Obtained estimates show that being older than 60 years of age is associated with about 40% likelihood of developing symptoms after SARS-CoV-2 infection and 1% risk of requiring intensive care. The analysis of potential SARS-CoV-2 transmission in Ethiopia suggests that the low prevalence and mortality observed during 2020 can be explained by combined effect of younger demography and a reduced transmission generated by school closures implemented in response to the pandemic. Provided estimates highlight that in this country, after the launch of vaccination in 2021, the highest fraction of severe cases is expected to arise from the interaction between children (who are the main responsible for the spread of the disease) with the elderly (representing the most vulnerable population segment). Remarkably, prioritizing the vaccination of the elderly emerged as the best strategy to reduce the number of critical patients, irrespectively to the limited number of doses made available to low-income settings.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 The impact of mathematical modeling on public health decisions

Mathematical modeling plays a pivotal role in understanding epidemiology and supporting public health responses to epidemics. A clear example is provided by the current COVID-19 crisis, during which mathematical models have been crucial to address fundamental issues related to disease-control policy decisions, unraveling hidden processes and quantifying key epidemiological parameters driving the disease spread, and providing data-driven perspectives on the potential impact of non-pharmaceutical interventions and alternative vaccination strategies.

The first attempt to model and explain disease patterns dates back to the twentieth century when the nonlinear dynamics of infectious disease transmission were firstly recognized. In 1906, Hemer understood that the decrease in the number of susceptible individuals alone could bring the epidemic to a halt (Hamer, 1906). However, the basic foundations of mathematical epidemiology are to be attributed to the work published in August 1927 by Kermack and McKendrick (Kermack and McKendrick, 1927), where the authors pioneered the idea of describing the dynamics of disease transmission in terms of a system of differential equations. Epidemiological modeling has made huge steps in understanding the mechanisms behind the spread of epidemics, incorporating stochasticity and the heterogeneous structures of the host population affecting the diffusion of different pathogens. Nowadays, approaches adopted could take into account different hidden factors which drive the transmission process, including human mobility, social mixing patterns, demographic structure, eco-climatic factors and they are often used to mimic changes in the infection spread caused by alternative intervention measures (e.g. school closures, vaccination, case isolation, etc.) (Keeling and Rohani, 2011; Dare et al., 2015; Marziano et al., 2021a; Li et al., 2017; Di Domenico et al., 2020; Kiem et al., 2021). In the last decades, epidemiological models have been integrated into computational frameworks complemented with detailed public health records, like contacts tracing, surveillance data and genomic sequencing data. Developed frameworks allowed helped to improve the appropriate management of uncertainty surrounding the model estimates and consequent epidemiological forecasts, and the development of plausible scenarios to explore expected outcomes to aid surveillance systems to cope with epidemic threats.

The extensive use of mathematical modeling evaluating intervention strategies markedly emerged in the 90s, when models were extensively used to plan the public response

strategies for mitigating the severity of potential influenza pandemics and evaluate vaccination programs to control childhood diseases epidemics (Gupta, Ferguson, and Anderson, 1998; Longini Jr and Halloran, 1996; Brisson et al., 2000). More recently, transmission models have been used to analyze data in real-time to assess the effectiveness of intervention measures and forecast possible future epidemic trajectories of Ebola, Zika and to support public decisions during the COVID-19 crisis (Wallinga and Teunis, 2004; Guanghong et al., 2004). During the ongoing pandemic, the analysis of contact-tracing data and detailed surveillance records have represented a fundamental aspect to robustly quantify hidden valuables influencing the infection transmission in the human population. However, the use of mathematical models may become even more crucial to analyze erratic data coming for countries with vulnerable economies and public health systems, where the surveillance system is extremely weak, as it is the case of several African countries.

This thesis aims at investigating appropriate modeling tools to provide solid estimates describing the clinical course of SARS-CoV-2 infections and the potential burden of COVID-19, taking into account the potential role played by the pathogen characteristics and by the heterogeneous socio-demographic structure of the population in the spread of COVID-19 disease. A specific focus of the thesis has been represented by epidemiological circumstances that might characterized the spread of SARS-CoV-2 infection across different geographical contexts of Ethiopia.

## 1.2 State of art

Mathematical modeling has been one of the cornerstones in the response to the COVID-19 pandemic. A lot of work have been done on the projection of the COVID-19 spread and the evaluation of the impact of different control measures (Marziano et al., 2021a; Guzzetta et al., 2021; Saad-Roy et al., 2020; Chinazzi et al., 2020; Ferguson et al., 2020; Hellewell et al., 2020; Kucharski et al., 2020; Trentini et al., 2021; Marziano et al., 2021b). This include studies focusing on estimating the effectiveness of the vaccination programs in reducing the burden on healthcare system and investigating the transmission potential of newly emerged SARS-CoV-2 variants (Marziano et al., 2021b; Harris et al., 2021; Subbarao et al., 2021; Sheikh et al., 2021; Thiruvengadam et al., 2021; Pouwels et al., 2021; Falsey et al., 2021). To address fundamental questions, a lot of work has developed to explore the hidden mechanisms and estimate key parameters driving the infection spread (Kiem et al., 2021; Verity et al., 2020; Vespignani et al., 2020; Poletti et al., 2020a; Poletti et al., 2021;Onder, Rezza, and Brusaferro, 2020; Yang et al., 2020). Several studies had shown that the risk of experiencing symptomatic, critical disease, or death after infection increases with the age (Poletti et al., 2020a; Poletti et al., 2021; Onder, Rezza, and Brusaferro, 2020; Yang et al., 2020; Kiem et al., 2021; Verity et al., 2020). However, the robustness of estimated risks strongly depends on the quality of data and the appropriate management of biases affecting the analyzed records, which often consist of data collected during the passive surveillance of the infesction or the disease. As a matter of fact, risk factors describing the natural clinical course of the infection remained poorly quantified (Davies et al., 2020; Wu et al., 2020). Difficulties in quantifying metrics required to estimate the disease burden of COVID-19 are related to challenges in assessing an unbiased sample of the infections which represents the appropriate denominator for estimating the robust risks of outcomes (e.g., deaths, severe disease,

respiratory symptoms) after infection (Verity et al., 2020; Poletti et al., 2020a; Poletti et al., 2021). Indeed, the quantitative estimates of the clinical course of the infection based only on confirmed cases could result in risk outcomes biased upward because of the higher probability of analyzing the symptomatic cases and infected individuals experiencing more severe symptoms (Verity et al., 2020; Poletti et al., 2020a; Poletti et al., 2021; Biggerstaff et al., 2020; Wu et al., 2020). An illustrative example of the huge uncertainty caused by this phenomenon is provided by the high variability around the available estimates of the proportion of symptomatic infections, ranging from 3% to 87% (Buitrago-Garcia et al., 2020; Byambasuren et al., 2020; Emery et al., 2020; Nikolai et al., 2020; Oran and Topol, 2020; Poletti et al., 2021).

The first goal of this thesis is to quantify age-specific probabilities of transitions between stages defining the natural history of SARS-CoV-2 infection from a sample of SARS-CoV-2 positive individuals identified in Italy between March and April 2020 among contacts of confirmed cases. The carried out analysis also provide estimates of time intervals between key events defining the temporal clinical progression of cases, as obtained from a larger sample of infections ascertained between February and July 2020.

Unraveling the real burden of disease associated with the transmission events occurring in the community may be an even more critical task in low-income countries that are characterized by a lack of reliable epidemiological records and limited availability of, and access to, healthcare resources and infrastructures (Gilbert et al., 2020; Poletti et al., 2018). A troubling example is given by Ethiopia, with a healthcare workforce that is five times lower than the minimum threshold defined by the WHO for Sustainable Development Goals health targets (*World Health organization. Health Workforce Requirements for Universal Health Coverage and the Sustainable Development Goals. Human Resource for Health Observers Series No. 17.*), and far below the African average (Haileamlak, 2018). This already difficult situation is exacerbated by the spread of COVID-19, which is sustained by transmission from a large share of asymptomatic infections, and by the recurrent emergence of SARS-CoV-2 variants, such as the hyper-transmissible Delta and Omicron variant, which has progressively replaced the historical strain in many countries (*Genomic epidemiology of novel coronavirus - Africa-focused subsampling*).

On the other hand, African countries have so far experienced a lower burden in terms of prevalence and mortality compared to Europe and the United States, as well as to numerous upper-middle-income countries in South America and Asia. While SARS-CoV-2 has spread rapidly worldwide causing unprecedented pressure on the healthcare system in most countries, it remains unclear why Africa was partially spared from a marked number of cases and deaths. Many factors could help to explain the experience of COVID-19 in Africa such as demography, sociocultural aspects, environmental exposures, genetics, and the immune system (Mbow et al., 2020). However, vulnerabilities of the health system in these countries should be considered as these can possibly lead to an underestimation of the number of people who have been affected by the infection and the disease. The lack of solid data is indeed a critical aspect for understanding the true impact of the disease, planning effective mitigation strategies and implementing relevant intervention measures.

The work presented in this thesis aims to assess how demographic factors and age-specific contact patterns can influence the impact of COVID-19 epidemics across different geographical contexts of the South West Shewa Zone (SWSZ) of the Oromia Region of Ethiopia.

Despite the tremendous achievements in the deployment of vaccination among high-income countries, the current vaccination coverage of Ethiopia is one of the lowest in Africa, with only 1.35% of the citizens being fully immunized (*Our world in data. COVID-19 Data Explorer*; *CovidVax*). Further vulnerabilities of this country are the high prevalence of comorbidities (e.g., malnutrition (Endris, Asefa, and Dube, 2017), tuberculosis, and malaria).

To explore appropriate vaccination strategies in settings with limited vaccine supply, a modeling work is here proposed to quantify the impact of different vaccination priority targets and vaccination strategies could have in reducing the number of SARS-CoV-2 infections and COVID-19 critical cases (e.g., requiring intensive care), taking into account the hidden factors that characterize different African socio-demographic contexts (e.g. in rural and urban areas).

## 1.3 Innovative aspects

In this thesis, to reduce potential biases in the identification of infections, the different risk ratios after infection are estimated based on a sample of SARS-CoV-2 positive contacts identified and tested independently of their symptoms and clinical signs The analyzed sample is not characterized by the underestimation of asymptomatic individuals and the lack of longitudinal records about the clinical history of study participants. The strengths of this study design rely on

1. the minimization of the risks of bias in the identification of infections led by the complete testing of all contacts of confirmed cases,

2. the daily follow-up of the infections for symptoms and critical disease in the weeks following the exposure to all analyzed confirmed infections, which allows to assess the exact outcome of potential patients experiencing the disease.

Provided metrics can be instrumental to refine model estimates aimed at assisting the design and evaluation of strategies to control or mitigate the COVID-19 pandemic.

A second innovative aspect of this thesis is given by the analysis of novel social contact data collected, before the pandemic, in the South West Shewa Zone of the Oromia region in Ethiopia. Data consist of individual records on the number and type of contacts experienced by study participants across different geographical contexts characterized by heterogeneous population density, work and travel opportunities, and access to primary care. The study consists of a cross-sectional survey with two-stage stratified random sampling by location and age group. These data, once combined with an appropriate modeling framework, are used to quantify how socio-demographic factors and observed mixing patterns can influence the expected COVID-19 disease burden. By comparing estimates obtained when including and excluding school contacts for the entire duration of the epidemic, the disease burden

averted by school closure is assessed across different geographical contexts.

Finally, leveraging on data on individuals' social mixing, a modeling work where potential sources of infection in terms of the age of the infector individuals are estimated over different phases of the pandemic and across different geographical contexts of the South West Shewa Zone. Different strategies of vaccination are simulated and compared under the assumption of limited vaccine supply, taking into account immunity acquired by both previous infection and vaccination. The model developed and consequent results could be instrumental to inform vaccination strategies in low-income settings.

## 1.4 Structure of the thesis

After this first introductory chapter, the thesis is structured as follows.

- In **chapter 2** I analyze a sample consisting of the line list of SARS-CoV-2 laboratory-confirmed individuals ascertained in Lombardy between February 20 and July 16, 2020, and regularly updated by the regional public health authorities. Information retrieved from this dataset was complemented with contact-tracing records collected between March 10 and April 27 and with the results of a serological survey targeting case contacts conducted between April 16 and June 15, 2020. These data were analyzed to estimate the likelihood of developing respiratory symptoms or fever, of being admitted to a hospital and an ICU, of developing a critical disease, and of dying after SARS-CoV-2 infection. These estimates are provide stratified by age and sex. In addition, 88,538 records on cases ascertained by regular surveillance activities were used to investigate the distribution of patients' length of stay in hospital and in ICU, and the time interval between the following key events: from symptom onset to diagnosis, from symptom onset to hospital and/or ICU admission, from symptom onset to death, and from hospital to ICU admission.

- In **chapter 3** I analyze social contact records collected at the end of 2019 in the South West Shewa Zone of the Oromia region in Ethiopia, and I used this novel dataset to quantify the influence of socio-demographic factors, observed mixing patterns, and the school closure mandate in the COVID-19 disease burden experienced in the first pandemic phase.

- In **chapter 4** I focus on quantifying the contribution of different ages in the spread of the SARS-CoV-2 transmission and critical diseases in the South West Shewa Zone after two years of pandemic. The impact of alternative vaccine priority targets is evaluated.

- **Chapter 5** is devoted to summarizing the obtained results, highlighting the overall conclusions and limitations of all works covered in the thesis.

# Chapter 2

# A quantitative assessment of epidemiological parameters required to investigate COVID-19 burden

## 2.1 Background

Mathematical modeling has been one of the cornerstones in the response to the COVID-19 pandemic (Chinazzi et al., 2020; Ferguson et al., 2020; Guzzetta et al., 2021; Hellewell et al., 2020; Kucharski et al., 2020; Marziano et al., 2021a; McCombs and Kadelka, 2020; Salje H, 2020; Trentini et al., 2021; Vespignani et al., 2020;Wu, Leung, and Leung, 2020). To provide solid estimates, models need to be properly calibrated based on empirical evidence (Biggerstaff et al., 2020; Lau et al., 2020; Ma et al., 2020; Salje H, 2020; Wood et al., 2021). While a lot of work has been done in this direction (Cereda et al., 2020; Lau et al., 2020; Hilton and Keeling, 2020; Ma et al., 2020; Park et al., 2020; Peiris et al., 2003; Riccardo et al., 2020; Zhang et al., 2020), metrics required to estimate the disease burden are still poorly quantified (Davies et al., 2020; Wu et al., 2020). Difficulties in deriving these quantities are related to challenges in defining unbiased denominators (i.e., the infections) for computing different risk outcomes (e.g., deaths, severe disease, respiratory symptoms) upon infection (Poletti et al., 2020a; Poletti et al., 2021; Verity et al., 2020). Indeed, as asymptomatic cases and infected individuals experiencing mild symptoms are, in general, more likely to remain undetected, quantitative estimates of the clinical course of the infection based only on confirmed cases could result in risk outcomes biased upward (Biggerstaff et al., 2020; Poletti et al., 2020a; Poletti et al., 2021; Verity et al., 2020; Wu et al., 2020).

In this work, we provide estimates of the probabilities of transition across the stages characterizing the clinical progression after SARS-CoV-2 infection, stratified by age and sex, as well as of the time delays between key events. To do this, we analyzed a sample of 1,965 SARS-CoV-2 positive individuals who were contacts of confirmed cases. These individuals were identified irrespective of their symptoms as part of contact tracing activities carried out in Lombardy (Italy) over the period from March 10 to April 27, 2020. These individuals were daily monitored for symptoms for at least two weeks after exposure to a COVID-19 case and either tested for SARS-CoV-2 via PCR in real time or retrospectively via IgG serological assays; their clinical history was also recorded. In addition to this highly detailed sample, we relied on the

epidemiological records of all the 95,371 SARS-CoV-2 PCR confirmed infections reported to the surveillance system between February and July 2020. This allowed us to provide a comprehensive quantitative assessment of all the main epidemiological parameters essential to model COVID-19 burden (see Figure 2.1), thus laying the foundation for future COVID-19 modeling efforts.

Estimates on age-specific risk outcomes after SARS-CoV-2 infection were validated against epidemiological records that have not been used to derive these quantities, leveraging on data from two serological surveys conducted in Italy (*Italian National Institute of Statistics2020. Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV-2*; Stefanelli et al., 2021) and on the national cumulative incidence reported up to April 2021 (*Istituto Superiore di Sanità, 2021*).

## 2.2 Methods

### 2.2.1 Study population

Lombardy represents the earliest and most affected region by the first COVID-19 epidemic wave experienced in Italy. Short after the detection of a first COVID-19 case on February 20, 2020, a ban of mass gatherings and the suspension of teaching in schools and universities was applied to the entire region. The interruption of non-essential productive activities and strict individual movement restrictions were imposed to the most affected municipalities. On March 8, 2020, after a rapid increase of cases, closure of all non-necessary businesses and industries and limitations of movements except in cases of necessity were extended to the entire region. A national lockdown was imposed on March 10, 2020. Suspended economic and social activities were gradually resumed between April 14 and May 18, 2020.

### 2.2.2 Data collection

Data analyzed here consists of the line list of SARS-CoV-2 laboratory confirmed infections ascertained in Lombardy between February 20 and July 16, 2020, and regularly updated by the regional public health authorities. Information retrieved from this dataset was complemented with contact-tracing records collected between March 10 and April 27 and with results of a serological survey targeting case contacts conducted between April 16 and June 15, 2020 (Poletti et al., 2020a ; Poletti et al., 2021). Data collection, integration, storage, and anonymization was managed by regional health authorities as part of surveillance activities and outbreak investigations aimed at controlling and mitigating the COVID-19 epidemic in Italy.

### 2.2.3 Definition of COVID-19 case

From February 21 to February 25, 2020, following the criteria initially defined by the European Centre for Disease Prevention and Control (ECDC), suspected COVID-19 cases were identified as:

1. patients with acute respiratory tract infection OR sudden onset of at least one of the following: cough, fever, shortness of breath AND with no other aetiology that fully explains the clinical presentation AND at least one of these other

conditions: a history of travel to or residence in China, OR patients among health care workers who has been working in an environment where severe acute respiratory infections of unknown etiology are being cared for;

2. OR patients with any acute respiratory illness AND at least one of these other conditions: having been in close contact with a confirmed or probable COVID-19 case in the last 14 days prior to onset of symptoms, OR having visited or worked in a live animal market in Wuhan, Hubei Province, China in the last 14 days prior to onset of symptoms, OR having worked or attended a health care facility in the last 14 days prior to onset of symptoms where patients with hospital-associated COVID-19 have been reported.

Confirmed cases were defined as suspect cases testing positive with a specific real-time reverse transcription polymerase chain reaction (RT-PCR) assay targeting multiple genes of SARS-CoV-2 (Cereda et al., 2020; Cohen and Kessel, 2020; Corman et al., 2020). From March 20, 2020 positivity to the nasopharyngeal swab was also granted for assays that tested a single gene. At any time, ascertained infections were defined as laboratory confirmed SARS-CoV-2 infections, irrespective of clinical signs and symptoms. Inconclusive swabs were repeated to reach the diagnosis.

### 2.2.4 Ascertainment of infections among close case contacts

All ascertained SARS-CoV-2 infections were considered as potential index cases for further spread of SARS-CoV-2. Close contacts of these individuals were therefore identified through standard interviews of cases, informed of their possible exposure and quarantined within 24-48 hours from a positive test result on the index case.

A close case contact was defined as a person living in the same household as a COVID-19 confirmed case; a person having had face-to-face interaction with a COVID-19 confirmed case within 2 meters and for more than 15 minutes; a person who was in a closed environment (e.g. classroom, meeting room, hospital waiting room) with a COVID-19 confirmed case at a distance of less than 2 meters for more than 15 minutes; a healthcare worker or other person providing direct care for a COVID-19 confirmed case, or laboratory workers handling specimens from a COVID-19 confirmed case without recommended personal protective equipment (PPE) or with a possible breach of PPE; a contact in an aircraft sitting within two seats (in any direction) of a COVID-19 confirmed case, travel companions or persons providing care, and crew members serving in the section of the aircraft where the index case was seated (passengers seated in the entire section or all passengers on the aircraft were considered close contacts of a confirmed case when severity of symptoms or movement of the case indicate more extensive exposure). Close case contacts were initially considered as contacts occurred between 14 days before and 14 days after the date of symptom onset of the index case. After March 20, 2020 the exposure period was shortened, ranging from 2 days before to 14 days after the symptom onset of the index case (*World Health Organization. Contact tracing in the context of COVID-19: interim guidance, 10 May 2020.*). For individuals unable to sustain the contact tracing interview, close contacts were identified by their parents, relatives or their emergency contacts. From February 20 to February 25, 2020 all contacts of confirmed infections were tested with RT-PCR, irrespective of clinical symptoms. From February 26 onward, the traced contacts were tested with RT-PCR only in case of symptom onset.

However, on April 16, 2020, regional health authorities initiated an IgG serological survey of quarantined case contacts without history of testing against SARS-CoV-2 infection to retrospectively identify all asymptomatic positive contacts. The test used to detect SARS-CoV-2 IgG antibodies was the LIAISONR SARS-CoV-2 test (DiaSorin), employing magnetic beads coated with S1 & S2 antigens. The antigens used in the tests are expressed in human cells to achieve proper folding, oligomer formation, and glycosylation, providing material similar to the native spikes. The S1 and S2 proteins are both targets to neutralizing antibodies. The test provides the detection of neutralizing antibodies with 98.3% specificity and 94.4% sensitivity at 15 days from diagnosis. Performance analyses validating the accuracy of this serological test can be found in Bonelli et al., 2020. Serological test results were binary and communicated to tested participants, who were categorized as seropositive if they had developed IgG antibodies.

All case contacts, irrespectively to the presence of a laboratory diagnosis, were followed up for at least 14 days after exposure to an index case and required by national regulations to report symptoms to local public health authorities. Symptomatic cases were defined as infected subjects showing fever $\geq$ 37.5$C$ or one of the following symptoms: dry cough, dyspnea, tachypnea, difficulty breathing, shortness of breath, sore throat, and chest pain or pressure. The definition of symptoms did not change throughout the period considered in this study. Clinical manifestations, admission to hospital or intensive care units and death among both ascertained infections and their close contacts were regularly updated by the regional health surveillance. In our study, individuals experiencing critical diseases were defined as positive patients who were either admitted to an intensive care unit or died with a diagnosis of SARS-CoV-2 infection. Positive subjects who developed a critical disease are hereafter simply denoted as critical cases. Hospitalized patients with a laboratory confirmation of SARS-CoV-2 infection are denoted as ascertained cases admitted to hospital.

### 2.2.5 Sample selection for computing risk outcomes

A large fraction of case contacts remained untested against SARS-CoV-2 infection, due to difficulties in maintaining a high level of testing during the contact tracing operations and to the relatively low coverage of IgG serological screening conducted on traced contacts. As asymptomatic infections ascertained by surveillance systems are likely under-represented, we selected a subsample of SARS-CoV-2 positive individuals who were tested irrespectively from their symptoms. In particular, we considered infections ascertained among case contacts identified between March 10 and April 27, 2020 and belonging to clusters whose individuals were all tested and daily followed up for symptoms. A fraction of these individuals, mainly symptomatic ones, was tested by RT-PCR during contact-tracing activities. The remaining fraction was confirmed via IgG serological assays collected at least one month after exposure, thus allowing the identification of asymptomatic infections. This study design allowed us to minimize the risks of bias in the identification of infections when computing the proportion of SARS-CoV-2 infections developing symptoms and severe conditions. The resulting subsample consisted of $1,965$ positive subjects

identified in 2,458 clusters of 3,947 close contacts. None of these records showed inconsistent data entries.

### 2.2.6 Statistical analysis

The aforementioned subsample of 1,965 positive individuals who were identified as contacts of confirmed cases was analyzed to estimate the likelihood of developing respiratory symptoms or fever $\geq$ 37.5C (SR), of being admitted to a hospital (HR) and an ICU (IR), of developing critical disease (CR) and of dying after SARS-CoV-2 infection (IFR). The same sample was considered to estimate the case fatality ratio (CFR). Age and sex specific ratios were computed as crude percentages; 95% confidence intervals were computed by exact binomial tests. Logistic regression models were used to estimate the corresponding risk ratios (RRs) using the case age group, sex and month of identification (March or April) as model covariates. For the regression analysis, the following age-groups were considered: 0-59 years, 60-74 years, 75+ years.

The entire sample of cases ascertained by regular surveillance activities (88,538 symptomatic individuals) was used to investigate temporal changes in the COVID-19 disease burden. In particular, we computed the age-specific crude percentage of ascertained cases admitted to hospital (asHR) and the percentage of ICU admissions among hospitalized cases (hIR) for four epidemic periods: before April, April, May and after May.

The same sample of cases was used to investigate the distribution of patients' length of stay in hospital and in ICU, and the time interval between the following key events: from symptom onset to diagnosis, from symptom onset to hospital and/or ICU admission, from symptom onset to death, and from hospital to ICU admission. The time at diagnosis was defined as the time of testing observed for positive individuals. As 3,855 out of 47,393 inpatients had inconsistent data entries on their temporal clinical progression after hospital admission, we excluded the corresponding data records when estimating time to key changes in patients' status, such as hospital or ICU admission and discharge. Specifically, we excluded inpatients with a date of hospital admission or of death preceding the date of symptom onset, patients with a date of ICU admission or death preceding their hospitalization and patients with a negative length of stay in ICU or in hospital. Estimates for the hospital and ICU length of stay and the time between key events are provided for two epidemic periods, defined by considering the date of peak in the COVID-19 incidence experienced during the first epidemic wave in Lombardy, namely March 16, 2020. Cases were aggregated on the basis of the initial date of the considered interval. Negative binomial distributions were used to separately fit each time interval of interest. A negative binomial distribution was considered to better reflect the data characteristics: time lags expressed as integer values (delays measured in days), and a non-negligible proportion of patients with null delays (events occurring within the same day). Specifically, the negative binomial distribution was preferred over the Poisson, truncated normal, Gamma, Weibull, and Log-normal distributions, given that these alternatives were associated with a lower goodness of fit in terms of Akaike Information Criterion or they requested additional assumptions to fit the available records (e.g., the Gamma distribution is defined for strictly positive values only).

To assess the robustness of the estimated risk outcomes with respect to the change in the definition of close contact occurred on March 20, 2020, we investigated how the analyzed metrics would change when considering infections ascertained after that date only. Since in our baseline analysis no assumptions were made on the time from the diagnosis of SARS-CoV-2 to hospital or to ICU admission, we also explored the effect of excluding patients reporting a delay from SARS-CoV-2 diagnosis to hospital or ICU admission greater than 30 days. Specifically, we analyzed the impact of this assumption on the estimated risk outcomes, the time intervals between key events, and the temporal changes in the probability of being admitted to hospital and ICU.

The statistical analysis was performed with the software R (version 3.6), using the "MASS" package. Figure 2.1 provides a schematic representation of all metrics considered to quantify COVID-19 burden.

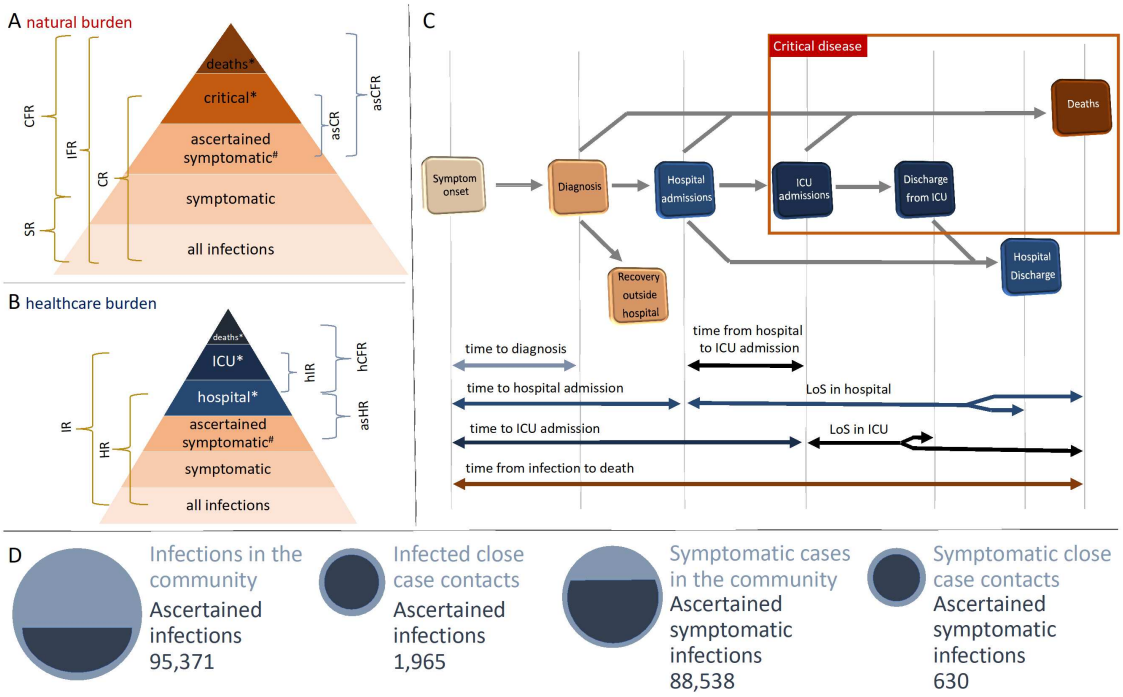


FIGURE 2.1: Schematic representation of transition probabilities characterizing possible disease outcomes after SARS-CoV-2 infection. These include the symptomatic ratio (SR), the ratio of critical cases (CR), the case (CFR) and infection fatality ratios (IFR) and similar quantities that could be estimated using ascertained symptomatic infections (asCR, asCFR) as the set of exposed individuals. B Schematic representation of transition probabilities characterizing the hospital (HR) and ICU (IR) admission among infected individuals, and of similar quantities that could be estimated using ascertained symptomatic infections (asHR) or hospital patients (hCFR, hIR) as the set of exposed individuals. C Schematic representation of time to key events defining the temporal clinical progression of cases. D Schematic representation of the differences in the ascertainment rates associated with SARS-CoV-2 infections and symptomatic cases in the community and among close contacts of identified cases, with the latter representing individuals who were all tested for SARS-CoV-2 infection and daily monitored for symptoms during their quarantine or isolation period.

### 2.2.7 Validation of age-specific risk outcomes

The adopted approach was validated by applying our estimates for age-specific risk outcomes given SARS-CoV-2 infection to seroprevalence data available for Italy and comparing the obtained results with the age distribution of critical cases and deaths observed in Lombardy during the first pandemic wave and throughout Italy up to April 2021. Combining the estimated risk outcomes with a serological study conducted in a specific period would be inappropriate to estimate the absolute number of patients associated with different outcomes at a different time. However, the rationale of applying the estimated risk outcomes to independent seroprevalence data (collected at a different time) was to test whether the provided estimates could

be used to reproduce the age profiles characterizing critical patients and deaths recorded over different periods.

Specifically, we computed the expected age distribution of critical cases $C(a)$ and deaths $D(a)$ as

$$C(a) = \frac{i(a)CR(a)}{\sum_a i(a)CR(a)}$$

and

$$D(a) = \frac{i(a)IFR(a)}{\sum_a i(a)IFR(a)}$$

where $i(a)$ is the number of SARS-CoV-2 IgG positive individuals identified in the age class a through serological surveys, $CR(a)$ and $IFR(a)$ represent our estimates for the probability of developing critical disease and the infection fatality ratio for the age class a. $i(a)$ was retrieved from: 1) a serological study conducted at the national level between May 25 and July 15, 2020 (*Italian National Institute of Statistics2020. Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV-2*) and 2) results of an extensive serological screening applied between May 5 and May 15, 2020 to 77% of individuals residing in a high-incidence area (approximately 8,000 residents) located in north-eastern Italy (Stefanelli et al., 2021). Resulting values for $C(a)$ were compared to the age distribution of all critical cases recorded in Lombardy between February 20 and July 16, 2020. Values obtained for $D(a)$ were compared to the age distribution of cumulative deaths recorded in Lombardy until July 16, 2020 and that observed at the national level between February 2020 and April 2021. The latter was obtained by using cumulative notification data stratified by age as provided by the Integrated National Surveillance System (NSS) (*Istituto Superiore di Sanità, 2021*). Validation of risk outcomes was carried by considering the following age-groups: 0-19, 20-39, 40-59, 60-69, 70+ years.

### 2.2.8 Ethical statement

Data collection and analysis were part of outbreak investigations during a public health emergency. Processing of COVID-19 data is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care, and therefore exempted from institutional review board approval (Regulation EU 2016/679 GDPR).

## 2.3 Results

### 2.3.1 Sample description

We analyzed a total of 95,371 laboratory confirmed infections ascertained between February and July 2020. Of these, 88,538 (92.8%, median age 65 years, IQR: 50-81) reported respiratory symptoms or fever $\geq 37.5C$, 47,393 (49.7%, median age 69 years, IQR: 55-80) were hospitalized, 19,020 (19.9%, median age 79 years, IQR: 70-86) developed critical disease (i.e., requiring ICU treatment or resulting in a fatal outcome) and 16,778 (17.6%, median age 81 years, IQR: 73-87) died with a diagnosis of SARS-CoV-2 (Table 2.1 ).

By combining the regional line list of all ascertained infections with contact-tracing records collected between March 10 and April 27, 2020, we obtained a subsample of 1,965 (median age 53 years, IQR: 32-64) contacts who resulted positive to SARS-CoV-2. Of these, 630 (32.1%, median age 57 years, IQR: 42.5 - 71) developed symptoms, 266 (13.5%, median age 64 years, IQR: 53.25 - 76) were hospitalized, 43 (2.2%, median age 76 years, IQR: 69 - 81) experienced critical disease conditions, 12 (0.6%, median age 68 years, IQR: 52.5 - 72) were admitted to ICUs, and 35 (1.8%, median age 78 years, IQR: 74.5 – 82.5) resulted in a fatal outcome; 31 (1.6%, median age 79 years, IQR: 75-84) subjects died without being admitted to ICU; 4 (0.2%, median age 73.5 years, IQR: 71.25-75) died after an ICU admission (Table 2.2).

| | Symptomatic cases | Hospitalized patients | | Critical cases | | Deaths | |
|---|---|---|---|---|---|---|---|
| | Count | Count | Risk ratio (95%CI) | Count | Risk ratio (95%CI) | Count | Risk ratio (95%CI) |
| **Age** | | | | | | | |
| **≥ 80** | 24,092 | 11,849 | Reference | 9,325 | Reference | 9,291 | Reference |
| **0-39** | 11,019 | 3,361 | 0.54 (0.52 - 0.56) | 158 | 0.03 (0.03 - 0.04) | 38 | 0.01 (0.01 - 0.01) |
| **40-59** | 25,910 | 12,037 | 0.77 (0.75 - 0.79) | 1,747 | 0.12 (0.12 - 0.13) | 733 | 0.05 (0.05 - 0.06) |
| **60-69** | 12,731 | 8,917 | 1.22 (1.19 - 1.24) | 2,642 | 0.35 (0.33 - 0.37) | 1,872 | 0.24 (0.23 - 0.26) |
| **70-79** | 14,784 | 11,228 | 1.39 (1.37 - 1.41) | 5,147 | 0.65 (0.63 - 0.68) | 4,844 | 0.62 (0.60 - 0.64) |
| **Unknown** | 2 | 1 | 1.13 (0.06 - 1.99) | 1 | 0.86 (0.05 - 2.40) | 0* | - |
| **Sex** | | | | | | | |
| **Female** | 46,234 | 19,318 | Reference | 7,323 | Reference | 6,804 | Reference |
| **Male** | 42,168 | 28,061 | 1.49 (1.47 - 1.51) | 11,682 | 1.87 (1.82 - 1.92) | 9,966 | 1.78 (1.73 - 1.84) |
| **Unknown** | 136 | 14 | 0.20 (0.11 - 0.33) | 15 | 1.28 (0.75 - 1.97) | 8 | 0.84 (0.38 - 1.57) |
| **Epidemic period** | | | | | | | |
| **Before April** | 56,288 | 37,391 | Reference | 14,473 | Reference | 12,530 | Reference |
| **April** | 21,022 | 6,909 | 0.52 (0.51 - 0.54) | 3,451 | 0.49 (0.47 - 0.51) | 3,239 | 0.48 (0.46 - 0.50) |
| **May** | 6,019 | 1282 | 0.36 (0.34 - 0.38) | 313 | 0.20 (0.18 - 0.22) | 276 | 0.19 (0.16 - 0.21) |
| **After May** | 3,596 | 591 | 0.27 (0.25 - 0.29) | 44 | 0.06 (0.04 - 0.08) | 39 | 0.06 (0.04 - 0.08) |
| **Unknown** | 1,613 | 1,220 | 1.15 (1.11 - 1.18) | 739 | 1.56 (1.45 - 1.66) | 694 | 1.63 (1.51 - 1.76) |

TABLE 2.1: Estimated risk ratios of hospital admission, experiencing critical disease, and fatal outcome among symptomatic cases, disaggregated by age, sex, and period.

* RR and 95%CI were not computed for insufficiently large sample size

| | Positives contacts | Symptomatic cases | | Critical cases | | Deaths | | | Hospitalized patients | | ICU-admitted patients | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Count | Count | Proportion (95% CI) | Count | Proportion (95% CI) | Count | IFR (95% CI) | CFR (95% CI) | Count | Proportion (95% CI) | Count | Proportion (95% CI) |
| **Age** | | | | | | | | | | | | |
| **0-14** | 219 | 39 | 17.8% (13-23.5%) | 0 | 0% (0-1.7%) | 0 | 0% (0-1.7%) | 0% (0-9%) | 4 | 1.8% (0.5-4.6%) | 0 | 0% (0-1.7%) |
| **15-19** | 22 | 6 | 27.3% (10.7-50.2%) | 0 | 0% (0-15.4%) | 0 | 0% (0-15.4%) | 0% (0-45.9%) | 2 | 9.1% (1.1-29.2%) | 0 | 0% (0-15.4%) |
| **20-39** | 377 | 99 | 26.3% (21.9-31%) | 2 | 0.5% (0.1-1.9%) | 0 | 0% (0-1%) | 0% (0-3.7%) | 18 | 4.8% (2.9-7.4%) | 2 | 0.5% (0.1-1.9%) |
| **40-59** | 662 | 213 | 32.2% (28.6-35.9%) | 5 | 0.8% (0.2-1.8%) | 2 | 0.3% (0-1.1%) | 0.9% (0.1-3.4%) | 89 | 13.4% (10.9-16.3%) | 3 | 0.5% (0.1-1.3%) |
| **60-69** | 331 | 106 | 32% (27-37.3%) | 5 | 1.5% (0.5-3.5%) | 3 | 0.9% (0.2-2.6%) | 2.8% (0.6-8%) | 49 | 14.8% (11.2-19.1%) | 3 | 0.9% (0.2-2.6%) |
| **70-79** | 240 | 94 | 39.2% (33-45.7%) | 17 | 7.1% (4.2-11.1%) | 16 | 6.7% (3.9-10.6%) | 17% (10.1-26.2%) | 59 | 24.6% (19.3-30.5%) | 4 | 1.7% (0.5-4.2%) |
| **≥ 80** | 114 | 73 | 64% (54.5-72.8%) | 14 | 12.3% (6.9-19.7%) | 14 | 12.3% (6.9-19.7%) | 19.2% (10.9-30.1%) | 45 | 39.5% (30.4-49.1%) | 0 | 0% (0-3.2%) |
| **Sex** | | | | | | | | | | | | |
| **Female** | 1,111 | 365 | 32.9% (30.1-35.7%) | 19 | 1.7% (1-2.7%) | 16 | 1.4% (0.8-2.3%) | 4.4% (2.5-7%) | 139 | 12.5% (10.6-14.6%) | 5 | 0.5% (0.1-1%) |
| **Male** | 854 | 265 | 31% (27.9-34.3%) | 24 | 2.8% (1.8-4.2%) | 19 | 2.2% (1.3-3.5%) | 7.2% (4.4-11%) | 127 | 14.9% (12.6-17.4%) | 7 | 0.8% (0.3-1.7%) |

TABLE 2.2: Estimated crude percentages of symptomatic, hospitalized, ICU admitted, and critical cases among SARS-CoV-2 positive individuals who were identified as contacts of confirmed cases as well as estimated risk of death among positive individuals (i.e., infections) and symptomatic case (i.e., infected and symptomatic) individuals who were identified as contacts of confirmed cases. Results are disaggregated by age and sex.

### 2.3.2 Metrics of COVID-19 burden

Age-specific transition probabilities characterizing the different outcomes after SARS-CoV-2 infection were estimated by considering infections occurred among close case contacts identified between March 10 and April 27, 2020. We found that the likelihood of developing respiratory symptoms or fever $\geq$ 37.5 C after SARS-CoV-2 infection (SR) was 27.9% (95%CI: 25.4-30.4%) under 60 years of age and 39.9% (95%CI: 36.2-43.6%) above (see Table 2.2). We estimated that, in the first age-group, 8.8% (95%CI: 7.3-10.5%) of infected individuals required hospital care (HR) and 0.4% (95%CI: 0.1-0.9%) were admitted to ICU (IR); the corresponding proportions in positive individuals older than 60 years were 22.3% (95%CI: 19.3-25.6%) and 1% (95%CI: 0.4-2.1%), respectively. A significantly higher risk of developing critical disease after infection (CR) was found above 60 years of age when compared to younger individuals: 5.3% (95%CI: 3.7-7.2%) vs 0.5% (95%CI: 0.2-1.1%). The infection fatality ratio (IFR) ranged between 0.2% (95%CI: 0.0-0.6%) in subjects younger than 60 years to 12.3% (95%CI: 6.9-19.7%) for those aged 80 years or more. The case fatality ratio (CFR) in these two age groups was 0.6% (95%CI: 0.1-2%) and 19.2% (95%CI: 10.9-30.1%). Although the case fatality ratio was higher for subjects older than 80 years compared to cases aged 60-79 years (namely, 9.5%, 95%CI: 5.8-14.4%), a significantly lower proportion of ICU admissions was found for the oldest age segment: 1.2% (95%CI: 0.5-2.5%) vs 0% (95%CI: 0-3.2%). A detailed age-stratification of all these quantities is provided in Table 2.2. The strong age dependency in the risk of

developing symptoms and most severe outcomes after SARS-CoV-2 infection was confirmed by a statistical analysis based on generalized linear models applied to infected case contacts and accounting for possible confounding factors (see Table 2.4). The regression analysis also highlighted a significantly higher risk ratio (RR) of hospital admission (RR: 1.34, 95%CI: 1.07-1.67), critical disease (RR: 2.16, 95%CI: 1.17-3.98), and death (RR: 2.15, 95%CI: 1.08-4.27) for infected males as compared to females (Table 2.4).

Figure 2.2 compares the age distributions of critical cases and deaths observed in Lombardy and in Italy with those resulting when applying our estimates for risk outcomes after SARS-CoV-2 infection to serological data available for the Italian context (Istituto Superiore di Sanità,2021, Italian National Institute of Statistics, 2020; Stefanelli et al., 2021). These findings highlight that, although estimates for CR and IFR were obtained from a relatively small sample of case contacts identified during the first pandemic phase (1,965 subjects), they can well capture the age profiles characterizing the entire line list of critical patients and deaths recorded in Lombardy during the first COVID-19 wave and the age distribution of all deaths officially recorded across the entire Italian territory until 29 April 2021.

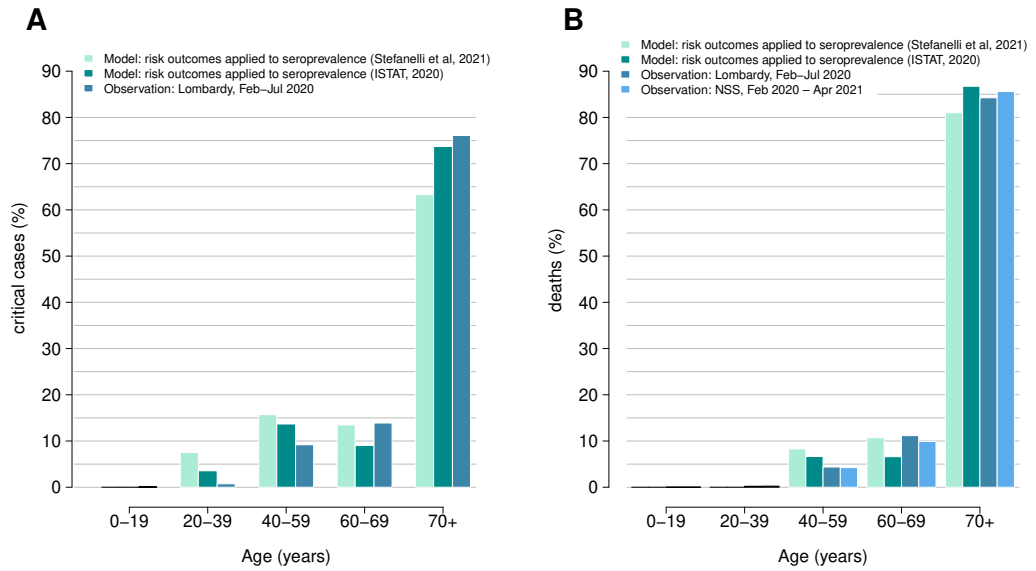


FIGURE 2.2: **A** Comparison between the age distributions of critical cases as obtained when applying estimated risk outcomes to available serological records with the one observed in Lombardy during the first COVID-19 wave. **B** Comparison between the age distributions of deaths as obtained when applying estimated risk outcomes to available serological records with the one observed in Lombardy during the first COVID-19 wave and the one associated to deaths occurred in Italy between February 2020 and April 2021, as reported by the Integrated National Surveillance System (NSS).

### 2.3.3 Temporal changes in the investigated risk metrics

Temporal changes in the risk of being admitted to hospital and ICUs were explored by analyzing records of the complete list of 88,538 symptomatic cases ascertained between February and July 2020 (see Table 2.1 and Table S2 for sample description). The analyzed data includes inpatients with inconsistencies in dates defining the temporal clinical progression after hospitalization. Crude ratios computed from ascertained symptomatic cases should be carefully interpreted because of possible biases due to higher ascertainment rates among more severe cases. However, the analysis of this large sample highlighted an increase of admission rates at different levels of intensity of care among the elderly (Figure 2.3). In particular, hospital admission ratios among ascertained symptomatic cases (asHR) aged more than 80 years increased from the 26.4% (95%CI: 25.5-27.2%) observed between April and May to 34.7% (95%CI: 30.5-39.1%) afterwards. Similarly, the ICU admission ratio among patients hospitalized (hIR) in this age group raised from the 0.9% (95%CI: 0.7-1.1%) observed between March and April to 2.3% (95%CI: 1.2-3.8%) afterwards.

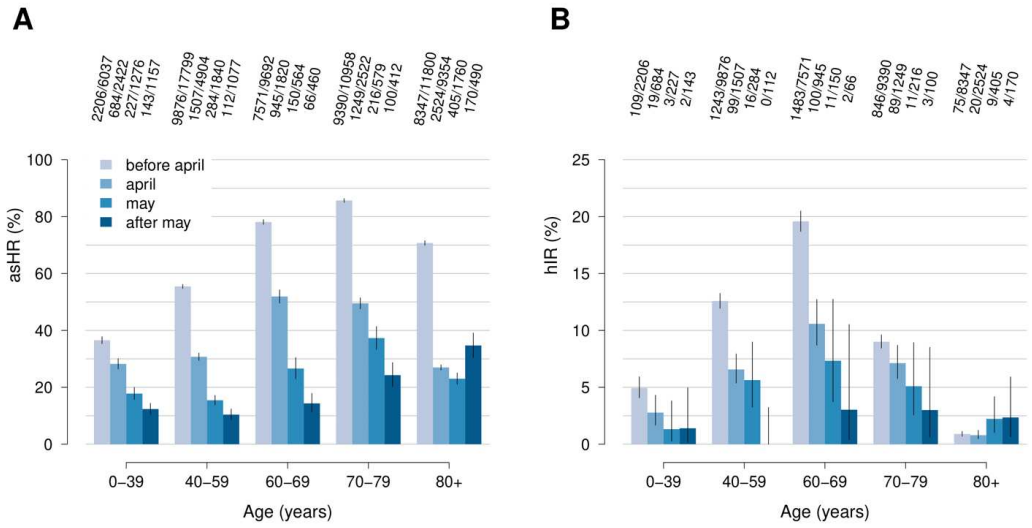


FIGURE 2.3: **A** Age-specific case hospital admission ratios among ascertained symptomatic cases (asHR). **B** Age-specific ICU admission ratios among hospitalized cases (hIR). Bars of different colors represent crude percentages observed across different epidemic periods; vertical lines represent 95% confidence intervals computed by exact binomial tests. Numbers shown in each panel represent the age-specific number of events observed in the data among exposed COVID-19 cases.

### 2.3.4 Time to key events

Time delays from symptom onset to diagnosis and death were investigated by analyzing all the 88,538 symptomatic infections ascertained in Lombardy between February and July 2020. The temporal clinical progression of inpatients was investigated by analyzing 43,538 hospitalized cases (Table 2.3), after having excluded 3,855 out of the 47,393 available inpatient records because of inconsistent dates of hospital or ICU admission/discharge. We estimated that, overall, the median delay between

symptom onset and diagnosis was 4 (IQR: 1-10) days. The median time from symptom onset to death was 12 (IQR: 7-21) days. Hospitalization of cases occurred 5 (IQR: 2-9) days after patients' symptom onset; admission to ICU occurred 10 (IQR: 6-15) days after symptom onset. The median time between hospital and ICU admission was 3 (IQR: 0-6) days. The median hospital length of stay was 10 (IQR: 3-21) days, while the median length of stay in ICU was 11 (IQR: 6-19) days. A negative binomial distribution was used to separately fit each time interval of interest (Figure 2.4).

When looking at these variables across different ages, we found a shorter delay between symptom onset and death in individuals older than 70 years (11-12 days vs 15-16 days at younger ages) and a shorter length of stay in ICU among patients aged 80 years or more (5 days vs 9-12 days at younger ages). We separately analyzed these quantities for cases who developed symptoms before and after March 16, 2020, corresponding to the peak in the number of hospitalized patients in Lombardy (Figure 2.5). We found a marked decrease in the time required to both diagnose and hospitalize COVID-19 patients after this date (from 7 to 2 days and from 7 to 3 days, respectively, Table 2.3). It is worth noting that the lag between the time of the test and the time when the test result became available remained approximately constant during the entire period considered (ranging from 2 to 4 days). Detailed estimates obtained on the temporal clinical progression of COVID-19 cases are reported in table 2.3.

By considering only positive contacts ascertained after March 20, 2020, when the definition of close contact changed, a lower likelihood of experiencing critical disease and death for positive individuals older than 70 years and females was found (see Table 2.6). Such differences may be linked to the enhancement of the tracing and treatment procedures during the first month of the COVID-19 epidemic, which may include a faster detection and diagnosis of SARS-CoV-2 infections and shorter time lags between diagnosis and hospitalization of severe patients. On the other hand, our estimates did not change when excluding patients with a delay from SARS-CoV-2 diagnosis to hospital or ICU admission greater than 30 days (Tables 2.7 and 2.8, and Figure 2.6).

| | Median days (IQR) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Time between symptom onset and diagnosis of SARS-CoV-2 | Time between symptom onset and death | Time between symptom onset and hospital admission | Time between symptom onset and ICU admission | Time between hospital and ICU admission | Hospital length of stay | ICU length of stay |
| **Age** | | | | | | | |
| **0-39** | 3 (0-11) | 16 (6-27) | 4 (1-8) | 8 (4-11) | 1 (0-5) | 4 (0-10) | 9 (4-15.75) |
| **40-59** | 5 (1-11) | 15 (9-26) | 6 (2-10) | 10 (6-13) | 2 (0-6) | 9 (1-18) | 11 (6-19) |
| **60-69** | 5 (1-10) | 16 (9-25) | 6 (2-10) | 11 (7-15) | 3 (0-7) | 13 (6-24) | 12 (6-20) |
| **70-79** | 5 (1-9) | 12 (7-20) | 5 (2-9) | 10 (6-16) | 4 (1-7) | 12 (5-24) | 10 (5-18) |
| **≥ 80** | 3 (0-8) | 11 (6-19) | 4 (1-8) | 9 (4-23) | 2 (0-11.5) | 10 (4-24) | 5 (3-10.75) |
| **Unknown** | 13 (9.5-16.5) | 0 (0-0) | 8 (8-8) | 8 (8-8) | 0 (0-0) | 14 (14-14) | 14 (14-14) |
| **Epidemic period** | | | | | | | |
| **Before 16 March** | 7 (3-12) | 13 (8-21) | 7 (3-10) | 10 (7-15) | 3 (0-6) | 10 (3-21) | 11 (6-19) |
| **After 16 March** | 2 (0-7) | 11 (6-20) | 3 (1-7) | 9.5 (5-14) | 4 (0-8) | 10 (3-23) | 11 (6-20) |
| **Unknown** | 0 (0-0) | 0 (0-0) | 0 (0-0) | 0 (0-0) | 0 (0-4) | 8 (2-20) | 6 (3.75-13) |
| **Overall** | 4 (1-10) | 12 (7-21) | 5 (2-9) | 10 (6-15) | 3 (0-6) | 10 (3-21) | 11 (6-19) |
| **Estimates obtained by fitting a negative binomial distribution to observed data** | | | | | | | |
| **overdispersion** | 0.445 | 1.638 | 0.831 | 2.036 | 0.57 | 0.746 | 1.517 |
| **mean** | 9.753 | 16.105 | 7.385 | 12.159 | 5.309 | 15.54 | 14.614 |

TABLE 2.3: Time intervals between key events as estimated from laboratory confirmed infections ascertained in Lombardy between February 20 and July 16, 2020

## 2.4 Discussion and conclusions

In this work, we provided a comprehensive assessment of the parameters regulating COVID-19 burden and natural history. The proposed analysis leveraged data on the infections ascertained in Italy between February and July 2020 to estimate the time between key events and the age- and sex- specific stage-to-stage transition probabilities characterizing the clinical progression of COVID-19.

Previous studies have highlighted that a significant share of SARS-CoV-2 infections is represented by symptom-free subjects and by individuals developing mild disease (Emery et al., 2020; Lavezzo et al., 2020; Poletti et al., 2021;Salje H, 2020; Wu et al., 2020). Therefore, using the number of notified or confirmed COVID-19 cases as an approximation of the number of infections would likely lead to overestimate the risk of disease and severe outcomes, undermining the comparability and generalizability of the obtained results. An illustrative example of the huge uncertainty caused by this phenomenon is provided by the high variability around the available estimates of the proportion of symptomatic infections, ranging from 3% to 87% (Buitrago-Garcia et al., 2020; Byambasuren et al., 2020; Emery et al., 2020; Nikolai et al., 2020; Oran and Topol, 2020; Poletti et al., 2021). To reduce potential biases in

the identification of SARS-CoV-2 infections, we estimated different risk ratios based on a sample of SARS-CoV-2 positive individuals who were identified as contacts of confirmed cases and tested irrespectively of their symptoms. A larger sample, consisting of all notified symptomatic cases, was used only to estimate the time to key events describing the clinical progression of cases and to highlight temporal changes in the risk of hospitalization and ICU admission. Our results confirmed findings from other studies on the strong age gradient in the likelihood of developing symptoms, critical disease, and death after infection (Onder, Rezza, and Brusaferro, 2020; Poletti et al., 2020a; Salje H, 2020; Verity et al., 2020; Yang et al., 2020). The estimated proportion of symptomatic cases among SARS-CoV-2 infections is within the range of estimates obtained in previous studies (Buitrago-Garcia et al., 2020, Nikolai et al., 2020) and particularly close to findings obtained in Emery et al., 2020. Our estimated CFR was lower compared to the one obtained in a previous study on other Italian data (Onder, Rezza, and Brusaferro, 2020), but slightly higher than those observed in other countries (Fu et al., 2020; Li et al., 2020; Verity et al., 2020; Yang et al., 2020). The estimated age-profile of the IFR closely resembles Verity et al., 2020. However, our aggregate (population-level) estimate of the IFR is generally higher than those obtained in other studies (O'Driscoll et al., 2021; Perez-Saez et al., 2021; Salje H, 2020; Verity et al., 2020). Such difference can be due to a variety of factors. First, Italy is one of the oldest countries in the world (average age: 45.7 years (*Italian National Institute of Statistics. Demographic indicators.*). Second, there may be between-country differences in the age distribution of SARS-CoV-2 infected individuals. Third, there are differences in the definition of COVID-19 death. In fact, in Italy, deaths occurring among SARS-CoV-2 positive subjects are classified as COVID-19-related deaths regardless of other conditions that might have caused the observed fatal outcome (Onder, Rezza, and Brusaferro, 2020). This has possibly led to overestimate the number of deaths caused by SARS-CoV-2, especially in the oldest segment of the population. Nonetheless, in Italy, a laboratory confirmation for SARS-CoV-2 infection is required to define a COVID-19 death. COVID-19 deaths are mainly represented by cases ascertained before their decease, while only few cases are ascertained post-mortem. In the subsample of positive case contacts used to estimate the IFR and CFR, all deaths were confirmed before their decease.

The proportion of hospitalized patients among positive individuals older than 60 years was almost double than that observed in France (Salje H, 2020). On the other hand, the proportion of ICU admissions and deaths among hospitalized cases was markedly lower in our sample (4.5% and 12.4% vs 19% and 18.1%, respectively), and we found a strong temporal decreasing pattern in the risk of hospital admission among ascertained symptomatic cases. This suggests that hospitalization criteria might have been highly heterogeneous across different countries and may also greatly vary over time.

The estimated time from symptom onset to laboratory diagnosis well compares with estimates obtained from Belgian patients (Faes et al., 2020). Although in line with previous findings from Belgium (Faes et al., 2020), the time from symptom onset to hospital admission we found is markedly shorter than those observed in France, in China and in the US (Bhatraju et al., 2020; Salje H, 2020; Zhou et al., 2020). This may be the consequence of the higher proportion of severe cases observed in Italy compared to other countries, which strictly relates to the older age-structure characterizing this country. This hypothesis is partially supported by the shorter hospital

length of stay and by the longer length of stay in ICU we found, compared to estimates from China (Guan et al., 2020; Zhou et al., 2020).
The relatively low ICU admission ratio we observed among the elderly was already highlighted in previous studies (Grasselli et al., 2020; Salje H, 2020). However, our findings clearly show that hospitalized patients aged 80 years or older faced the highest risk of fatal outcome, but also the lowest likelihood of being admitted to ICU. The increased ratio of ICU admission among inpatients we found for this age group after April 2020 suggests that elective ICU admission has been initially adopted in Lombardy due to saturation of healthcare resources. The lower delay between symptom onset and admission to hospital observed after March 16, 2020, and the progressive temporal increase in the likelihood of hospital admission among older patients strongly suggest that reducing the pressure on the regional healthcare system markedly improved its capacity to rapidly identify and treat severe patients (Trentini et al., 2020) (see Figure 2.5).
Our estimates of the risk ratios of hospital and ICU admissions after infection should be interpreted cautiously. In fact, rather than being purely biological features, such quantities strongly depend on the available healthcare resources, on the temporal changes in the number of patients seeking care, and on the protocols adopted to face a brisk upsurge of COVID-19 cases. Consequently, using these estimates to investigate the healthcare burden over different phases of the pandemic could produce misleading results. Additionally, due to temporal changes in the ascertainment rates of infections, we were not able to quantify the reduced risk of severe outcomes determined by timely detection, diagnosis and treatment of cases, nor to evaluate the role played by the progressive enhancement in the treatment procedures in reducing the risk of disease. A further limitation affecting our study is the lack of data to disentangle the role played by patients' comorbidities in shaping the risk of severe diseases. Finally, it is important to stress that estimates reported here are associated with the historical and dominant variant of the virus that circulated during 2020, in the absence of vaccination. As such, estimated metrics may not apply to new emerging SARS-CoV-2 variants (Davies et al., 2021; Kiem et al., 2020; Volz et al., 2021) and may not reflect the risk of developing COVID-19 disease among infections occurring among vaccinated individuals. Although disease parameters may be specific for the time and place of the data collection (northern Italy's first COVID-19 wave), we showed that estimated risk outcomes after SARS-CoV-2 infection well compare with data associated with broader time periods and geographical locations.

Metrics defining the natural history of SARS-CoV-2 infection were estimated from positive individuals who belonged to clusters of contacts, who were all tested and daily followed up for symptoms and for severe outcomes. A fraction of these individuals, mainly consisting of symptomatic ones, was tested via RT-PCR during contact-tracing activities. The remaining case contacts were retrospectively tested via IgG serological assays collected at least one month after exposure, thus allowing the identification of asymptomatic infections as well. Despite the heterogeneous testing procedure, we believe that the strengths of this study design rely on: (1) the minimization of the risks of bias in the identification of infections (contacts were identified and tested independently of their clinical signs), and (2) the daily follow-up of the infections for symptoms and critical disease in the weeks following the exposure to a confirmed infection. Therefore, the analyzed sample does not suffer

the usual limitations of surveillance data (i.e., underestimation of asymptomatic individuals) and of serological data (i.e., lack of longitudinal records about the clinical history of study participants). Despite the aforementioned limitations, the provided metrics can be instrumental to refine model estimates. In particular, our findings could be used to assist the design and evaluation of forthcoming vaccination efforts and the development of appropriate strategies to control the COVID-19 pandemic until a sufficiently large proportion of the population has become immune.

## 2.5 Supplementary Figures and Tables

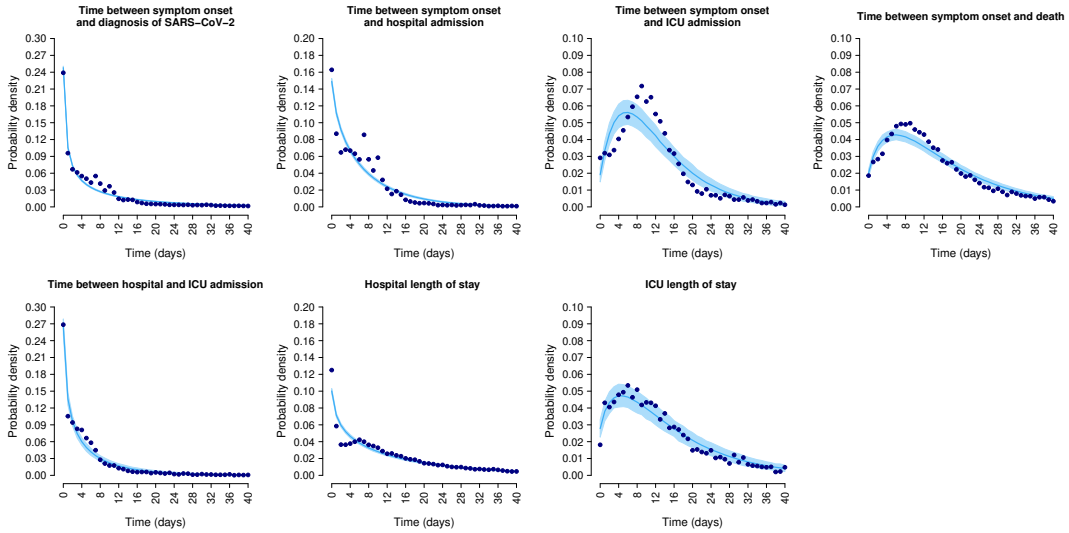


FIGURE 2.4: Estimated and observed distributions of time intervals between key events. (Top row) Blue dots represent the observed distribution of time from symptom onset to diagnosis, hospitalization, ICU admission and death. Light blue lines show the mean frequencies obtained by simulating 1,000 different datasets with size equal to the number of observations in the data on the basis of a negative binomial model. Shaded areas represent the corresponding 95% prediction interval. (Bottom row) As for the top row, but for the time between hospital and ICU admission, for the length of stay in hospital and ICU.

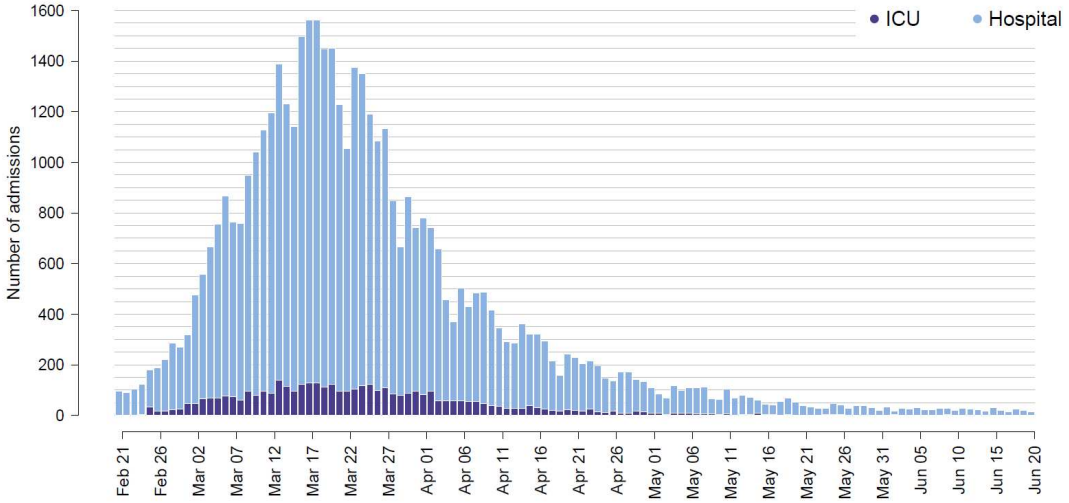


FIGURE 2.5: Number of hospital (light blue) and ICU (dark blue) admissions of COVID-19 cases in Lombardy, Italy between February 21 and June 20, 2020.

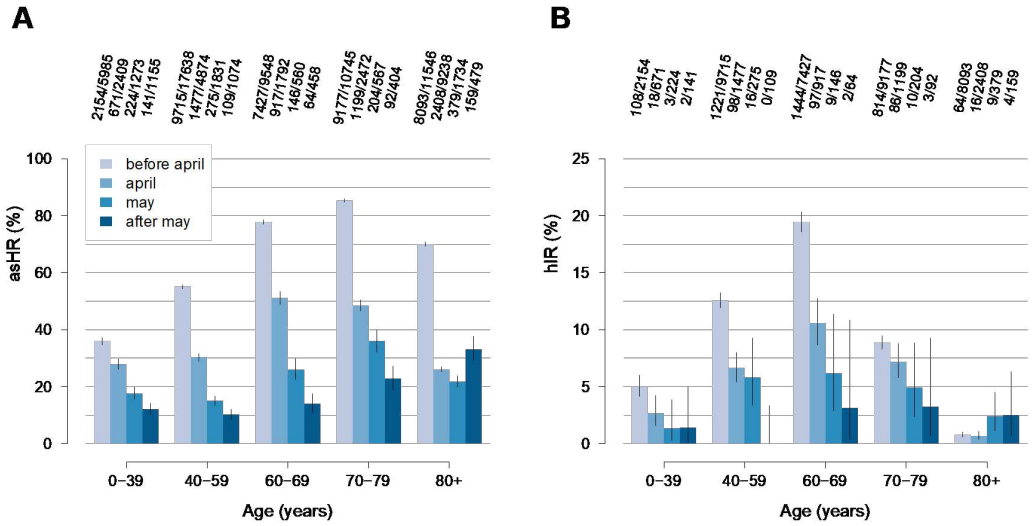


FIGURE 2.6: **A** Age-specific case hospital admission ratios among ascertained symptomatic cases (asHR) as estimated by excluding patients with a delay from SARS-CoV-2 diagnosis to hospital or ICU admission greater than 30 days. **B** As A but for the age-specific ICU admission ratios among hospitalized cases (hIR). Bars of different colors represent crude percentages observed across different epidemic periods; vertical lines represent 95% confidence intervals computed by exact binomial tests. Numbers shown in each panel represent the age-specific number of events observed in the data among exposed COVID-19 cases.

| | Positives contacts | Symptomatic cases | | Critical cases | | Deaths | | Hospitalized patients | | ICU-admitted patients | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Count | Count | Risk ratio (95%CI) | Count | Risk ratio (95%CI) | Count | Risk ratio (95%CI) | Count | Risk ratio (95%CI) | Count | Risk ratio (95%CI) |
| **Age** | | | | | | | | | | | |
| **≥ 75** | 215 | 118 | Reference | 26 | Reference | 26 | Reference | 75 | Reference | 2 | Reference |
| **0-59** | 1,280 | 357 | 0.51 (0.41-0.62) | 7 | 0.04 (0.02-0.09) | 2 | 0.01 (0.00-0.04) | 113 | 0.24 (0.18-0.33) | 5 | 0.39 (0.08-2.67) |
| **60-74** | 470 | 155 | 0.60 (0.48-0.74) | 10 | 0.18 (0.08-0.36) | 7 | 0.12 (0.05-0.27) | 78 | 0.48 (0.35-0.65) | 5 | 1.19 (0.26-7.86) |
| **Sex** | | | | | | | | | | | |
| **Female** | 1,111 | 365 | Reference | 19 | Reference | 16 | Reference | 139 | Reference | 5 | Reference |
| **Male** | 854 | 265 | 0.97 (0.85-1.11) | 24 | 2.16 (1.17-3.98) | 19 | 2.15 (1.08-4.27) | 127 | 1.34 (1.07-1.67) | 7 | 2.04 (0.64-6.86) |
| **Epidemic period** | | | | | | | | | | | |
| **March** | 1,500 | 467 | Reference | 35 | Reference | 30 | Reference | 200 | Reference | 8 | Reference |
| **April** | 465 | 163 | 1.14 (0.98-1.31) | 8 | 0.71 (0.30-1.47) | 5 | 0.50 (0.17-1.22) | 66 | 1.07 (0.81-1.38) | 4 | 1.60 (0.43-5.04) |

TABLE 2.4: Estimated risk ratios of developing symptoms, hospital and ICU admission, experiencing critical disease, and fatal outcome among positive contacts of confirmed cases as identified during contact tracing operations. Results are disaggregated by age, sex, and period.

| | Hospitalized patients | ICU-admitted patients | | Deaths | |
|---|---|---|---|---|---|
| | Count | Count | Risk ratio (95%CI) | Count | Risk ratio (95%CI) |
| **Age** | | | | | |
| **≥ 80** | 11,849 | 110 | Reference | 6,370 | Reference |
| **0-39** | 3,361 | 136 | 4.49 (3.51 - 5.73) | 31 | 0.02 (0.01 - 0.02) |
| **40-59** | 12,037 | 1,397 | 11.33 (9.53 - 13.55) | 681 | 0.09 (0.08 - 0.10) |
| **60-69** | 8,917 | 1,615 | 16.99 (14.43 - 20.07) | 1,740 | 0.30 (0.29 - 0.32) |
| **70-79** | 11,228 | 965 | 8.03 (6.70 - 9.70) | 4,323 | 0.64 (0.62 - 0.66) |
| **Unknown** | 1 | 1* | - | 0* | - |
| **Sex** | | | | | |
| **Female** | 19,318 | 891 | Reference | 4,586 | Reference |
| **Male** | 28,061 | 3,324 | 2.19 (2.04 – 2.35) | 8,557 | 1.46 (1.41 – 1.50) |
| **Unknown** | 14 | 9 | 14.79 (8.60 – 19.19) | 2 | 2.59 (0.78 – 3.69) |
| **Epidemic period** | | | | | |
| **Before April** | 37,391 | 3,757 | Reference | 10,886 | Reference |
| **April** | 6,909 | 327 | 0.63 (0.56 – 0.71) | 1,708 | 0.72 (0.68 – 0.76) |
| **May** | 1,282 | 50 | 0.53 (0.39 – 0.69) | 158 | 0.36 (0.30 – 0.42) |
| **After May** | 591 | 11 | 0.25 (0.13 – 0.44) | 30 | 0.15 (0.10 – 0.21) |
| **Unknown** | 1,220 | 79 | 0.74 (0.58 – 0.92) | 363 | 0.90 (0.81 – 1.00) |

TABLE 2.5: Estimated risk ratios of ICU admission and fatal outcome among hospitalized patients, disaggregated by age, sex, and period.

* RR and 95%CI were not computed for insufficiently large sample size.

| | Positives contacts | Symptomatic cases | | Critical cases | | Deaths | | | Hospitalized patients | | ICU-admitted patients | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Count | Count | Proportion (95%CI) | Count | Proportion (95%CI) | Count | IFR (95%CI) | CFR (95%CI) | Count | Proportion (95%CI) | Count | Proportion (95%CI) |
| **Age** | | | | | | | | | | | | |
| **0-14** | 142 | 27 | 19% (12.9-26.4%) | 0 | 0% (0-2.6%) | 0 | 0% (0-2.6%) | 0% (0-12.8%) | 3 | 2.1% (0.4-6%) | 0 | 0% (0-2.6%) |
| **15-19** | 21 | 5 | 23.8% (8.2-47.2%) | 0 | 0% (0-16.1%) | 0 | 0% (0-16.1%) | 0% (0-52.2%) | 1 | 4.8% (0.1-23.8%) | 0 | 0% (0-16.1%) |
| **20-39** | 215 | 58 | 27% (21.2-33.4%) | 2 | 0.9% (0.1-3.3%) | 0 | 0% (0-1.7%) | 0% (0-6.2%) | 14 | 6.5% (3.6-10.7%) | 2 | 0.9% (0.1-3.3%) |
| **40-59** | 360 | 128 | 35.6% (30.6-40.7%) | 2 | 0.6% (0.1-2%) | 1 | 0.3% (0-1.5%) | 0.8% (0-4.3%) | 53 | 14.7% (11.2-18.8%) | 1 | 0.3% (0-1.5%) |
| **60-69** | 186 | 66 | 35.5% (28.6-42.8%) | 2 | 1.1% (0.1-3.8%) | 1 | 0.5% (0-3%) | 1.5% (0-8.2%) | 32 | 17.2% (12.1-23.4%) | 1 | 0.5% (0-3%) |
| **70-79** | 121 | 46 | 38% (29.3-47.3%) | 6 | 5% (1.8-10.5%) | 6 | 5% (1.8-10.5%) | 13% (4.9-26.3%) | 25 | 20.7% (13.8-29%) | 2 | 1.7% (0.2-5.8%) |
| **≥ 80** | 73 | 46 | 63% (50.9-74%) | 5 | 6.8% (2.3-15.3%) | 5 | 6.8% (2.3-15.3%) | 10.9% (3.6-23.6%) | 26 | 35.6% (24.7-47.7%) | 0 | 0% (0-4.9%) |
| **Sex** | | | | | | | | | | | | |
| **Female** | 589 | 200 | 34% (30.1-37.9%) | 5 | 0.8% (0.3-2%) | 4 | 0.7% (0.2-1.7%) | 2% (0.5-5%) | 73 | 12.4% (9.8-15.3%) | 2 | 0.3% (0-1.2%) |
| **Male** | 529 | 176 | 33.3% (29.3-37.5%) | 12 | 2.3% (1.2-3.9%) | 9 | 1.7% (0.8-3.2%) | 5.1% (2.4-9.5%) | 81 | 15.3% (12.3-18.7%) | 4 | 0.8% (0.2-1.9%) |

TABLE 2.6: Risk metrics estimated from SARS-CoV-2 positive contacts identified after 20 March 2020.

| | Positives contacts | Symptomatic cases | | Critical cases | | Deaths | | | Hospitalized patients | | ICU-admitted patients | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Count | Count | Proportion (95%CI) | Count | Proportion (95%CI) | Count | IFR (95%CI) | CFR (95%CI) | Count | Proportion (95%CI) | Count | Proportion (95%CI) |
| **Age** | | | | | | | | | | | | |
| **0-14** | 219 | 39 | 17.8% (13-23.5%) | 0 | 0% (0-1.7%) | 0 | 0% (0-1.7%) | 0% (0-9%) | 4 | 1.8% (0.5-4.6%) | 0 | 0% (0-1.7%) |
| **15-19** | 22 | 6 | 27.3% (10.7-50.2%) | 0 | 0% (0-15.4%) | 0 | 0% (0-15.4%) | 0% (0-45.9%) | 2 | 9.1% (1.1-29.2%) | 0 | 0% (0-15.4%) |
| **20-39** | 377 | 99 | 26.3% (21.9-31%) | 2 | 0.5% (0.1-1.9%) | 0 | 0% (0-1%) | 0% (0-3.7%) | 18 | 4.8% (2.9-7.4%) | 2 | 0.5% (0.1-1.9%) |
| **40-59** | 662 | 213 | 32.2% (28.6-35.9%) | 5 | 0.8% (0.2-1.8%) | 2 | 0.3% (0-1.1%) | 0.9% (0.1-3.4%) | 89 | 13.4% (10.9-16.3%) | 3 | 0.5% (0.1-1.3%) |
| **60-69** | 330 | 105 | 31.8% (26.8-37.1%) | 5 | 1.5% (0.5-3.5%) | 3 | 0.9% (0.2-2.6%) | 2.9% (0.6-8.1%) | 48 | 14.5% (10.9-18.8%) | 3 | 0.9% (0.2-2.6%) |
| **70-79** | 240 | 94 | 39.2% (33-45.7%) | 17 | 7.1% (4.2-11.1%) | 16 | 6.7% (3.9-10.6%) | 17% (10.1-26.2%) | 59 | 24.6% (19.3-30.5%) | 4 | 1.7% (0.5-4.2%) |
| **≥ 80** | 112 | 71 | 63.4% (53.8-72.3%) | 14 | 12.5% (7-20.1%) | 14 | 12.5% (7-20.1%) | 19.7% (11.2-30.9%) | 43 | 38.4% (29.4-48.1%) | 0 | 0% (0-3.2%) |
| **Sex** | | | | | | | | | | | | |
| **Female** | 1,109 | 363 | 32.7% (30-35.6%) | 19 | 1.7% (1-2.7%) | 16 | 1.4% (0.8-2.3%) | 4.4% (2.5-7.1%) | 137 | 12.4% (10.5-14.4%) | 5 | 0.5% (0.1-1%) |
| **Male** | 853 | 264 | 30.9% (27.9-34.2%) | 24 | 2.8% (1.8-4.2%) | 19 | 2.2% (1.3-3.5%) | 7.2% (4.4-11%) | 126 | 14.8% (12.5-17.3%) | 7 | 0.8% (0.3-1.7%) |

TABLE 2.7: Risk metrics estimated from SARS-CoV-2 positive contacts as obtained by excluding patients with a delay from SARS-CoV-2 diagnosis to hospital or ICU admission greater than 30 days.

| | Median days (IQR) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Time between symptom onset and diagnosis of SARS-CoV-2 | Time between symptom onset and death | Time between symptom onset and hospital admission | Time between symptom onset and ICU admission | Time between hospital and ICU admission | Hospital length of stay | ICU length of stay |
| **Age** | | | | | | | |
| **0-39** | 3 (0-11) | 16 (6-27.25) | 4 (1-8) | 8 (4-11) | 1 (0-5) | 4 (0-10) | 9 (4-16) |
| **40-59** | 5 (1-11) | 16 (9-26) | 6 (2-10) | 10 (6-13) | 2 (0-5) | 9 (1-18) | 12 (6-19) |
| **60-69** | 5 (1-10) | 15 (9-25) | 6 (2-10) | 10 (7-15) | 3 (0-6) | 13 (6-24) | 12 (6-21) |
| **70-79** | 5 (1-9) | 12 (7-20) | 5 (2-9) | 10 (6-15) | 3 (1-7) | 12 (5-24) | 10 (5-18) |
| **≥ 80** | 3 (0-8) | 11 (6-19) | 4 (1-8) | 7 (3-15.25) | 1 (0-5) | 10 (4-24) | 5 (3-11) |
| **Unknown** | 13 (9.5-16.5) | 0 (0-0) | 8 (8-8) | 8 (8-8) | 0 (0-0) | 14 (14-14) | 14 (14-14) |
| **Epidemic period** | | | | | | | |
| **Before 16 March** | 7 (3-12) | 13 (8-21) | 7 (3-10) | 10 (7-15) | 3 (0-6) | 10 (3-20) | 11 (6-19) |
| **After 16 March** | 2 (0-7) | 11 (6-20) | 3 (1-7) | 9 (5-14) | 3 (0-7) | 10 (3-23) | 11 (6-20) |
| **Unknown** | 0 (0-0) | 0 (0-0) | 0 (0-0) | 0 (0-0) | 0 (0-4) | 8 (2-20) | 7 (4-13.5) |
| **Overall** | 4 (1-10) | 12 (7-20) | 5 (1-9) | 10 (6-14) | 3 (0-6) | 10 (3-21) | 11 (6-19) |

TABLE 2.8: Time intervals between key events as estimated when excluding patients with a delay from SARS-CoV-2 diagnosis to hospital or ICU admission greater than 30 days.

# Chapter 3

# Modeling the interplay between demography, social contact patterns and SARS-CoV-2 transmission in the South West Shewa Zone of Oromia Region, Ethiopia

## 3.1 Background

Despite limited access to healthcare (Gilbert et al., 2020; Poletti et al., 2018) and relatively milder social distancing restrictions compared to those imposed in most high-income countries (Loembé et al., 2020; *International Monetary Fund. Policy responses to Covid-19. 2020.*), coronavirus disease 2019 (COVID-19) mortality rates have been relatively low throughout Africa (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*). As of January 24 2021, the World Health Organization (WHO) reports 2,462,083 diagnosed cases and 57,902 deaths in the continent (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*). However, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission dynamics have been highly heterogeneous across different African countries in terms of timing and implemented interventions (Makoni, 2020).

In sub-Saharan Africa, Ethiopia is second only to South Africa in terms of number of recorded cases and deaths, with an overall case fatality ratio (CFR) of about 1.5% compared to about 2.2% in the rest of the world (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*). The first COVID-19 case was confirmed on March 13, 2020 and, less than a month later, the Ethiopian Prime Minister declared a state of emergency in the country on April 8, 2020 (*World Health Organization. Covid-19 Response Bullettin Ethiopia. 2020.*). Since then, rigorous contact tracing, isolation, and compulsory quarantine have been established (Mohammed et al., 2020; *Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 37 weekly bulletin.*). Borders and school closure were implemented, public institutions and firms operated at minimum capacity or under complete closure, and people were advised to stay at home (Mohammed et al., 2020). However, in November 2020, schools reopened in the entire country, and social gatherings were allowed again. As of January 24, 2021, 133,298 SARS-CoV-2 infections and 2,063 deaths (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard.*

*2020.*) were ascertained in the entire country, with thousands of cases reported in all the 12 regions of Ethiopia (*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 37 weekly bulletin.*). In Ethiopia, a syndromic surveillance is carried out to identify SARS-CoV-2 infected individuals. Samples from suspected cases and case contacts are collected at different health facilities displaced in the country (including health centers serving the most rural areas) and cases are confirmed via real-time reverse transcription–polymerase chain reaction (RT-PCR) test. Collected samples are analyzed by 38 national, regional, hospital, and private laboratories (*UNICEF. Ethiopia COVID-19 Situation Report No. 3. 2020.*). Both suspected and laboratory confirmed cases are admitted to isolation centers and discharged after a negative laboratory test (*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 37 weekly bulletin.*). Although swab testing was initially applied to both symptomatic patients and all close contacts of cases, it is possible that, due to limited resources and the increased number of cases in the country, only symptomatic case contacts are currently tested. Active monitoring of cases conducted by Ethiopian Public Health Institute suggested that 52% of the identified positive cases were asymptomatic (*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 6 weekly bulletin.*). As of January 10, 2021, the overall rate for positive laboratory test results since the first detection of the epidemic in the country was 6.9% (*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 37 weekly bulletin.*).

The possible spread of SARS-CoV-2 in rural areas of the country is especially dangerous because of the sparse presence of well-resourced health facilities implying long travel distances for remote populations, which is an important barrier to universal access to primary care (Poletti et al., 2018). Moreover, the healthcare workforce in Ethiopia is 5 times lower than the minimum threshold defined by the WHO for Sustainable Development Goals health targets (*World Health organization. Health Workforce Requirements for Universal Health Coverage and the Sustainable Development Goals. Human Resource for Health Observers Series No. 17.*), and far below the African average (Haileamlak, 2018). Recent modeling studies investigated the impact of control measures, such as self-isolation and temporary lockdowns, in a number of sub-Saharan African countries, highlighting the difficulties in defining effective, feasible and sustainable strategies for suppression or mitigation of COVID-19 epidemics (Van Zandvoort et al., 2020; Quaife et al., 2020; Brand et al., 2020; Walker et al., 2020). In this work, we aim to assess how demographic factors and age-specific mixing patterns can influence the impact of COVID-19 epidemics across different geographical contexts of the South West Shewa Zone (SWSZ) of the Oromia Region of Ethiopia, characterized by different levels of access to healthcare. So far, 21,133 cases were reported in the Oromia Region. The interventions implemented to control the epidemic were part of the national strategy designed by Ministry of Health targeting all districts of the country, including the SWSZ. National measures undertaken between April and mid-September 2020 included the suspension of teaching activities at schools and universities. More stringent measures, including interruption of economic activities, restrictions on the use of public transport and social gatherings (churches, mosques, markets etc.) were partially adopted as well (Mohammed et al., 2020).

## 3.2 Method

### 3.2.1 Study desing

We conducted a survey based on individual interviews to estimate age-specific mixing patterns in four districts (woreda) of the SWSZ. About 40% of the SWSZ population is below 15 years of age and about 68% lives in remote rural settlements, 18% in rural villages, and 14% in the largest town of the area (Woliso Town, 53,065 inhabitants). The districts targeted by our study encompass a population of 449,460 inhabitants and represent the main catchment area of the St. Luke Hospital located in Woliso Town, a well-resourced health facility acting as the referral hospital for the entire Zone (Poletti et al., 2018). The study consists in a cross-sectional survey with two-stage stratified random sampling by location and age group. The survey was conducted in eight different sites, choosing two neighborhoods (kebele) for each district under study, in such a way to capture contact patterns in areas characterized by different population densities, work and travel opportunities, and access to the healthcare infrastructure. Three types of geographical contexts were considered: remote settlements (consisting of scattered subsistence farming settlements), rural villages (consisting of concentrated clusters of households served by a main road, and better access to main public services), and urban neighborhoods inside Woliso Town (significantly higher population density and full access to public services (*United States Department of Agriculture. Economic Research Service.*)).

For each site, a target sample size of 105 study participants was set on the basis of findings from previous contact surveys (Melegaro et al., 2017; Waroux et al., 2018) to provide the desired precision in the mean number of contacts (see Appendix: Sections 1 and 2 (Waroux et al., 2018; Cohen, 2013; Horby et al., 2011)). Households and study participants were randomly sampled using predefined quotas for each site, sex, and age group. A household was defined as a group of individuals living under the same roof and sharing the same kitchen on a daily basis. One individual per household was interviewed. If the study participant was a student, additional shorter interviews were performed to complement the data with information about close contacts occurring at school.

### 3.2.2 Data collection

Participants were asked to recall information on the frequency, location, type of social encounters from the day preceding their interview, providing the age (or age range when exact age was unknown) and their relationship for each listed contact. A contact was defined as an interaction between two individuals, either physical (when involving skin-to-skin contact), or non-physical (when involving a two-way conversation with five or more words in the physical presence of another person, but no skin-to-skin contact) (Melegaro et al., 2017; Waroux et al., 2018). The participants' age, sex, education and occupational status were recorded along with details on their household composition. In the SWSZ, schools may host up to 100 students within a single class. To avoid inaccurate reporting of the number of school contacts, participants were only asked to count the total number of physical contacts they had at school in the previous day, without further details. Information on the age of

students attending the targeted schools for different grades was also collected. Interviews were carried out between November and December 2019, i.e. prior to the COVID-19 pandemic. Schools were regularly open during the survey period.

### 3.2.3 Contact patterns and data analysis

For each type of geographical context, we computed the mean number of contacts reported by respondents after grouping by age (six 10-year age groups from 0 to 59 years and one age group for individuals aged 60 years or older) and by contact setting (households, schools, and the general community). Since for many study participants it was difficult to distinguish encounters occurred because of their job from other random contacts, all social interactions occurring outside family and schools were aggregated with contacts occurring in the general community. Age-specific contact matrices were computed considering both physical and non-physical contacts and were adjusted for reciprocity as in (Melegaro et al., 2017). Variability due to sampling of study participants was explored by computing 1,000 bootstrapped contact matrices (Zhang et al., 2020), where each bootstrap consisted in sampling with replacement a number of interviews equal to the original sample size, choosing the age of the participant with probability proportional to the Ethiopian age distribution (*United Nations Department of Economic and Social Affairs. 2019 UN World Population Prospects.*). The proportions of the SWSZ population living in remote settlements, rural villages and in urban neighborhoods were used as sampling weights to compute an average contact matrix for the entire SWSZ. Full details about the study design, data collection and the analysis of contact patterns are provided in the Appendix: Sections 1-7.

### 3.2.4 Transmission model

We simulated SARS-CoV-2 spread in the SWSZ, using an age-structured Susceptible-Infectious- Recovered (SIR) compartmental model with three consecutive stages of infectiousness, in such a way to reproduce a gamma-distributed generation time of mean 6.6 days (Cereda et al., 2020; Marziano et al., 2021a; Guzzetta et al., 2020). The model was run separately for each geographical context (i.e. the remote, rural and urban neighborhoods), using estimates of the population age structure and of the age-specific contact matrix computed from survey data (see Appendix: Sections 4-6). These data were collected in the absence of any restrictions imposed to control the infection spread. Because school closure in all of Ethiopia was mandated much before the exponential growth of reported COVID-19 cases, transmission of SARS-CoV-2 in the SWSZ was simulated by removing contacts occurring at school and considering only household and community contacts. In the model, 1,000 values of the per-contact transmission rate were considered by matching the reproduction number computed through the next-generation matrix approach (Diekmann, Heesterbeek, and Metz, 1990) with random samples from the posterior distribution of the reproduction number estimated from the curve of reported cases in Ethiopia during the phase of exponential growth (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.* Wallinga and Lipsitch, 2007). As the same public measures and restrictions were applied across different geographical contexts in Ethiopia, heterogeneous transmission of SARS-CoV-2 was assumed to be driven by differences in the demographic and contact structures in urban, rural and

remote neighborhoods. The same per-contact transmission rate was therefore assumed across different settings of the SWSZ and estimated using the sum of contact matrices obtained for the urban, rural and remote neighborhoods, weighted by the percentage of SWSZ population living in each geographical context. We included school contacts to estimate the theoretical SARS-CoV-2 transmission potential in the absence of a school closure mandate. We considered susceptibility to SARS-CoV-2 infection to vary with age. We adopted the posterior distributions estimated in Zhang et al. (Zhang et al., 2020) for the relative probability of developing infection upon effective exposure to an infectious case, where the age-group 15-64 years is taken as a reference; an average relative susceptibility of 0.33 (95%CI: 0.24-0.47) was considered for children under 15 years of age, and of 1.47 (95%CI: 1.16-2.06) for older adults (above 65 years) (Zhang et al., 2020). These estimates are aligned with other independent studies (reviewed in Viner et al. (Viner et al., 2021)). We assumed the same infectiousness across individuals of different ages (see Appendix: Section 4 (Lavezzo et al., 2020)). We computed projections of the number of SARS-CoV-2 infections, cases with respiratory symptoms or fever, and COVID-19 critical cases (either requiring mechanical ventilation or resulting in a fatal outcome), based on available estimates of the age-specific risks (Poletti et al., 2020b). By comparing estimates obtained when including and excluding school contacts for the entire duration of the epidemic, we computed the overall percentage of infections, symptomatic and critical cases that could be averted by school closure. To explore the robustness of our findings with respect to model assumptions, five separate sensitivity analyses were carried out assuming: 1) a Susceptible-Exposed-Infectious-Recovered (SEIR) model structure; 2) a 20% increase or a 20% decrease of the net reproduction number; 3) different per-contact transmission rates across geographical settings; 4) homogeneous susceptibility by age; 5) a lower infectiousness of children (see Appendix: Section 8). As the probability of developing symptoms after infection markedly increases with age (Poletti et al., 2020b; Buitrago-Garcia et al., 2020), the latter sensitivity is similar to exploring the effect of differential infectiousness among symptomatic and asymptomatic cases.

## 3.3 Results

### 3.3.1 Social contact data

A total of 938 study participants were interviewed with 43% of them living in rural remote settlements, 35% in rural villages, and 22% from urban neighborhoods (Table 3.1). 227 participants were students, 22.9% of whom were between 5 and 9 years of age, 71.8% between 10 and 19 years, and 4.9% older. School attendance rates among the study participants aged 5-18 years was 67%, 80% and 77% in remote, rural and urban sites, respectively. The median class size ranged from 70 children per class in rural villages to 90 in remote settlements. Only 27% of our study participants reported travels outside their village in the last month; 87.3% reported they were never admitted to the local hospital (see Appendix: Section 7).

| | Number of study participants | | | | |
|---|---|---|---|---|---|
| VARIABLE* | OVERALL n (%) | REMOTE n (%) | RURAL n (%) | URBAN n (%) | ETHIOPIA (%) [24] |
| **Total** | | | | | |
| | 938 (100.0) | 400 (42.6) | 326 (34.8) | 212 (22.6) | - |
| **Age** | | | | | |
| <10y | 382 (40.7) | 160 (40) | 137 (42) | 85 (40.1) | 27.3 |
| 10-19y | 198 (21.1) | 85 (21.2) | 66 (20.2) | 47 (22.2) | 24.1 |
| 20-29y | 92 (9.8) | 40 (10) | 32 (9.8) | 20 (9.4) | 18.4 |
| 30-39y | 117 (12.5) | 50 (12.5) | 42 (12.9) | 25 (11.8) | 12.0 |
| 40-49y | 59 (6.3) | 26 (6.5) | 18 (5.5) | 15 (7.1) | 7.9 |
| 50-59y | 40 (4.3) | 17 (4.2) | 13 (4) | 10 (4.7) | 4.9 |
| 60y + | 50 (5.3) | 22 (5.5) | 18 (5.5) | 10 (4.7) | 5.3 |
| **Occupation** | | | | | |
| Pre-school | 309 (32.9) | 129 (32.2) | 109 (33.4) | 71 (33.5) | - |
| Student | 226 (24.1) | 85 (21.2) | 87 (26.7) | 54 (25.5) | - |
| Manual/Office/Shop worker | 62 (6.6) | 5 (1.2) | 30 (9.2) | 27 (12.7) | - |
| Housewife | 137 (14.6) | 66 (16.5) | 47 (14.4) | 24 (11.3) | - |
| Agriculture** | 112 (11.9) | 84 (21) | 25 (7.7) | 3 (1.4) | - |
| Unemployed/Retired | 44 (4.7) | 9 (2.3) | 12 (3.7) | 23 (10.8) | - |
| Other | 48 (5.1) | 22 (5.5) | 16 (4.9) | 10 (4.7) | - |
| **Sex** | | | | | |
| Female | 478 (51) | 206 (51.5) | 170 (52.1) | 102 (48.1) | 50.0 |
| Male | 460 (49) | 194 (48.5) | 156 (47.9) | 110 (51.9) | 50.0 |

TABLE 3.1: Characteristics of study participants and relative percentages in the Ethiopian population.

* No missing data for any of the three listed variables.

** The percentage of male adults (18-64yo) working in agriculture is 45.2%; in the remote, the rural and the urban settings this percentage is 81%, 28% and 7%, respectively

Age and sex were also recorded for all the 4,635 household members of the 938 study participants. The mean household size in remote settlements was 5.5 (95% CI: 5.3-5.7), significantly larger (Tukey test $p<0.001$) than in rural villages (4.6, 95% CI: 4.4-4.8) and in urban neighborhoods (4.4, 95%CI: 4.2-4.6), while no significant difference in the household size was found between the latter two settings (Tukey test $p=0.48$). Overall, 5,690 non-school contacts were reported by the 938 study participants (median 6 contacts per person, range 1-26, see Table 3.2). Of these, 79.9% were physical and 43.0% involved a single social interaction during the day.

Mean number of contacts per day (excluding school contacts)

| VARIABLE | | OVERALL mean (95% CI) | REMOTE mean (95% CI) | RURAL mean (95% CI) | URBAN mean (95% CI) |
|---|---|---|---|---|---|
| **Total** | | 6.07 (5.88-6.26) | 6.19 (5.87-6.51) | 5.73 (5.44-6.02) | 6.35 (5.96-6.73) |
| **Age** | | | | | |
| | <10y | 5.57 (5.32-5.83) | 5.67 (5.23-6.12) | 5.21 (4.84-5.58) | 5.96 (5.46-6.47) |
| | 10-19y | 6.48 (6.02-6.94) | 6.33 (5.6-7.06) | 6.30 (5.63-6.98) | 7.00 (5.96-8.04) |
| | 20-29y | 5.77 (5.28-6.26) | 5.8 (5.04-6.56) | 5.72 (4.84-6.6) | 5.80 (4.82-6.78) |
| | 30-39y | 6.99 (6.41-7.57) | 6.84 (5.96-7.72) | 7.05 (6.00-8.09) | 7.20 (6.05-8.35) |
| | 40-49y | 6.86 (6.08-7.65) | 7.23 (5.85-8.61) | 6.67 (5.51-7.82) | 6.47 (5.01-7.92) |
| | 50-59y | 5.80 (4.90-6.70) | 5.76 (4.66-6.87) | 5.77 (3.95-7.59) | 5.90 (3.74-8.06) |
| | 60y + | 5.84 (4.69-6.99) | 7.73 (5.54-9.91) | 3.56 (2.84-4.27) | 5.80 (4.26-7.34) |
| **Sex** | | | | | |
| | Male | 6.15 (5.87-6.43) | 6.15 (5.69-6.61) | 6.02 (5.55-6.49) | 6.34 (5.77-6.9) |
| | Female | 5.99 (5.73-6.24) | 6.23 (5.79-6.67) | 5.46 (5.11-5.82) | 6.36 (5.84-6.89) |
| **Occupation** | | | | | |
| | Pre-school | 5.42 (5.17-5.66) | 5.51 (5.11-5.91) | 5.08 (4.72-5.44) | 5.76 (5.24-6.29) |
| | Student | 6.57 (6.10-7.04) | 6.51 (5.63-7.38) | 6.24 (5.61-6.88) | 7.20 (6.25-8.15) |
| | Manual/Office/Shop worker | 7.23 (6.34-8.12) | 7.0 (3.72-10.28) | 7.47 (6.06-8.87) | 7.0 (5.78-8.22) |
| | Housewife | 5.76 (5.34-6.17) | 5.67 (5.06-6.28) | 5.43 (4.75-6.1) | 6.67 (5.68-7.65) |
| | Agriculture | 7.02 (6.35-7.68) | 7.35 (6.53-8.16) | 5.92 (4.86-6.98) | 7.00 (5.04-8.96) |
| | Unemployed/Retired | 5.18 (4.4-5.96) | 6 (3.88-8.12) | 4.42 (3-5.83) | 5.26 (4.25-6.27) |
| | Others | 5.83 (5.26-6.4) | 6 (5.14-6.86) | 5.69 (4.68-6.7) | 5.7 (4.42-6.98) |
| **Setting** | | | | | |
| | Household | 2.8 (2.68-2.92) | 2.94 (2.74-3.15) | 2.48 (2.28-2.67) | 3.02 (2.8-3.24) |
| | Community | 3.27 (3.09-3.45) | 3.25 (2.95-3.54) | 3.25 (2.97-3.53) | 3.33 (2.98-3.67) |
| **Traveled to a different neighborhood in the month prior the interview** | | | | | |
| | Yes | 6.21 (5.83-6.59) | 6.08 (5.34-6.81) | 5.66 (5.16-6.16) | 7.22 (6.46-7.97) |
| | No | 6.01 (5.79-6.23) | 6.22 (5.86-6.57) | 5.76 (5.4-6.13) | 5.93 (5.51-6.35) |
| **Contacts outside neighborhood (%)** | | | | | |
| | 0-14y | 0.67% (0.4-1.04) | 0.38% (0.14-0.94) | 0.19% (0.03-0.77) | 1.80% (1.03-3.09) |
| | 15-59y | 3.98% (3.23-4.89) | 4.54% (3.36-6.07) | 3.3% (2.19-4.9) | 3.91% (2.47-6.07) |
| | 60y+ | 2.74% (1.28-5.53) | 1.18% (0.20-4.63) | 0.00% (0.00-0.00) | 10.34% (4.28-21.84) |

TABLE 3.2: Mean number of recorded daily contacts, excluding contacts at school, by age, across different geographical contexts.

For all sites, contacts outside school were predominantly reported between family members (46.1%), neighbors (25.2%), and other relatives outside the household

(13.1%), while the remaining 15.5% of contacts occurred with friends, schoolmates outside school, or other unspecified categories. Individuals with a recent history of travel outside their neighborhood did not report an increased number of contacts, except for urban residents (t-test p=0.004). The mean number of contacts (excluding school contacts) reported by participants was lower in rural villages (5.73, 95%CI 5.44-6.02) with respect to both urban neighborhoods (6.35, 95%CI 5.96-6.73) and remote settlements (6.19, 95%CI 5.87-6.51). In particular, the mean number of daily contacts reported by the elderly (60+ years old) was much higher in remote settlements and urban neighborhoods than in rural villages (7.7 and 5.8 vs 3.6, see Table 3.2).

Students reported 1,372 additional contacts in schools, resulting in a mean number of 6.1 (95%CI 4.98-7.16) daily physical contacts per child (median 3, interquartile range 0-10). There were limited differences in the mean number of school contacts across geographical contexts (6.31, 95%CI 4.13-8.50 in remote settlements; 5.70, 95%CI 4.19-7.21 in rural towns; 6.54, 95%CI 4.25-8.84 in urban neighborhoods).

The analysis of contacts by age clearly shows that subjects below 30 years of age tend to interact mostly with individuals of similar age (assortative mixing). The highest contact rates were found between school aged children (10-19 years), young adults (20-39 years) and between children below 10 years and their parents (Figure 3.1, and Appendix: Sections 6 and 7). However, a marked intergenerational mixing both within households and in the community was found, especially in remote settlements.

The average overall number of daily contacts reported by our study participants (7.5 contacts), the share of contacts experienced with household members (46.1% including all ages) and the proportion of school contacts for children between 5 and 21 years of age (40.3%) are in line with estimates obtained by similar studies conducted in Zimbabwe, Uganda and Kenya (Melegaro et al., 2017; Waroux et al., 2018; Kiti et al., 2014), where the number of contacts per day was found in the range 7-11, the proportion of contacts at home was 50-66% and around 50% of contacts of school-aged children were recorded between schoolmates. The high level of mixing between the elderly and both young adults and children has been already highlighted for Ethiopia by the synthetic contact matrices estimated in Prem et al. (Prem, Cook, and Jit, 2017).

### 3.3.2 Effect of demography and age-specific contacts on COVID-19 epidemics

From the epidemic curve of reported cases, we estimated a net reproduction number R of 1.62 (95%CI 1.55-1.70) over approximately 6 weeks of exponential growth starting from May 1, 2020 when schools were closed in the entire country (see Appendix: Section 4). We relied on this estimate of R to simulate COVID-19 epidemics in the SWSZ considering no school contacts. If school contacts are included, we estimate R to increase to 3.15 (95%CI 2.22-4.20, see Appendix: Section 4), which is comparable with estimates of the basic reproduction number from other parts of the world (Riccardo et al., 2020; Munayco et al., 2020; Muniz-Rodriguez et al., 2020; Park et al., 2020).

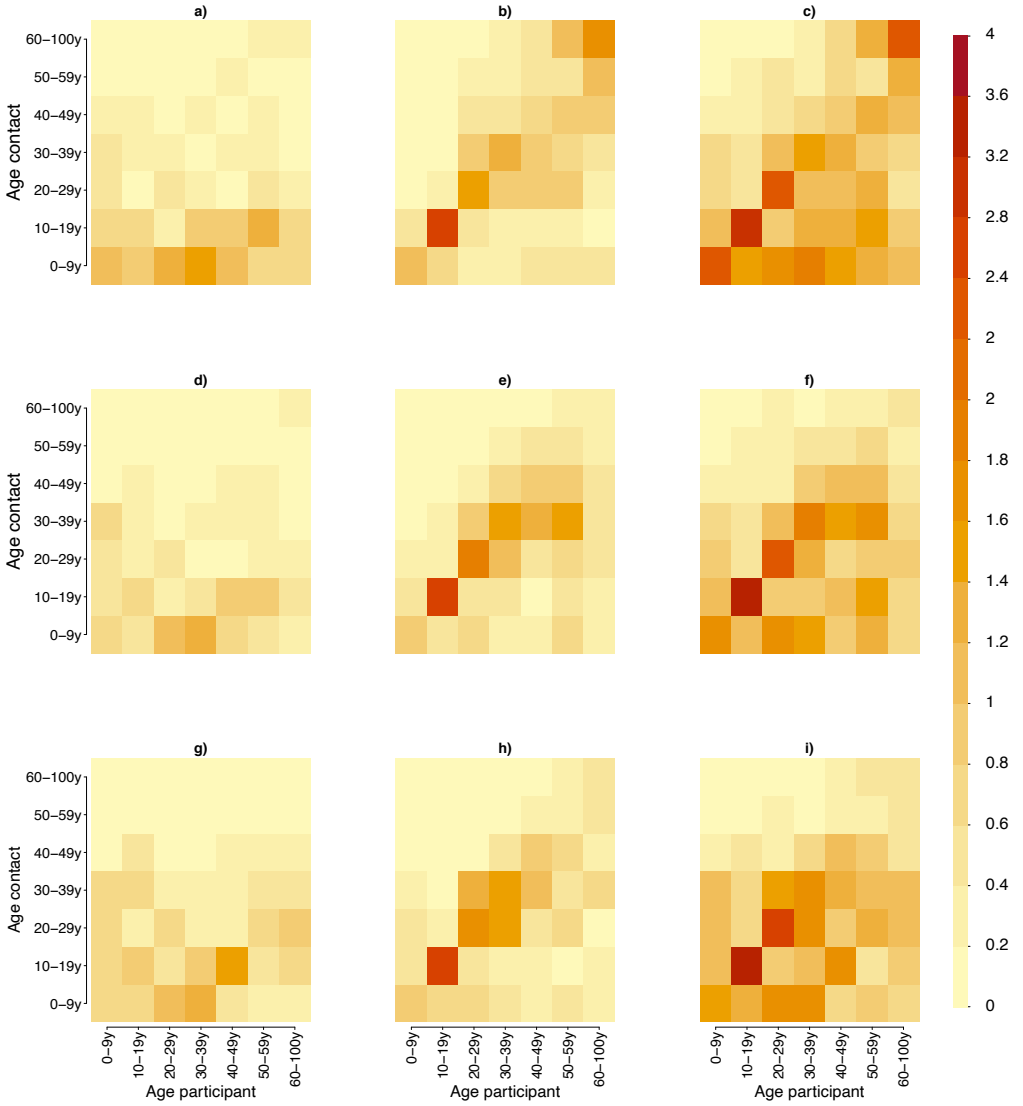


FIGURE 3.1: Contact matrix representing the mean number of daily contacts reported by a participant in the age group i with individuals in the age group j in household (a), in the general community (b), and both (c) in remote settlements. d-f, g-i The same quantities estimated for rural villages and for the urban neighborhoods, respectively

Our simulation results show that, had schools remained closed for the entire duration of the epidemic and had no other interventions been enacted, 12.11% (95%CI 10.78-13.51), 12.13% (95%CI 10.57-13.55), and 13.12% (95%CI 11.62-14.96) of the population residing in rural, remote, and urban settings respectively would have developed respiratory symptoms or fever because of COVID-19. The fraction of critical cases (requiring mechanical ventilation and/or resulting in a fatal outcome) is estimated between 0.28% and 0.41% of the overall population (Figure 3.2). The highest prevalence of critical cases (between 4.4% and 5.4% on average) is expected within subjects aged 60 years or older. This age segment represents only about 5% of the total population in SWSZ but is expected to represent 7 to 14% of symptomatic cases and 43% to 63% of all critical cases.

Remote settlements are expected to suffer a higher overall burden of critical cases (0.40% of the total population, 95%CI: 0.37-0.41%) compared to rural villages (0.33%, 95%CI: 0.31-0.35%) and urban neighborhoods (0.31%, 95%CI 0.29-0.33%). This difference is explained by a higher proportion of the elderly in the population (see Appendix: Sections 6 and 7), but also by their higher number of daily contacts and the higher intergenerational mixing (Figure 3.1 **c)**) compared to the other settings, which results in a higher attack rate of infections, symptomatic cases, and critical disease in this age group (Figure 3.2). Urban neighborhoods, where highest contact rates at younger ages were recorded, are expected to have the highest attack rate of infections (57.3%, 95%CI: 49.6-66.7) and symptomatic cases (13.1%, 95%CI: 11.6-15.0). However, since a large proportion of the overall number of infections (81.8%, 95%CI: 76.1-85.3) is concentrated on children and younger adults (up to 40 years of age), this does not result in a high overall proportion of critical disease. Finally, rural villages have lower attack rates among the elderly because of the significantly lower number of contacts reported by that age group in this geographical context (Figure 3.1 **f)**, Table 3.2).

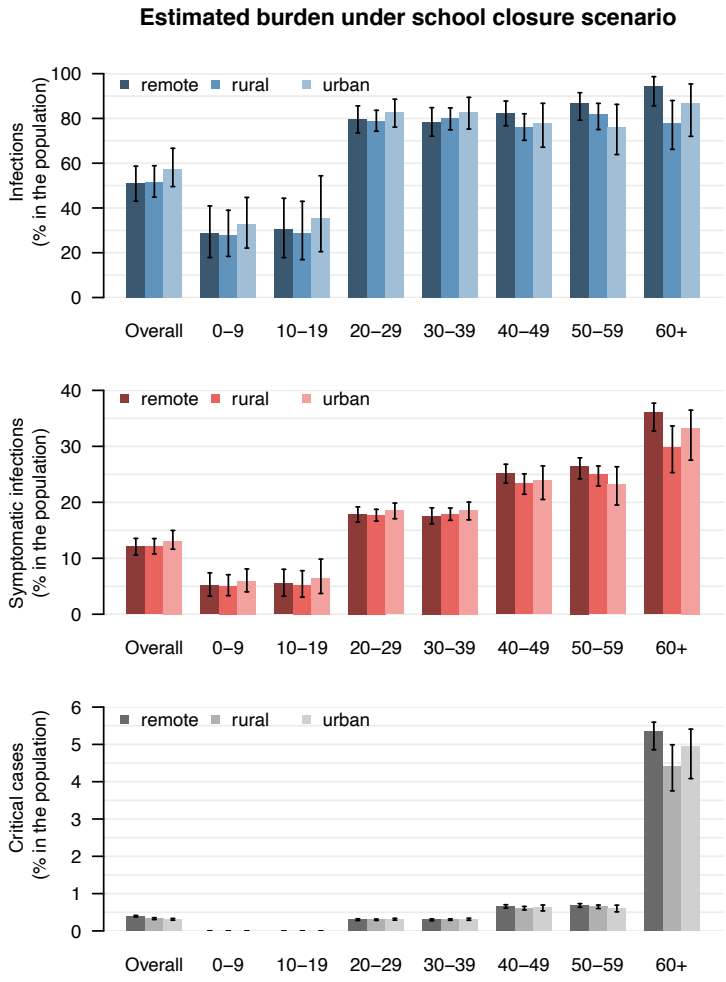


FIGURE 3.2: Estimated attack rates of infection (a), symptomatic cases (b), and critical disease (c), overall and by age group in different geographical contexts of the SWSZ, as expected at the end of an epidemic mitigated by school closure alone. Outputs were obtained by simulating 1000 different epidemics where the per-contact transmission rate is set to reproduce, when neglecting contacts occurring at school, random samples of the distribution of the net reproduction number estimated from national surveillance data: 1.62 (95% CI 1.55–1.70) (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*). Black lines represent 95% credible intervals

Figure 3.3 shows the impact of maintaining the school closure mandate on the burden of COVID-19 in the SWSZ in terms of percentages of infections, symptomatic and critical cases averted with respect to a hypothetical scenario of an unmitigated SARS-CoV-2 epidemic.

According to our estimates, the beneficial impact of school closure would be in 26.9% (95% CI: 20.7-32.8), 29.9% (95%CI: 19.5-38.7) and 25.1% (95%CI: 18.2-30.8) averted symptomatic cases, and 10.6% (95%CI: 8.1-13.7), 6.3% (95% CI: 3.8-10.3) and 8.1% (95%CI: 4.7-12.1) averted critical cases respectively in rural, remote and urban contexts. As expected, the larger effect of the intervention in terms of averted infections

is observable in younger ages, while the indirect effect of school closure on the elderly is highlighted by the expected high fractions of averted critical cases among individuals aged 50 or over.

A comparison of model estimates obtained in our baseline analysis with those obtained in the sensitivity analyses is provided in Figure 3.4. These results suggest that our estimates on the overall fraction of critical cases expected by maintaining the school closure mandate are robust against all alternative assumptions considered, ranging between 0.25%-0.37%, 0.23%-0.42% and 0.25%-0.34% for rural villages, remote settlements and urban areas, respectively.

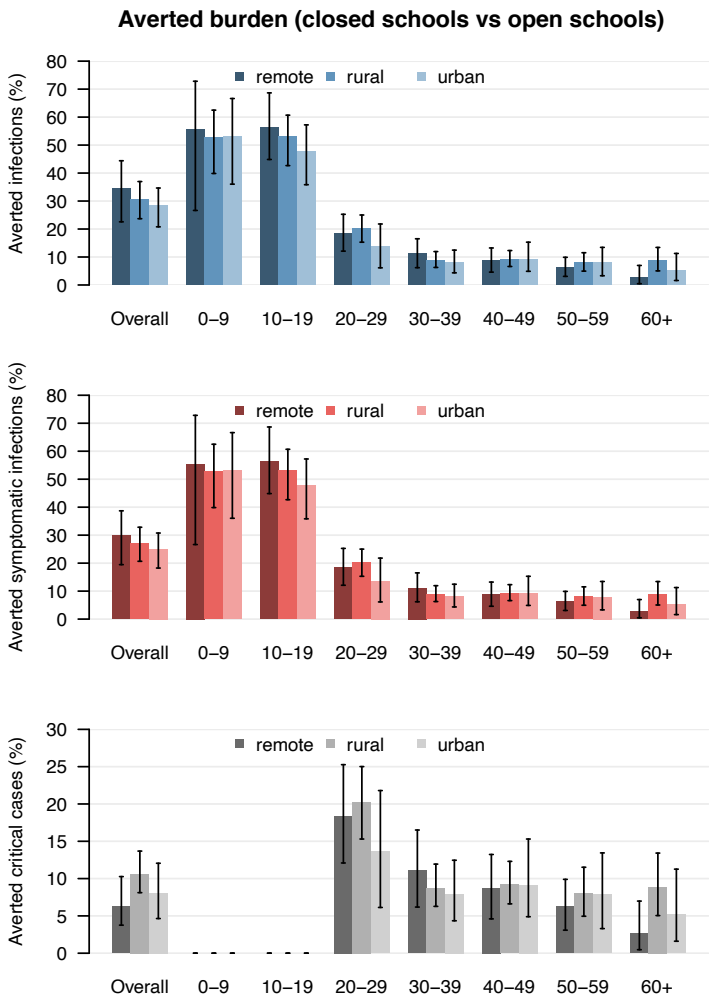


FIGURE 3.3: Estimated percentage of a averted infections, b symptomatic infections, and c critical cases, overall and by age group in different geographical contexts of the SWSZ with respect to a hypothetical scenario without school closure. Black lines represent 95% credible intervals

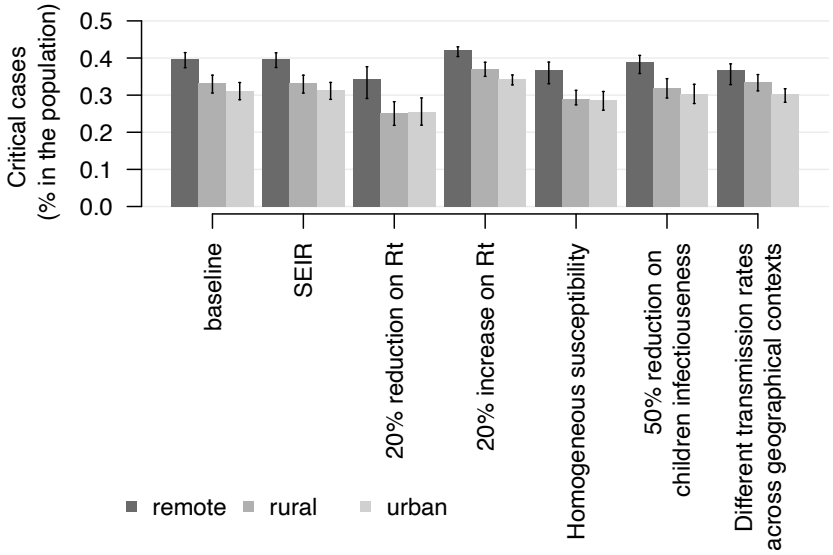


FIGURE 3.4: Comparison of the estimated overall percentage of critical cases in different geographical contexts of the SWSZ in the baseline and sensitivity analyses. Black lines represent 95% credible intervals

## 3.4 Discussion

Our analysis explored the effect of demographics and social contact patterns on COVID-19 burden in the South West Shewa Zone of the Oromia region, Ethiopia. Data collected within an interview-based survey highlighted differences in demographic structure and in age-specific contacts between urban neighborhoods, rural villages, and remote settlements, and were used to inform an epidemic model simulating the transmission dynamic of SARS-CoV-2. On the basis of the trajectory of COVID-19 cases observed in the country up to June 12, 2020, we estimated that between 3.1 and 4.0 patients per 1,000 inhabitants may experience critical disease (i.e., requiring mechanical ventilation and/or resulting in a fatal outcome) at the end of an epidemic mitigated by school closure alone. Considering the low availability and accessibility of healthcare, especially in remote and rural settlements, and the lack of intensive care units to treat critical patients (Poletti et al., 2018; Murthy, Leligdowicz, and Adhikari, 2015), it is possible that a large fraction of those cases would result in a fatal outcome, adding up to the already high background mortality rate in the region (estimated at about 6.4 per 1,000 per year (*The World Bank. World Bank Open Data.*)).

Considering the extreme scenario where all critical cases would result in a fatal outcome, we obtain an estimate of the infection-fatality ratio (IFR) ranging between 0.55% in urban neighborhoods and 0.78% in remote settlements. Such estimates are generally lower than the IFR estimated from serological studies for higher income countries (Pollán et al., 2020; Poletti et al., 2020a). This difference is partially due to the younger age structure of the Ethiopian population, where only 5% of individuals

are older than 60 years (compared to over 20% in most of Europe (*Eurostat. Population Structure and Ageing.*)). However, by simply adjusting the age-specific IFR to the local demographics, Ghisolfi et al. (Ghisolfi et al., 2020) estimated a four-fold reduction in the overall IFR in Eastern Africa with respect to European countries, which is around 2 times lower than our estimates. In fact, our simulations not only account for the demography of the population, but also for its mixing patterns. Indeed, we found that in the SWSZ the effect of a younger population is partially compensated by high infection attack rates in the elderly, which derive from the intense intergenerational mixing and the larger number of contacts observed among the elderly. In particular, we show that these characteristics are especially marked in remote settlements, where the highest incidence of critical disease is expected to occur. Although our analysis is limited to the SWSZ, we expect that similar arguments may be generalizable to settings with similar socio-demographic conditions. Our results suggest that, in the SWSZ, school closures might have reduced by 48.9% the SARS-CoV-2 reproduction number and by 28.3-34.6% the infection attack rate that would have been expected in the absence of any intervention. In line with observations from other settings (Zhang et al., 2020), school closure was estimated to be insufficient to prevent the spread of the infection. Recently published studies have shown that the lockdown implemented in Kenya reduced individuals' social interactions by 60-70% compared to the pre-pandemic period (Quaife et al., 2020), but it is difficult to extrapolate these data to Ethiopia, where social distancing measures were comparatively milder. Data on how contacts outside school may have changed in Ethiopia during the COVID-19 epidemic are still lacking.

To properly interpret the results presented in our study, it is important to consider the following limitations. First, the target study population may be not representative of all Ethiopia, and in particular of epidemic patterns observed in highly urbanized areas such as the capital Addis Abeba. Second, the net reproduction number was estimated from national surveillance data (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*). This data reports cases aggregated at the country level and may suffer from a number of biases: it does not account for reporting delays; the growth over time in the number of cases may partly be ascribable to the increase in testing capacity; total cases represent the superimposition of different, asynchronous epidemics in multiple parts of the country, a majority of which coming from the highly urbanized Addis Abeba area (*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 37 weekly bulletin.*). More in general, estimates of time varying reproduction numbers from data where the symptoms' onset time-series is approximated with the notification date series may inaccurately describe the early infection dynamics and could fail in assessing the impact of containment measures. However, we show that, when assuming no restriction to school contacts, the reproduction number estimated by the model is in the range 2.43-3.52, comparable with estimates of the SARS-CoV-2 basic reproduction number from other countries (Riccardo et al., 2020; Munayco et al., 2020; Park et al., 2020; Murthy, Leligdowicz, and Adhikari, 2015). Moreover, our conclusions remain robust when considering a 20% increase or a 20% decrease of the reproduction number. In this case, we estimated an attack rate of critical cases ranging from 0.25 to 0.37 for rural villages and from 0.34 to 0.42 for remote settlements (see Figure 3.4). Third, the model lacks of spatial structure. The finding

from the survey that about 97% of recorded contacts have occurred within the participant's neighborhood of residence (Table 3.2) suggests that local containment or confinement of COVID-19 outbreaks in rural regions of Ethiopia may be favored by low human mobility. On the other hand, the observation of a large number of cases in all regions of Ethiopia (*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 37 weekly bulletin.*) may imply that a significant widespread diffusion of the epidemic, possibly sustained by a high fraction of asymptomatic infections (Figure 3.2), is ongoing. Fourth, the role played by children in the transmission of SARS-CoV-2 infections is still poorly understood and highly debated (Zhang et al., 2020; Bi et al., 2020). In the main analysis we assumed that the probability of transmission is homogeneous across all ages. As asymptomatic infections are more prevalent at younger ages, this also reflects the assumption that symptomatic and asymptomatic cases are characterized by the same infectiousness. However, an alternative assumption in which children are assumed half as infectious as adults would result in similar attack rates of critical cases (see Appendix: Section 8). These results are also robust with respect to the assumption of a homogeneous susceptibility across age groups (see Appendix: Section 8). Finally, in absence of direct data from sub-Saharan Africa, the age-specific susceptibility and proportions of infections resulting in symptomatic cases or critical disease were estimated from data from China or Europe (Zhang et al., 2020; Poletti et al., 2020b). However, the high prevalence of comorbidities which are uncommon in higher income countries (e.g., malnutrition (Endris, Asefa, and Dube, 2017), tuberculosis, and malaria) and inequalities in the access to primary care represent additional vulnerabilities for African settings (Poletti et al., 2018) and may result in an underestimation of the expected disease burden. Since the number of COVID-19-related deaths may be under ascertained in low-income countries, further research is warranted regarding the disease severity in sub-Saharan populations, potentially leveraging excess mortality data once they will become available.

## 3.5 Conclusions

This study provides novel data on mixing patterns in rural Ethiopia and highlights the potential impact of COVID-19 epidemics in less urbanized regions of the country. We provide estimates on the potential burden of COVID-19 in the SWSZ under the assumption of a mitigated, but not controlled epidemic. We conclude that, although the overall mortality might be generally lower in sub-Saharan Africa compared to high income settings, thanks to younger demographics (Ghisolfi et al., 2020; Dowd et al., 2020; Hilton and Keeling, 2020), this effect may be partially offset in rural areas by higher attack rates in elderly individuals, due to high rates of intergenerational mixing. The observed contact patterns suggest that elderly individuals in remote settlements may be even more exposed to the risk of infection (and thus of critical disease), which is especially worrysome in light of the major obstacles in access to healthcare for those populations (Poletti et al., 2018).

## 3.6 Appendix

### 3.6.1 Study desing

The study population consisted of individuals residing in four districts (woreda) of the South West Shewa Zone (SWSZ) in the Oromia region of Ethiopia. These woredas count 449,460 inhabitants and represent the main catchment area of the St. Luke Hospital located in Woliso Town. The St. Luke Hospital is a well-resourced health facility and represents the referral hospital for the entire SWSZ, serving a population of about 1.3M inhabitants with 200 beds and an annual average bed-occupation rate of 84% (Poletti et al., 2018). Data on individuals' mixing patterns and local demography were collected through a cross-sectional survey, by adopting a two-stage stratified random sampling of study participants by location and age group. For each woreda, two neighbourhoods (kebeles) were identified as representative of the considered woreda, chosen as extremes illustrative socio-demographic contexts within the woreda in terms of urbanization, population density, work and travel opportunities, and distance to healthcare facilities. The target sample size was uniformly distributed across the 8 selected kebeles. The sample stratification was designed to capture different activity levels (e.g. movements, schooling/working, etc.) and the different role played by individuals in the community (e.g. household heads, women, etc.), taking into account the local schooling system (age at enrolment in pre-primary, primary, and secondary school). Individuals of all ages living in the selected sites were considered eligible for inclusion in the study. A target sample size was defined for the following age groups: <1 year old, 1-3 years old, 4-10 years old, 11-14 years old, 15-29 years old, 30-49 years old, >50 years old. Random sampling of households and study participants was applied, using a list of predefined quotas for each site, sex and age group. Specifically, the target sample for each age group and location was equally divided into males and females. One individual per household was selected and interviewed. If the study participant was temporarily outside the household, another attempt was made later in the day or within three days from the first visit. After the second attempt, the study participant was replaced.

### 3.6.2 Sample size definition

For each age group i, we chose an equal sample size $n_i$ in such a way to detect, given a specified power p and significance level set at 0.05, a significant difference in the average number of contacts between at least two out the seven age groups defined above in a one-way ANOVA (Cohen, 2013). The optimal sample size can be computed as a function of the power of the test, the significant level and the effect size f, which in turn can be calculated using the following formula,

$$f = \sqrt{\frac{\sum_{i=1}^{k} \frac{1}{k}(\mu_i - \mu)^2}{\sigma^2}}$$

where k is the number of groups, $\mu_i$ is the expected average number of contacts for age group $i$, $\mu$ is the expected average number of contacts in the overall population, and $\sigma^2$ is the expected constant error variance within groups. As shown in the Figure 3.5, setting a power of 80%, $k = 7$ and a significance level at 0.05, a sample size of 120 in each group would correspond to the optimal sample size for $f = 0.13$,

which can be considered as a sufficiently small effect size. Indeed, by considering values of $\mu_i$, $\mu$ and $\sigma^2$ obtained in previous studies on social contacts (Waroux et al., 2018; Horby et al., 2011) , the effect size would be around 0.17. Based on previous findings available at the time (Horby et al., 2011), the considered sample size enabled the detection of 20% difference in the average number of daily contacts by age group.

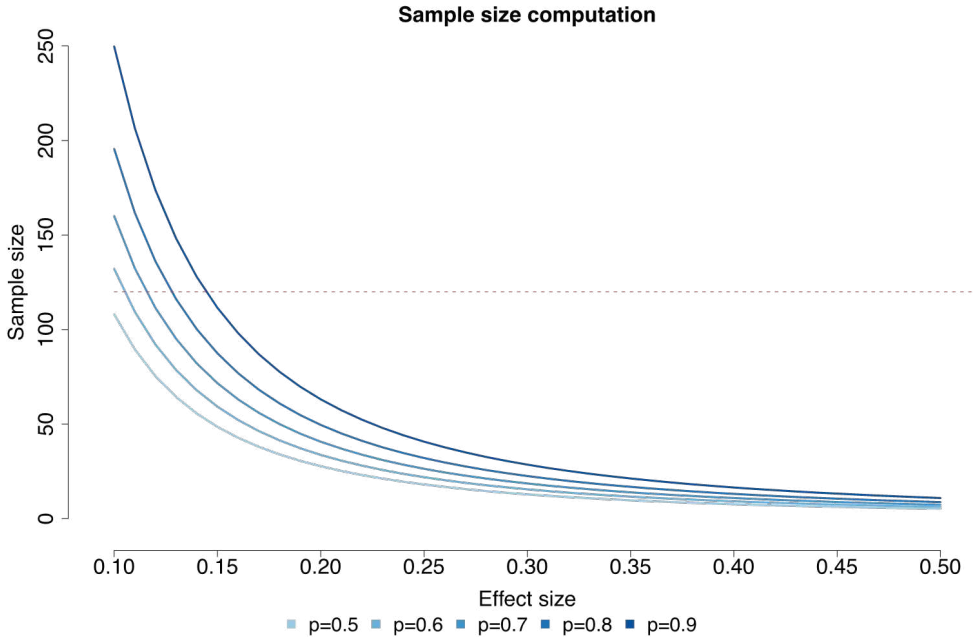


FIGURE 3.5: **Sample size definition** Optimal sample size computed for different values of effect size and power of the test (p), assuming a significance level of 0.05. The horizontal line represents the target sample size defined in our study

In the final sampling scheme, senior adults (30-49 years old) were slightly over-sampled and the elderly (>50 years old) slightly under-sampled, due to the different relative frequency in the population.

### 3.6.3 data collection

For each study participant we collected data on their age, sex, household size, household composition, place of residence and the full list of contacts they experienced in the day preceding their interview. Specifically, the frequency and type (either physical or non-physical) of each social encounter was collected, along with the age and relationship with each listed contact and the transmission setting and the kebele (neighborhood) where the interaction occurred. The day of the week in which the interview was administered to each study participant was also recorded. Data collection was performed through interviews of the study participants by field investigators, who directly inserted anonymized answers into an electronic dataset based on Survey CTO software, installed on tablets. Questions were designed in English and translated into the predominant local language (Oromyffa). Data entry conflicts

and inconsistencies were identified automatically by the system and resolved as the data entry progressed. Results of a preliminary pilot study on a sample of 20 people, recruited in a different site from those used in the survey, were used to optimize the interview, address logistic challenges and refine the operational guidelines for the data collectors conducting the interviews. Following the results obtained for students in the pilot, we decided to collect only data on physical contacts at school. In particular, only the overall number of contacts experienced by students at school was collected. Responses gathered during the pilot were excluded from the analysis. Quality of data collection was then verified by administering 78 individual interviews in a remote settlement of the SWSZ which was outside the original target sites. These data were included in the analysis of contact patterns.

### 3.6.4 Transmission model and reproduction numbers

We developed a transmission model for the spread of SARS-CoV-2 infection, based on an age-structured susceptible-infectious-removed (SIR) scheme. Contact data collected with 938 individual interviews was used to inform the model with the age-specific mixing patterns in the South West Shewa Zone across different geographical contexts and transmission settings. To this aim, participants and contacts were grouped in six 10-year age classes plus an additional class including all individuals aged 60 years or older. When exact age of the contactee was unknown, the midpoint of the age range provided during the interview was used to assign the contactee to an age class. We then computed age-specific contact matrices $C^x_{a,\tilde{a}}$ representing the average number of contacts reported by one individual in age group $a$ with contactees in age group $\tilde{a}$ in the setting $x$. Considered transmission settings included the household, the school and the general community. Contacts at work were aggregated with all other contacts occurring in the community, since for people employed in agriculture (about 33% in Ethiopia) and many other occupations (e.g. street vendors and people participating to community markets) it was difficult to disentangle encounters occurred because of their job from other random contacts. Only physical contacts were considered for school; both physical and non-physical social interactions were considered for other transmission settings. Sample variability was explored using bootstrap sampling, as detailed in section 6. Contact matrices were separately computed for three different geographical contexts by aggregating interviews conducted in remote settlements (n=400), rural villages (n=326) and the two urbanized neighbors of Woliso Town ($n = 212$) and corrected for reciprocity as detailed in the following section.

In the model, infectious contacts within and between age classes may occur in three different transmission settings (household H, schools S, community C), and are combined in an overall contact matrix, specific for each geographical context, according to the following equation:

$$M_{a,\tilde{a}}(t) = C^H_{a,\tilde{a}} + \sigma_s C^S_{a,\tilde{a}} + C^C_{a,\tilde{a}} \tag{3.1}$$

where:

- $C^H_{a,\tilde{a}}$, $C^S_{a,\tilde{a}}$, $C^C_{a,\tilde{a}}$ are the contact matrices for the transmission settings described above;

- $\sigma_s$ is a parameter which is set equal to 0 to consider the transmission dynamics under the school closure mandate, and equal to 1 to assess the transmission dynamics when schools are open;
- $M_{a,\tilde{a}}$ represents the age-specific contact matrix, whose entries describe the mean number of persons in age group $\tilde{a}$ encountered by an individual of age group $a$ per day across different settings.

The proportions of the SWSZ population living in each geographical context were used as sampling weights to compute average contact matrices for the entire SWSZ.

In the model, we assumed asymptomatic and symptomatic individuals to be equally infectious, as suggested by an early analysis of virological data from Lombardy (Cereda et al., 2020) and Veneto (Lavezzo et al., 2020). The transmission model considers three consecutive infectious compartments to reproduce a gamma-distributed generation time (Cereda et al., 2020; Guzzetta et al., 2020). The force of infection for subjects of age a is defined as:

$$\lambda_a(t) = \beta r_a \sum_{\tilde{a}} \hat{r}_{\tilde{a}} M_{a,\tilde{a}} \frac{\alpha_I I_{\tilde{a}}(t) + \alpha_J J_{\tilde{a}}(t) + \alpha_K K_{\tilde{a}}(t)}{N_{\tilde{a}}} \tag{3.2}$$

where:

- $\beta$ is a scaling factor shaping the number of potentially infectious contacts resulting in infection;
- $r_a$ is the relative susceptibility to SARS-CoV-2 infection at age $a$;
- $\hat{r}_{\tilde{a}}$ is the relative infectiousness at age $\tilde{a}$;
- $I_{\tilde{a}}(t)$ , $J_{\tilde{a}}(t)$ and $K_{\tilde{a}}(t)$ represent the number of individuals of age $\tilde{a}$ in the three stages of infection $I$, $J$,$K$, at time $t$;
- $\alpha_I$, $\alpha_J$ and $\alpha_K$ are adjusting factors for individuals' infectiousness during the three stages of infection $I$,$J$ and $K$;
- $N_{\tilde{a}}$ represents the total number of individuals in age group $\tilde{a}$ ;

In the baseline analysis, we assumed that, compared to adults aged 20-59 years ($r_a$=1), individuals aged <20 years are 67% less susceptible to infection (i.e. $r_a$=0.33; 95%CI 0.24-0.47) and those aged $\geq$60 years are 47% more susceptible to infection ($r_a$=1.47; 95%CI 1.16-2.06) (Zhang et al., 2020); homogeneous susceptibility to SARS-CoV-2 infection across ages was considered for sensitivity analysis ($r_a$=1 for all $a$). In the baseline analysis, individuals of different ages were considered equally infectious ($\hat{r}_{\tilde{a}}$=1 for all $\tilde{a}$). For sensitivity analysis, we assume that individuals aged 0-19y are 50% less infectious than other individuals ($\hat{r}_{\tilde{a}}$=0.5 when $\tilde{a}$<19; $\hat{r}_{\tilde{a}}$=1 for $\tilde{a}$ $\geq$20). Finally, we assumed that recovering from infection provides full immunity against re-infection for at least the duration of our simulations (2 years).

Transitions across different epidemiological classes can be summarized by the following differential system:

$$\begin{cases} S'_a(t) = -\lambda_a(t)S_a \\ I'_a(t) = \lambda_a(t)S_a - \gamma I_a(t) \\ J'_a(t) = \gamma I_a(t) - \gamma J_a(t) \\ K'_a(t) = \gamma J_a(t) - \gamma K_a(t) \\ R'_a(t) = \gamma K_a(t) \end{cases} \tag{3.3}$$

where:

- $S$ represents the number of individuals susceptible to SARS-CoV-2 infection;
- $\gamma$ is the recovery rate associated with each stage of infection: $I$, $J$, $K$;
- $R$ represents the number of individuals who recover from the infection.

We assumed that the average generation time of SARS-CoV-2 can be approximated with the observed average serial interval, which was estimated in 6.6 days (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*). The adjusting factors $\alpha_J$, $\alpha_J$ and $\alpha_K$ were set equal to 0.014, 0.9 and 0.086 respectively, in such a way to reproduce a distribution of the generation time consistent with that of the observed serial interval, i.e. a Gamma distribution with shape 1.87 and rate 0.28 (Cereda et al., 2020; Marziano et al., 2021a).

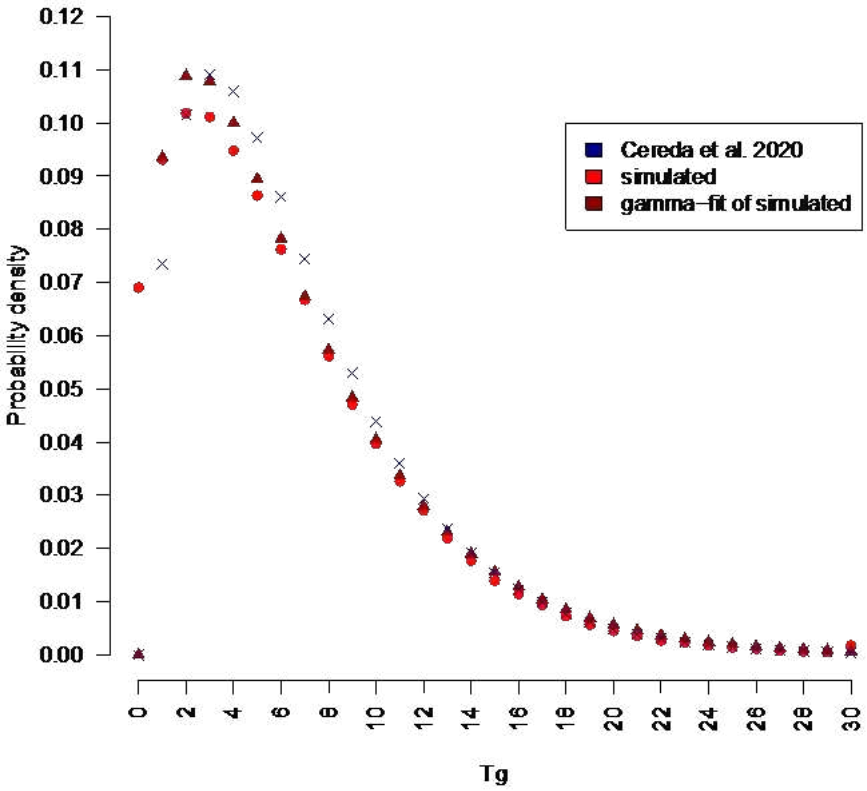


FIGURE 3.6: **SARS-CoV-2 generation time** Distribution of the SARS-CoV-2 generation time (red) as simulated in our model when assuming $\gamma$=0.303 days$^{-1}$, $\alpha_I$=0.014, $\alpha_J$=0.9 and $\alpha_K$=0.086 compared to the distribution of the SARS-CoV-2 serial interval as observed in Italy (blue)

Reproduction numbers ($R$) associated with different geographical contexts were computed by using the Next Generation Matrix approach (Zhang et al., 2020). The parameter $\beta$ was assumed to be equal across different geographical contexts and calibrated by considering the average contact matrix for the entire South West Shewa Zone, by computing the model's Next Generation Matrix under the assumption of school closure, and by assuming that the resulting R is equal to the reproduction number estimated from the initial (from May 1 to June 12) exponential growth characterizing the reported COVID-19 cases in Ethiopia (mean 1.62; 95%CI: 1.55-1.70, see Figure 3.7) (Wallinga and Lipsitch, 2007). Alternative values of R (i.e +/- 20%) were considered for sensitivity.

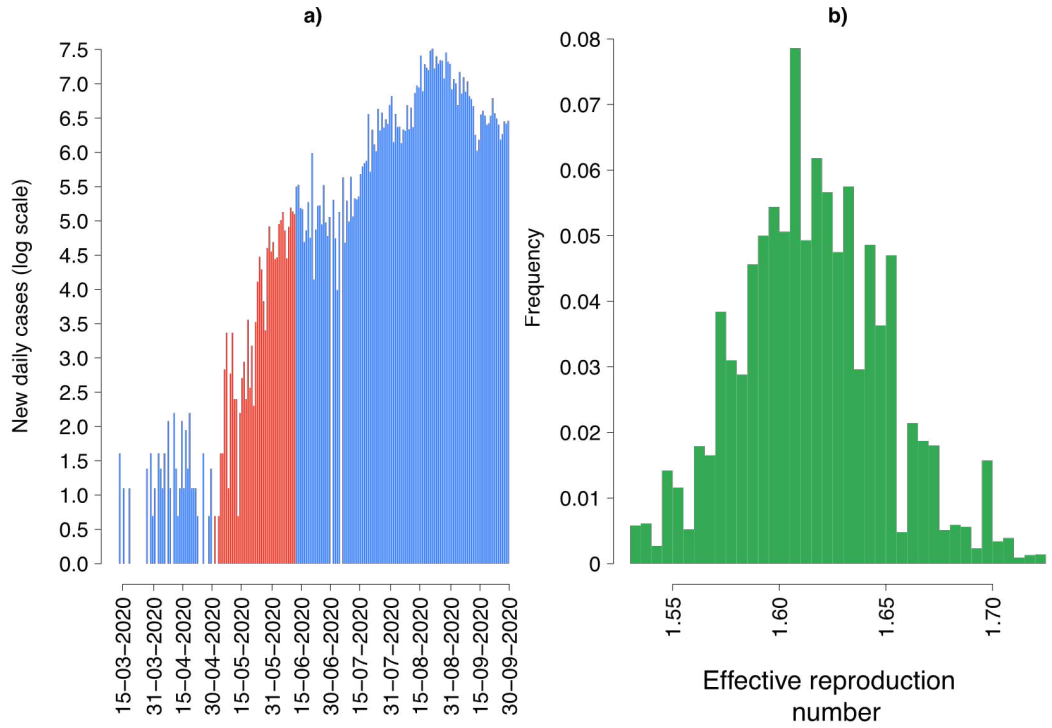


FIGURE 3.7: **Transmissibility potential. a)** Daily COVID-19 cases reported in Ethiopia (*World Health organization. Health Workforce Requirements for Universal Health Coverage and the Sustainable Development Goals. Human Resource for Health Observers Series No. 17.*). The red bars show the exponential phase considered to estimate the SARS-CoV-2 reproduction number in Ethiopia. **b)** Estimates of R obtained from the exponential growth of cases observed between May 1 and June 12.

Dynamic transmission of SARS-CoV-2 was investigated separately for the three geographical contexts (remote settlements, rural villages and urban neighbors) by considering a population stratified into 7 age groups (six 10-year age groups from 0 to 59 years and one age group for individuals aged 60 years or older). The age distribution of household members of study participants was used to define the population age-structure across different geographical contexts. Simulation results shown in the main text and in the following sections were obtained by using a stochastic version of the model described above and 1,000 stochastic runs accounting for variability in available estimates of $r_a$ (Zhang et al., 2020), uncertainty in the derived contact matrices and the uncertainty in the estimated value of R from surveillance data (Figure 3.7). Each simulation was initialized with 5 infections every 10,000 inhabitants, assigned randomly across age classes. Figure 3.8 shows a comparison between the weekly number of COVID-19 cases reported in Ethiopia during the period of exponential growth (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*) when all teaching activities were suspended in the entire country and the weekly number of infections occurring in the SWSZ as simulated by our model by excluding school contacts.

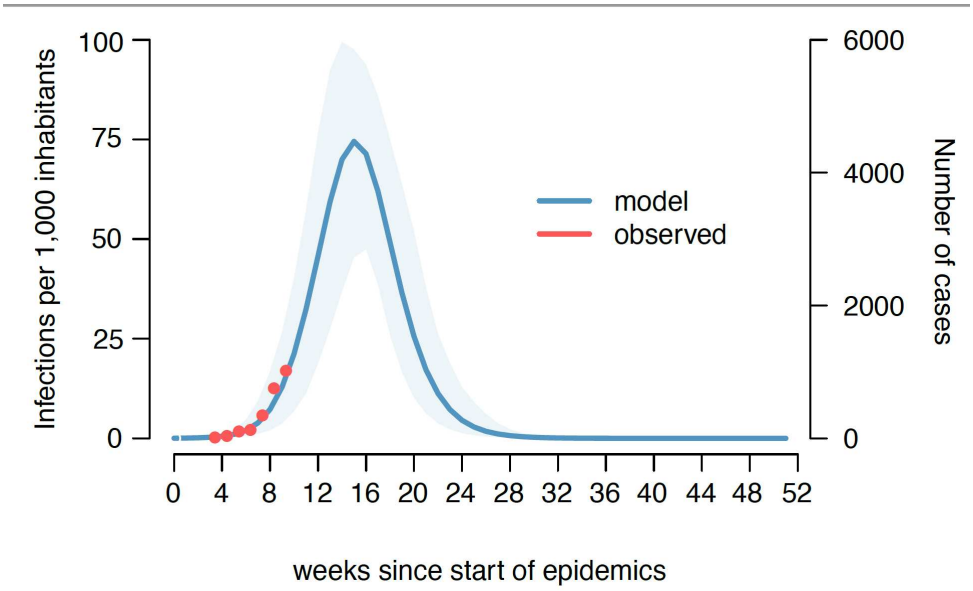


FIGURE 3.8: **Epidemic growth.** Number of infections per 1,000 inhabitants as estimated by the model (blue line: mean; shaded area: 95% credible intervals CI) compared to the growth observed in the number of weekly COVID-19 cases reported in Ethiopia (red dots) used for the estimation of the reproduction number.

Age specific attack rates for symptomatic infections and critical cases were obtained by applying estimates for the absolute probability of developing symptoms (respiratory or fever), and critical disease (either requiring mechanical ventilation or resulting in death) after infection, as provided in (Poletti et al., 2020b).

### 3.6.5 Adjustment of contact matrices for reciprocity

In order to robustly estimate the average number of observed contacts per person per day, we need to consider that the sample age distribution is different from the population age distribution and to take into account the probability of an individual to be included in the sample. All the considered contact matrices were therefore adjusted for reciprocity, by applying the same approach used in (Melegaro et al., 2017) and detailed as follows. Let $P_a$ denote the number of participants in the $a$-th age class and let $c_{(a,\tilde{a})}(i)$ denote the number of contacts a specific study participant $i$ of age $a$ has with individuals of age $\tilde{a}$. The total number of contacts $T_{(a,\tilde{a})}$ that all study participants of age $a$ have with individuals of age $\tilde{a}$ can be computed as

$$T_{(a,\tilde{a})} = \sum_{i=1}^{P_a} c_{(a,\tilde{a})}(i)$$

The average contacts an individual of age $a$ has with individuals of age $\tilde{a}$ can be approximated by the average contacts that a participant of age $a$ with individuals of age $\tilde{a}$ as follows:

$$C_{(a,\tilde{a})} = \frac{T_{(a,\tilde{a})}}{P_a}$$

In principle,$T_{(a,\tilde{a})}$ can be different from $T_{(\tilde{a},a)}$. In order to correct matrices for symmetry we should take into account the probability of an individual to be included in the sample. To do this, we corrected the total number of contacts that all study participants of age $a$ have with individuals of age $\tilde{a}$ as a weighted average of the total contacts reported by participants of different ages as follows:

$$T^{\textbf{corrected}}_{(a,\tilde{a})} = \frac{P_a N_a T_{(a,\tilde{a})} + P_{\tilde{a}} N_{\tilde{a}} T_{(\tilde{a},a)}}{P_a + P_{\tilde{a}}}$$

where $N_a$ is the size of the age group a in the population targeted by our contact survey. The adjusted average contacts an individual of age $a$ has with individuals of age $\tilde{a}$ can be finally computed as

$$C^{\textbf{corrected}}_{(a,\tilde{a})} = \frac{T^{\textbf{corrected}}_{(a,\tilde{a})}}{N_a}$$

### 3.6.6 Uncertainty in contact matrices

In order to take into account sample variability, we computed 1,000 bootstrapped contact matrices for each geographical context and transmission setting. At each bootstrap iteration, we sampled with replacement 400, 326 and 212 interviews from those obtained in remote settlements, rural villages and urban neighborhoods respectively, choosing the age of the participant with probability proportional to the age distribution of the Ethiopian population (*United Nations Department of Economic and Social Affairs. 2019 UN World Population Prospects.*). Then, we counted for each participant $i$ of age group $a$ the number of contacts reported with individuals of age $\tilde{a}$ in the setting $x$, $c^x_{(a,\tilde{a})}(i)$, and estimated the average number of contacts occurring in the setting $x$ between ages $a$ and $\tilde{a}$from the following equation:

$$C^x_{(a,\tilde{a})} = \frac{\sum_{i=1}^{P_a} c^x_{(a,\tilde{a})}(i)}{P_a} \tag{3.4}$$

where $P_a$ is the number of sampled participants of age group $a$. Contact matrices resulting by averaging entries of 1,000 bootstrap $c^x_{(a,\tilde{a})}$ after the correction for reciprocity described in the previous section are reported in Figure 3.9 and Figure 3.10.

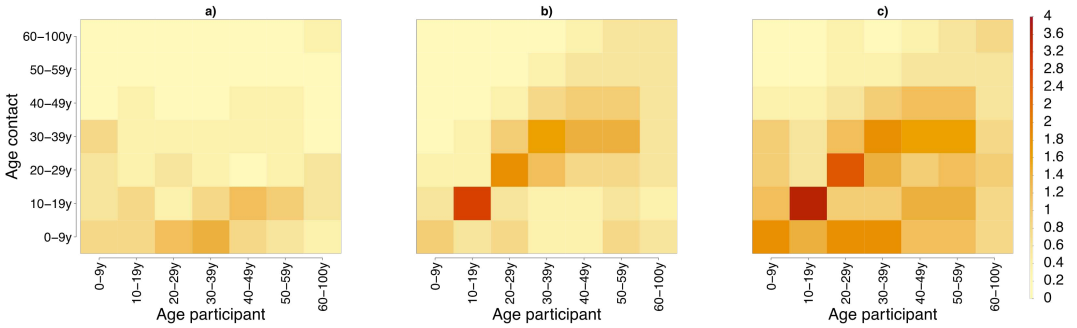


FIGURE 3.9: **Contact matrices by settings.** Age-specific contact matrices as obtained by averaging 1,000 bootstrapped contact matrices representing the average number of daily contacts reported by participants in the age group i with individuals in the age group j in household **a)**, in the general community **b)** and both **c)** in the SWSZ.

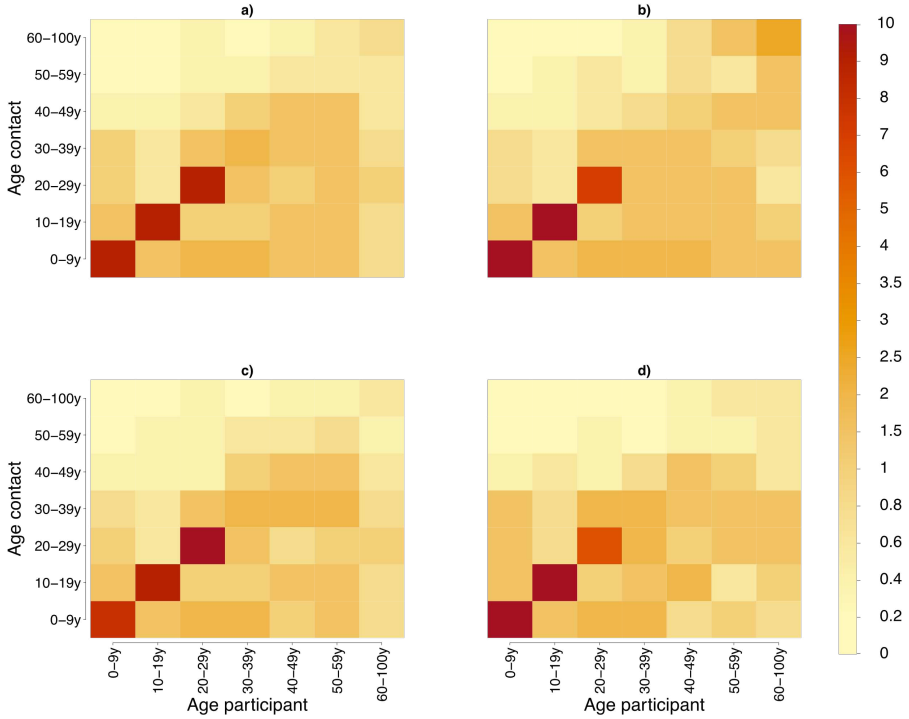


FIGURE 3.10: **Contact matrices by geographical context.** Age-specific contact matrices as obtained by averaging 1,000 bootstrapped contact matrices, representing the estimated average number of daily contacts that an individual in the age group *i* experience with individuals in the age group *j* across all settings (including schools) in the entire SWSZ **a)** in remote settlements **b)**, rural villages (c) and urban sites **c)**.

### 3.6.7 Additional results on contact patterns

The mean number of daily contacts per person was analyzed with respect to a set of covariates, including age, sex, type of work and geographical context of the study

participant, and day of the week in which the encounter occurred. A statistical comparison of mean values was carried out using either t-tests or ANOVA if the strata are more than two. Differences among three or more group means were assessed by a post-hoc analysis based on the Tukey test. A Kolmogorov–Smirnov (KS) test was used to compare distributions across different strata.

In our sample, 51% of individuals were female, and no significant differences were found in the sample age distribution across different geographical contexts (pairwise KS test p-value>0.28). Differences in the age distribution of all household members of study participants are reported in Figure 3.11.

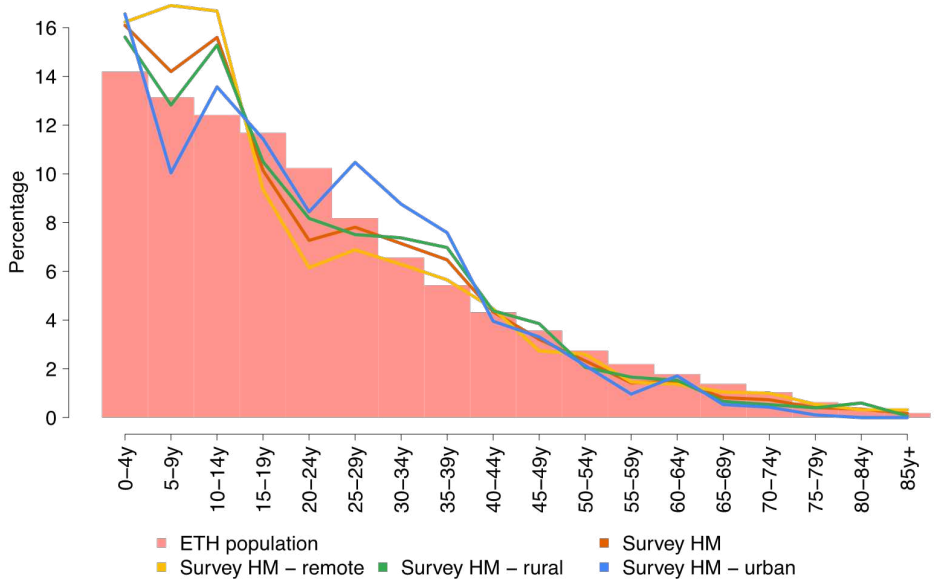


FIGURE 3.11: **Population age structure.** Age distribution of household members (HM) of study participants residing in the three geographical contexts and in the overall SWSZ with respect to the age distribution of the Ethiopian population reported in (*United Nations Department of Economic and Social Affairs. 2019 UN World Population Prospects.*).

Among study participants residing in remote settlements, 88.5% of male adults reported to work in agriculture. Although agriculture remains the main occupation in rural villages (30.6%), 38.8% of male adults living in these sites reported to be office, shop or manual worker; 30% of adults living in urban neighbourhoods were unemployed. In all sites, more than 60% of adult females were housewives and only 5% of working adults reported travels to a different kebele to reach their workplace. Only 9.0% of study participants accessed a health facility in the month preceding their interview. This latter percentage varies largely across age and geographical context, with percentages ranging from 7.7% among children living in remote settlements to 22.3% among those living in urban neighbors, respectively. 87.3% of the participants reported they were never admitted to the local hospital.

Highest contact rates were recorded among individuals aged 35-44 (7.22 95%CI 6.51-7.93), lowest in younger children (5.16 95%CI 4.87-5.45). However, the average number of daily contacts reported by individuals aged 65 years or more was similar to those reported by individuals aged 25-34 years (6.41, p-value>0.99). The number of daily contacts reported by people employed in agriculture was also remarkably high (mean: 7.02) when compared to office workers (mean: 8.07) and retired individuals (mean: 4.67). A similar number of daily contacts was found in males and females (6.15 vs 5.99, p-value=0.40). The number of contacts experienced during the weekends were not significantly different from those experienced during the week (6.12 vs 6.05, p-value=0.74). A significantly larger proportion of contacts outside the household was found among study participants living in rural villages (56.8%) than in those living in remote settlements (52.5% p-value =0.013) or in the more urbanized neighborhoods (52.4%, p-value =0.035). The percentage of contacts occurring outside the kebele of residence was very low in all sites: 1.5% in rural towns, 2.1% in remote settlements and 2.9% in urban neighborhoods. However, adult males residing in the urban neighborhoods and the rural towns (representing the 10.7% of the sample) were twice more likely to travel outside of their neighborhood compared to those living in remote settlements (p-value <0.001).

### 3.6.8 Sensitivity analyses

We conducted a set of sensitivity analyses to evaluate how estimates of the potential COVID-19 burden change across different geographical contexts:

- Sensitivity 1: susceptibility to infection is homogeneous across all different age classes (i.e., assuming $r_a$=1 for any a in Eq.(3.2));
- Sensitivity 2: the infectiousness of individuals aged between 0 and 19 years is 50% lower compared to older individuals (i.e., assuming $\hat{r}_{\tilde{a}}$ =0.5 when $\tilde{a}$<20 and $\hat{r}_{\tilde{a}} = 1$ for $\tilde{a} \geq$20 in Eq.(3.2)) and that susceptibility to infection is heterogeneous by age, as defined for the baseline analysis;
- Sensitivity 3-4: the reproduction number in the SWSZ is decreased and increased by 20% with respect to the value used in the baseline analysis, while susceptibility to infection and infectiousness is the same as in the baseline analysis.
- Sensitivity 5: the same reproduction number is used for the three geographical contexts; this sensitivity was conducted by estimating different per-contact transmission rates for each setting in such a way that the model reproduction number in each setting encompasses the estimates the distribution of the reproduction number obtained from fitting the exponential growth observed in the national surveillance data.
- Sensitivity 6: the SARS-CoV-2 transmission dynamics follows a SEIR scheme (instead of the SIR scheme adopted for our baseline analysis); the SEIR scheme was simulated by setting the adjusting factors $\alpha_I$, $\alpha_J$ and $\alpha_K$ at 0,1 and 0, in such a way that an individual, once infected, remains in the latent compartment for 3.3 days before becoming infectious (consistent with an average incubation period of 5-6 days and an average period of pre-symptomatic transmission of 2 days).

Figure 3.12 shows the estimated attack rates of infection, symptomatic cases, and critical disease in a hypothetical epidemic with school closure, by assuming that the reproduction number in the entire SWSZ is 1.62 (95%CI 1.55-1.70), as estimated from surveillance data (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*), and under the hypothesis of homogeneous susceptibility by age. Figure 3.13 shows the same quantities under the hypothesis that the infectiousness of individuals younger than 20 years of age is half of all other individuals.

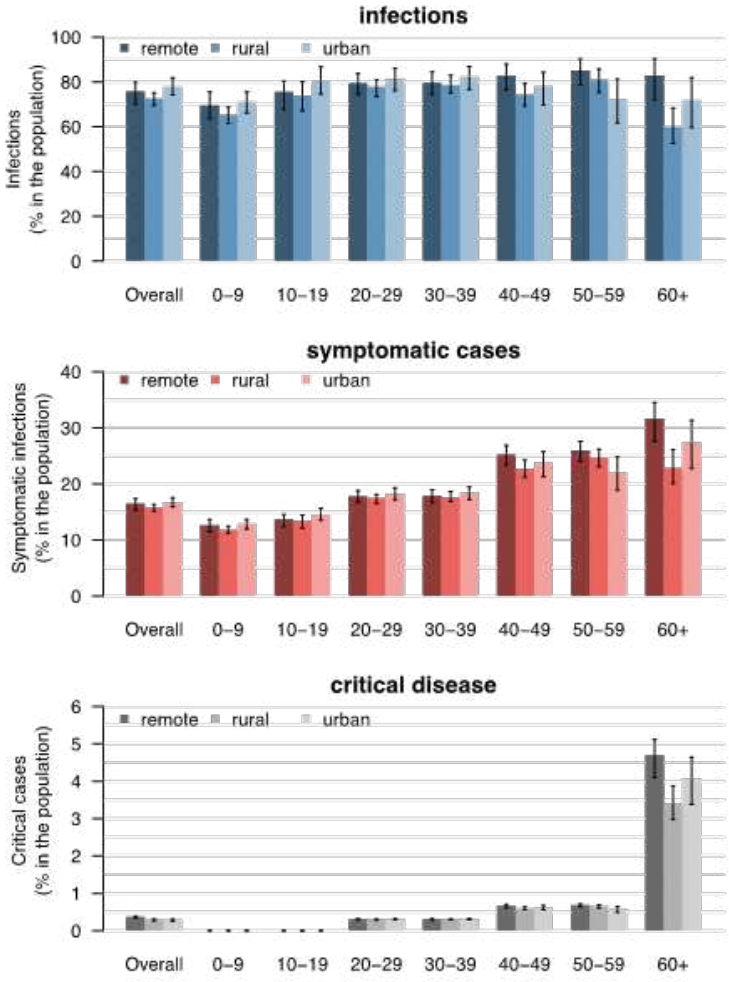


FIGURE 3.12: **Sensitivity 1.** Estimated attack rates of infection (top), symptomatic cases (middle), and critical disease (bottom), overall and by age group in different geographical contexts, as expected at the end of an epidemic mitigated by school closure alone and under the hypothesis of homogeneous susceptibility. Outputs were obtained by simulating 1,000 different epidemics where the per-contact transmission rate is set to reproduce, when neglecting contacts occurring at school, random samples of the distribution of the net reproduction number estimated from national surveillance data 1.62 (95%CI 1.55-1.70).

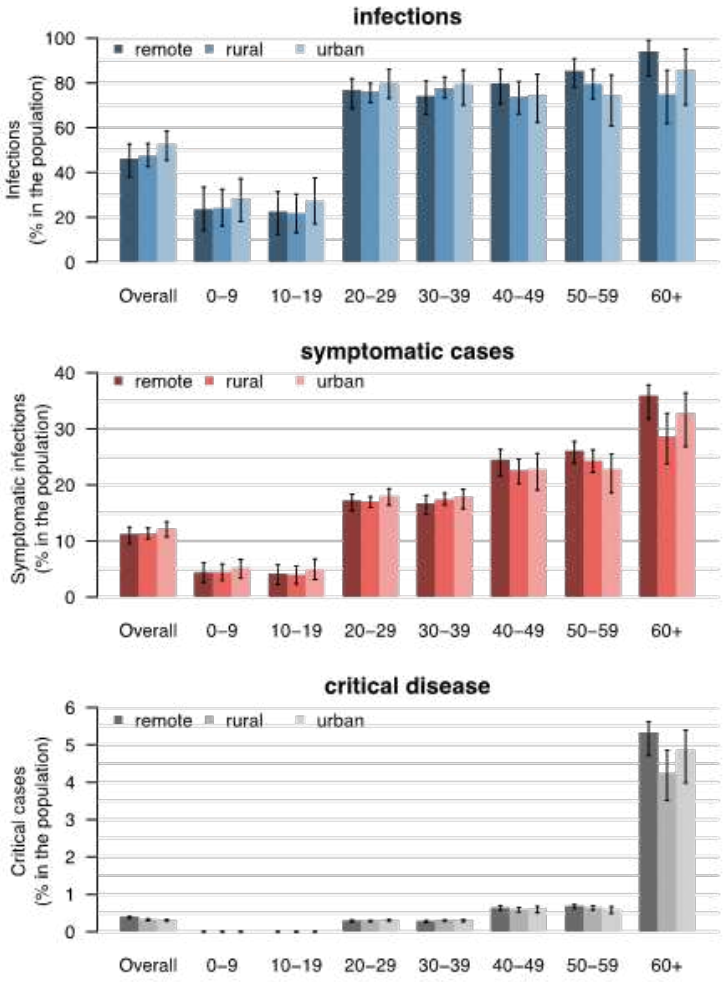


FIGURE 3.13: **Sensitivity 2.** Estimated attack rates of infection (top), symptomatic cases (middle), and critical disease (bottom), overall and by age group in different geographical contexts, as expected at the end of an epidemic mitigated by school closure alone and under the hypothesis that the infectiousness of individuals younger than 20 years of age is half of all other individuals. Outputs were obtained by simulating 1,000 different epidemics where the per-contact transmission rate is set to reproduce, when neglecting contacts occurring at school, random samples of the distribution of the net reproduction number estimated from national surveillance data 1.62 (95%CI 1.55-1.70).

Figures 3.14 and 3.15 show the estimated attack rates of infection, symptomatic cases, and critical disease in a hypothetical epidemic with school closure when the average reproduction number in the entire SWSZ is decreased or increased by 20%.

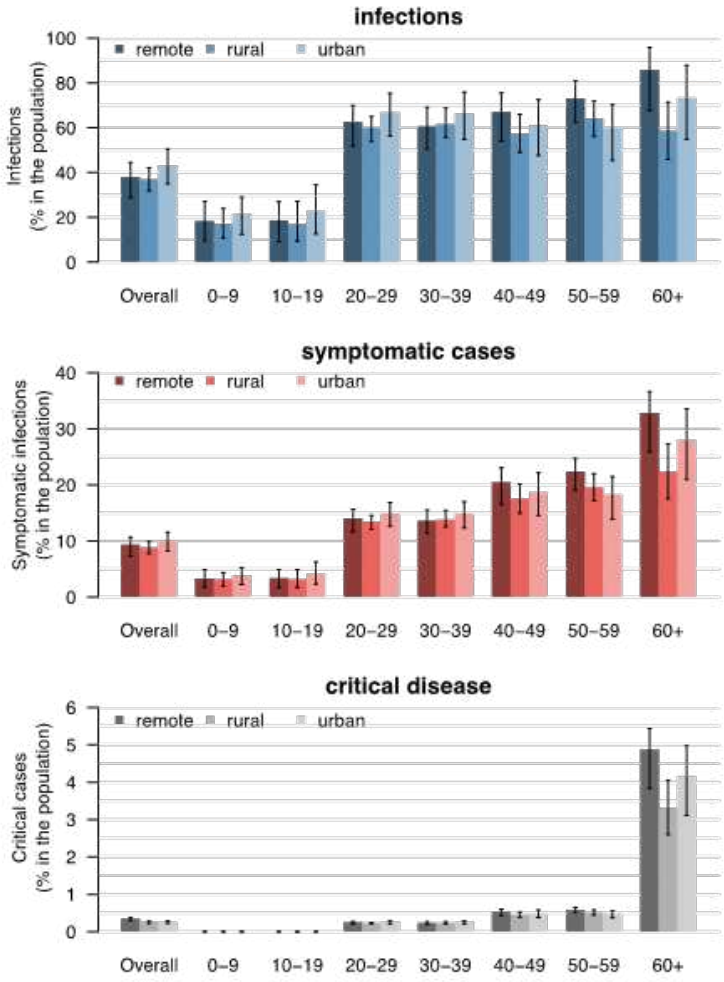


FIGURE 3.14: **Sensitivity 3.** Estimated attack rates of infection (top), symptomatic cases (middle), and critical disease (bottom), overall and by age group in different geographical, as expected at the end of an epidemic mitigated by school closure alone and under the assumption of 20% decrease of the reproduction number with respect to the baseline analysis.

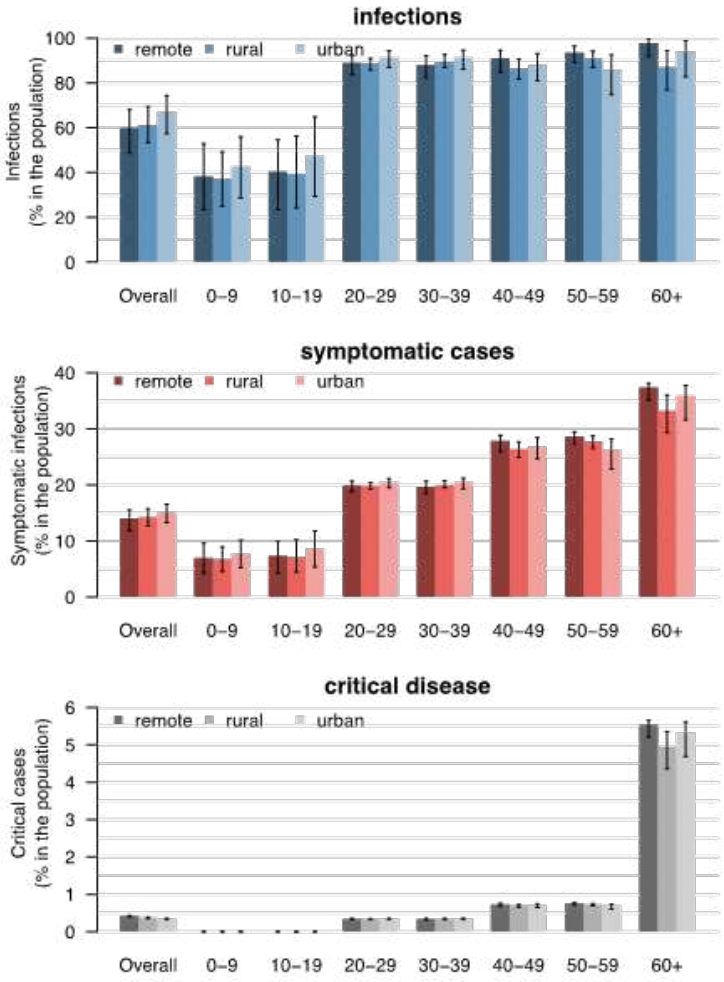


FIGURE 3.15: **Sensitivity 4.** Estimated attack rates of infection (top), symptomatic cases (middle), and critical disease (bottom), overall and by age group in different geographical, as expected at the end of an epidemic mitigated by school closure alone and under the assumption of 20% increase of the reproduction number with respect to the baseline analysis.

Figure 3.16 shows the estimated attack rates of infection, symptomatic cases, and critical disease when assuming the same reproduction number across remote, rural and urban contexts.

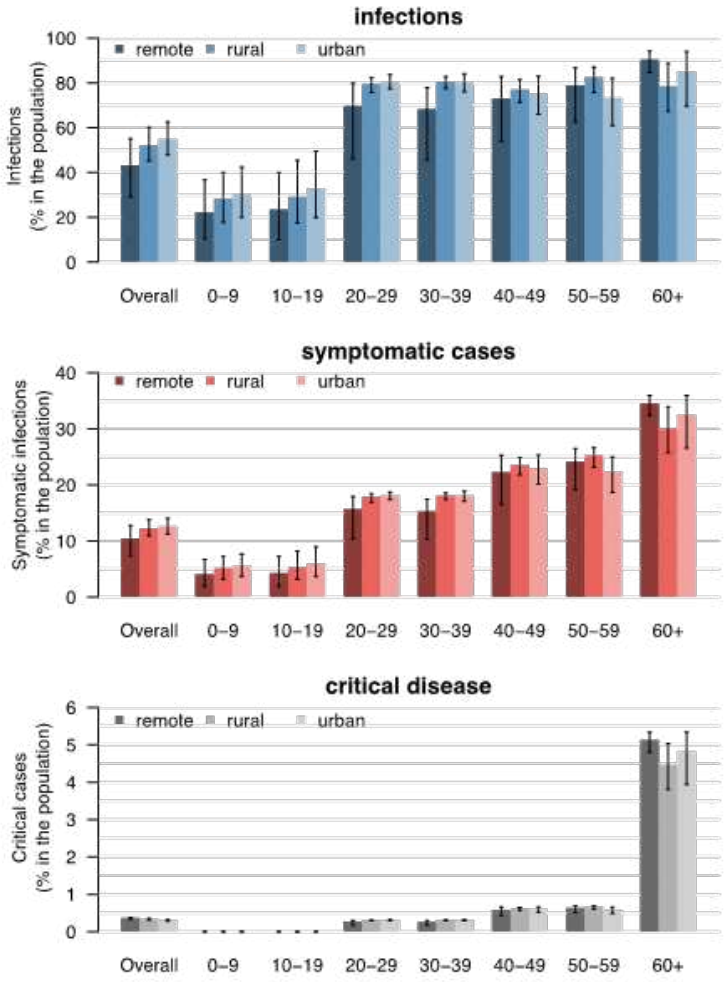


FIGURE 3.16: **Sensitivity 5.** Estimated attack rates of infection (top), symptomatic cases (middle), and critical disease (bottom), overall and by age group in different geographical, as expected at the end of an epidemic mitigated by school closure alone and under the assumption of an equal reproduction number across the three geographical contexts.

Figure 3.17 shows the epidemic curves in each geographical context simulated by using our baseline SIR model with three infectious compartment and by using an alternative SEIR scheme for the SARS-CoV-2 transmission dynamics. Estimates presented were obtained by assuming that the reproduction number in the entire SWSZ is 1.62 and by initializing the epidemic with the same number of infected individuals. No remarkable differences are observed in the patterns and magnitude of epidemic curves obtained with the two alternaticve scenarios.

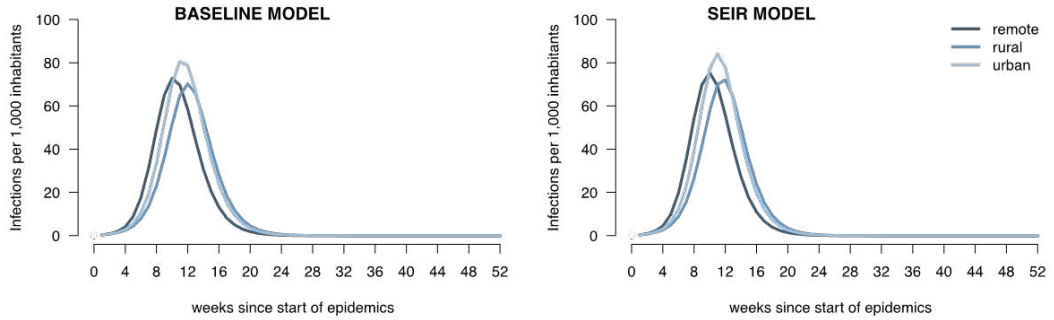


FIGURE 3.17: **Sensitivity 6.** Mean number of SARS-CoV-2 infections per 1,000 inhabitants in the three geographical contexts under a hypothetical scenario mitigated by school closure only, as simulated under SIR and SEIR schemes when assuming a reproduction number of 1.62 (95%CI 1.55-1.70).

.

# Chapter 4

# Priority ages targets for COVID-19 vaccination under limited vaccine supply: the case of South West Shewa Zone, Ethiopia

## 4.1 Introduction

After almost two years in the pandemic, the observed burden of the coronavirus disease 2019 (COVID-19) has been relatively low throughout Africa compared to high-income countries (Trentini et al., 2021). The main reasons behind this phenomenon may rely on the lack of reliable records on the real number of cases occurring in low-income settings and on the lower likelihood among young individuals of experiencing severe disease after SARS-CoV-2 infection (Trentini et al., 2021; Ofotokun and Sheth, 2021; Burki, 2021; Zardini et al., 2021). In Africa, approximately 40% of people aged less than 15 years, compared to a global mean of 26% (*Population of Africa in 2020, by age group*). The impact of COVID-19 in the African countries may have been vastly underestimated due to lacking testing capacity (Ofotokun and Sheth, 2021; Burki, 2021). For instance, a recent post-mortem surveillance study revealed that, contrary to expectations, deaths with COVID-19 were common among patients of a tertiary care referral hospital in Zambia (around 20% among deceased patients compared to less than 9% tested before death) (Mwananyanda et al., 2021).

Given the increasing spread of COVID-19 in the younger and healthier populations (Nachega et al., 2021), the repeated emergence of hyper transmissible lineages of SARS-CoV-2 (Okereke, 2021; Viana et al., 2022; Wolter et al., 2022; Gozzi et al., 2022;Abdullah et al., 2021), and the inequitable access to vaccination across countries (Loembé and Nkengasong, 2021), there is an urgent need to identify appropriate strategies for minimizing the COVID-19 burden in sub-Saharan settings. To mitigate the ongoing pandemic, unprecedented social distancing measures have been applied worldwide, including in low-income settings (Quaife et al., 2020; Van Zandvoort et al., 2020; Trentini et al., 2021; Walker et al., 2020; Tshangela et al., 2020). However, the implementation of drastic restrictions may have disproportionate effects on the already vulnerable economies of the African countries (Tshangela et al., 2020; Quaife et al., 2020; Van Zandvoort et al., 2020).

Mass immunization programs still represent the main public health strategy to reduce the burden caused by the circulation of SARS-CoV-2. While most high-income

countries have rapidly progressed in the deployment of multiple vaccine doses, as of January 7, 2022, only 9.7% of the total African population has been fully vaccinated (*Our world in data. COVID-19 Data Explorer*; *CovidVax*). At this date, Ethiopia shows the second highest cumulative number of infections and deaths of the whole African continent (*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.*). Although the national vaccination campaign was launched in March 2021 (*World Health Organization. Ethiopia introduces COVID-19 vaccine in a national launching ceremony.*), the current vaccine uptake of Ethiopia is one of the lowest of Africa, with only 1.35% of the citizens being fully immunized (*Our world in data. COVID-19 Data Explorer*; *CovidVax*). Further vulnerabilities of this country are represented by the high prevalence of comorbidities (e.g., malnutrition and tuberculosis), the limited healthcare resources, and the strong inequalities in access to primary care across geographical contexts (Poletti et al., 2018; Endris, Asefa, and Dube, 2017).

In this work, we estimate the contribution of different ages in generating SARS-CoV-2 infections and COVID-19 cases associated with most severe outcomes over different phases of the pandemic and across different areas of the South West Shewa Zone (SWSZ) of Ethiopia. The impact of alternative priority targets for vaccination is evaluated by considering different conditions of vaccine supply. To do this, we developed and simulated a transmission model for SARS-CoV-2 informed with data on age-specific mixing patterns recently collected across different geographical contexts of the SWSZ, characterized by heterogeneous population density, age structure, and access to healthcare (Trentini et al., 2021). The impact of different immunization strategies is evaluated in terms of the number of infections and critical cases that could be averted after the rollout of the national vaccination program. Obtained results could be used to identify the most effective strategies for the deployment of vaccines in emergency contexts characterized by limited vaccine supply.

## 4.2 Methods

The SARS-CoV-2 transmission dynamics is simulated by using a deterministic age-structured SIR model. Susceptibility to SARS-CoV-2 infection is assumed to vary with age according to estimates made available by Hu et al. (Hu et al., 2021). Specifically, taking the age group of 20-59 years as the reference, the relative susceptibility for individuals aged 0-19 years is set at 0.59 (95%CI: 0.35-0.92) and at 1.75 (95%CI: 1.07-2.81) for the individuals above 60 years of age. Homogenous infectiousness among individuals of different ages and an average generation time of 6.6 days are assumed.

The developed model keeps track of the contribution of infectors of different ages in causing secondary infections and cases developing severe outcomes in the population across different socio-economic contexts. Age-specific risks of developing critical disease after SARS-CoV-2 infection are taken from (Zardini et al., 2021). Critical disease cases are defined as positive individuals who would either require intensive care or likely result in a fatal outcome.

The adopted approach leverages on age-specific contact matrices recently estimated from records collected across different geographical areas of the SWSZ of the Oromia

Region, Ethiopia (Trentini et al., 2021). These sites consisted of rural villages, dispersed subsistence farming settlements, and urban neighborhoods of Woliso Town (Trentini et al., 2021; *United States Department of Agriculture. Economic Research Service.*). The model is run separately for each geographical context, assuming a constant total population. A different population age structure is considered for the three geographical settings under study (rural, remote, and urban).

The contribution of different ages in generating secondary infections and critical cases is estimated by considering two different pandemic phases. As for the first phase, lasting until March 2021, we considered the emergence of SARS-CoV-2 in a fully naïve population of individuals and analyzed the epidemic dynamics under the dominance of the historical strains of SARS-CoV-2, under a school closure mandate, and in the absence of vaccination. School closure was assumed for the entire period as this represented the prevalent restriction adopted by the national government until the vaccination program was launched. The spread of infection was simulated by considering an initial reproduction number of 1.62 (95%CI: 1.55–1.70), as estimated from the exponential growth of cases reported in Ethiopia from May to mid-June 2020 (Trentini et al., 2021). The transmission dynamics during this pandemic phase was simulated until a given proportion of the overall population gets infected. Such proportion was defined as setting specific according to serological prevalence levels estimated for March 2021 in the Jimma Zone: 31% in rural and remote sites and 45% in urban areas (Gudina et al., 2021b). Different seroprevalence values were considered for sensitivity analysis. Reliability of results obtained by the adopted approach was assessed by comparing the age distribution of the cumulative number of infections obtained by model simulation with the age distribution of infections ascertained with real-time reverse transcription polymerase chain reaction (RT-PCR) as reported between 13 March and 13 September 2021 in the Oromia Region (Gudina et al., 2021a). The second pandemic phase that we considered aims at reflecting the transmission of SARS-CoV-2 after the launch of the national vaccination program in March 2021. To explore the potential impact of COVID-19 vaccination, we simulated the SARS-CoV-2 circulation after school reopening and under the assumption that the population was partially immunized by natural infection. An initial age-specific immunity profile equal to the one obtained after running the first pandemic phase was assumed to reflect the epidemiological situation at the launch of the national immunization campaign. To account for the replacement of historical strains by hyper-transmissible variants occurred in 2021 (*Genomic epidemiology of novel coronavirus - Africa-focused subsampling*), we assumed a transmission rate mirroring a basic reproduction number of 6 (Liu et al., 2021); alternative values were considered for sensitivity analysis. The impact of different vaccination strategies on the burden of COVID-19 is assessed in terms of the potential attack rate of infections and critical cases expected after the launch of vaccination in March 2021 in the absence of restrictions on the individuals' contacts. Four different situations are analyzed. First, we consider a scenario where the number of administered vaccines is negligible, and we evaluate the impact of pre-existing immunity levels on the disease spread. Given the low vaccine uptake recorded in Ethiopia, this scenario may reflect what might have occurred in the first months following the launch of vaccination. Second, we assume that a limited number of doses is available, and we investigate whether the most effective strategy to mitigate the disease spread is either to vaccinate only the individuals over 50 years old, mimicking the initial priority target

defined by the Ethiopian vaccination program (*The World Bank. World Bank Open Data.*), or to vaccinate all the individuals eligible for vaccination (>10 years of age). Specifically, simulations obtained by assuming a specific coverage level in the narrow age target (>50 years of age) are compared with simulations obtained when the same number of doses are uniformly administered to the larger share of the population (e.g., >10 years). Third, we assume that all individuals older than 50 years have already completed the vaccination and we project the potential impact of expanding vaccination to other age groups. In this case, the impact of administering the vaccine only to individuals aged 30-50 years is compared with a scenario where the corresponding number of doses is uniformly distributed to all eligible ages (10-50 years). Finally, to provide a comprehensive view of the overall benefits of vaccination, we consider a variety of combinations of coverage values among subjects over 50 years and individuals aged between 10 and 50 years, irrespectively of the number of doses required to achieve the considered targets.

In the model, vaccinated individuals are defined as subjects who received two doses of vaccine and who therefore experience a lower risk of infection and of developing severe outcomes (Marziano et al., 2021b; Harris et al., 2021; Subbarao et al., 2021; Sheikh et al., 2021; Thiruvengadam et al., 2021; Pouwels et al., 2021; Falsey et al., 2021). The SARS-CoV-2 infectiousness of breakthrough infections (i.e., infections occurring among vaccinee) is assumed to be reduced by 50% (Marziano et al., 2021b). Since ChAdOx1 nCoV-19 was the dominant vaccine employed in Ethiopia during 2021, the vaccine efficacy against infection and critical diseases is set at 65% and 71.5%, respectively (Subbarao et al., 2021; Sheikh et al., 2021; Thiruvengadam et al., 2021; Pouwels et al., 2021; Falsey et al., 2021; Andrews et al., 2021). In a sensitivity analysis, different values for the vaccine efficacy against the infection and the critical disease were considered to reflect either the use of more effective vaccines or possible immune escape phenomena from vaccination led by recently emerged variants (Viana et al., 2022; Wolter et al., 2022; Abdullah et al., 2021; Gozzi et al., 2022) .

Epidemiological transitions are modeled by the following system of ordinary differential equations:

$$\begin{cases} \dot{S}_a = -r_a S_a \sum_{\tilde{a}} \lambda_{a,\tilde{a}} \\ \dot{S}^v_a = -(1 - VE^{\text{inf}}) r_a S^v_a \sum_{\tilde{a}} \lambda_{a,\tilde{a}} \\ \dot{I}_{a,\tilde{a}} = r_a \lambda_{a,\tilde{a}} S_a - \gamma I_{a,\tilde{a}} \\ \dot{I}^v_{a,\tilde{a}} = (1 - VE^{\text{inf}}) r_a \lambda_{a,\tilde{a}} S^v_a - \gamma I^v_{a,\tilde{a}} \\ \dot{R}_{a,\tilde{a}} = \gamma I_{a,\tilde{a}} \\ \dot{R}^v_{a,\tilde{a}} = \gamma I^v_{a,\tilde{a}} \end{cases}$$

where $a$ defines the chronological age of the individuals, $r_a$ is the relative susceptibility in the age class $a$, $S_a$ represents susceptible individuals who have never been vaccinated, $S^v_a$ represents vaccinated individuals who experienced a reduced force of infection, $VE^{\text{inf}}$ is the vaccine efficacy against the infection, $I_{a,\tilde{a}}$ and $I^v_{a,\tilde{a}}$ are the individuals of age $a$ among unvaccinated and vaccinated individuals who get the infection upon contacts with individuals of age $\tilde{a}$, $R_{a,\tilde{a}}$ and $R^v_{a,\tilde{a}}$ represent the number of recovered individuals from these two classes of infections, $1/\gamma$ is the average duration of the infectivity period, and $\lambda_{a,\tilde{a}}$ is the contribution of age $\tilde{a}$ to the force of

infection experienced by susceptible individuals of age $a$. The latter is defined as

$$\lambda_{a,\tilde{a}} = M_{a,\tilde{a}}\left(\beta\frac{\sum_{\tilde{a}} I_{a,\tilde{a}}}{N_{\tilde{a}}} + \frac{\beta}{2}\frac{\sum_{\tilde{a}} I^{v}_{a,\tilde{a}}}{N_{\tilde{a}}}\right)$$

where $M_{a,\tilde{a}}$ represents the average number of daily contacts that an individual of age class $a$ has with persons of age group $\tilde{a}$, $N_{\tilde{a}}$ is the total population in the age class $\tilde{a}$ , and $\beta$ is the SARS-CoV-2 transmission rate.

## 4.3 Results

### 4.3.1 SARS-CoV-2 transmission in the pre-vaccination period

The age distribution of the infections estimated by the model under the assumption of a fully susceptible population and the school closure mandate well compares with the one of SARS-CoV-2 infections ascertained in the Oromia Region between March and September 2021 (Gudina et al., 2021a) (Figure 4.1 **A**). The retrospective reconstruction of infections that occurred from the beginning of the pandemic to March 2021 suggests that, in all considered sites, the proportion of individuals who gained natural immunity during the first pandemic phase was markedly higher among subjects aged over 60 years: ranging from 47.7% (95%CI: 32.6-68.0%) in rural area to 73.5% (95%CI: 49.9-94.1%) in the remote settlements (Figure 4.1 **B**). Our estimates also suggest that the level of natural immunity acquired under 60 years of age was relatively higher in urban neighborhoods compared to other settings. According to our simulations, SARS-CoV-2 transmission during the first pandemic year might have been mainly assortative, i.e., characterized by a similar age between the infectors and their secondary cases (see Figure 4.2). Nonetheless, we estimated that during this period the highest fraction of SARS-CoV-2 infections was caused by infectors under 30 years: 46.1-58.7% across all the considered socio-economic contexts. Remarkably, in remote settlements, 48.7% of overall infections occurred over 60 years of age might have been caused by infected individuals of similar age. Accordingly, assortative contacts between the elderlies in remote settlements might have caused a higher percentage of critical cases compared to what estimated for the other geographical contexts (35.7% among individuals aged >60 years vs 9.5% and 9.1% in the rural and the urban settings respectively, see Figure 4.2). This may be partially explained by the older population structure characterizing less urbanized populations, and the higher number of community contacts between individuals aged 60 years and subjects with a similar age identified for this socio-economic context (see Figure 4.7 and 4.8).

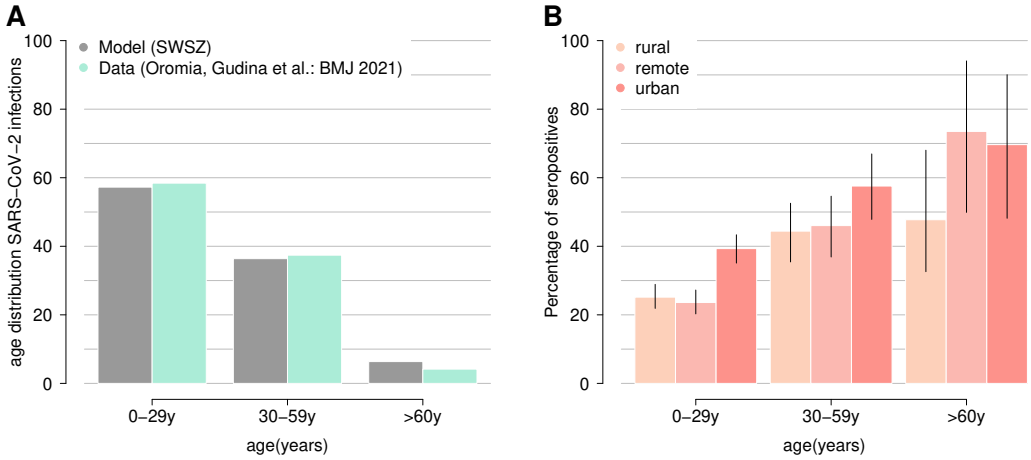


FIGURE 4.1: **A** Comparison between the age distribution of all confirmed cases reported between 13 March and 13 September 2021 in the Oromia Region (Gudina et al., 2021a) and that of the cumulative infections as obtained with a model mimicking the school closure and the achievement of the immunity profile estimated by Gudina et al. (Gudina et al., 2021b). **B** Estimated age-specific percentage of the population who acquired natural immunity to SARS-CoV-2 at the beginning of the vaccination campaign (March 2021) in rural, remote, and urban areas of the SWSZ.

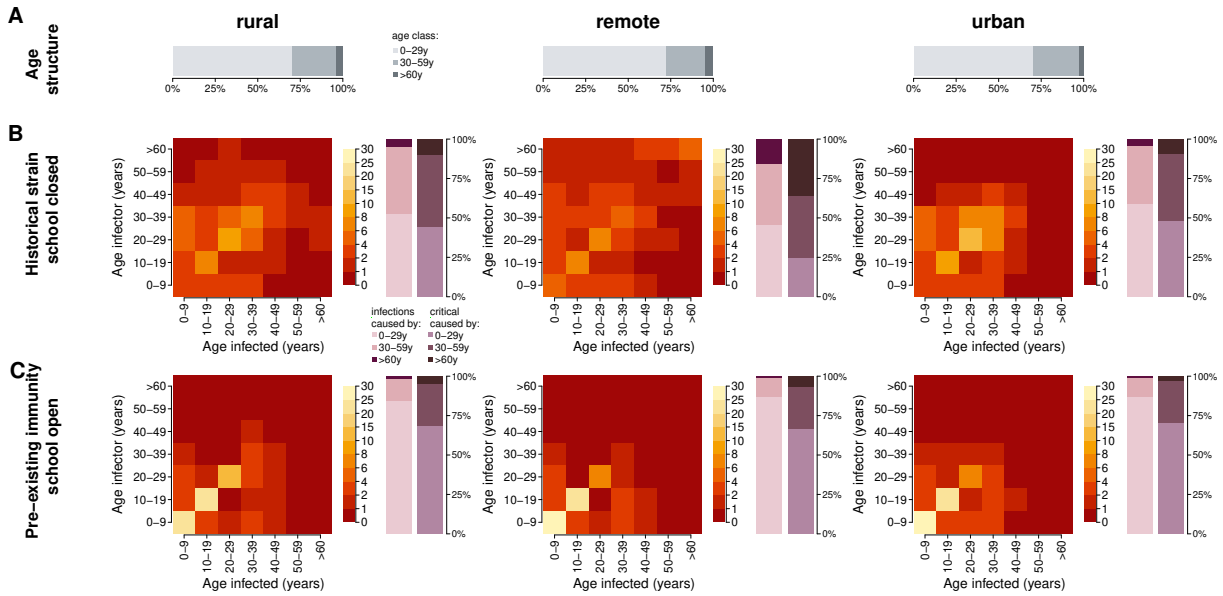


FIGURE 4.2: Age distribution of the population residing in rural villages, remote settlements, and urban neighbourhoods **A**. Matrices representing the estimated average contribution of different ages in the spread of SARS-CoV-2; bar plots represent the corresponding overall proportion of infections and critical cases caused by different age groups of infectors (0-29, 30-59, >60 years) in the pre-vaccination era **B** and at the launch of the national vaccination program **C**. The two scenarios are simulated by considering an average basic reproduction of 3 and 6, respectively.

### 4.3.2 SARS-CoV-2 transmission at vaccination launch

To mimic the COVID-19 epidemiology during the first months following the launch of the national vaccination program, we simulated the SARS-CoV-2 transmission by assuming that the vaccine uptake achieved in the entire population was negligible. However, pre-existing immunity levels as estimated for March 2021 were fully considered and an increased viral transmissibility was assumed to account for the emergence of new hyper-transmissible lineages of SARS-CoV-2. Our results suggest that the natural immunity acquired in the first pandemic phase and the reopening of teaching activities have redrawn the contribution of different ages in the spread of SARS-CoV-2 infection (see Figure 4.2 **C**). Specifically, we found that, after March 2021, the contribution of individuals under 30 years in causing new infections and critical cases might have increased to 84.5-87.3% and 66.7-70.6% respectively (see Figure 4.2 **C**). Accordingly, the contribution of the elderly in generating SARS-CoV-2 secondary infections and critical cases was rebounded to 0.8-1.3% and to 3.1-6.6%, respectively.

### 4.3.3 The expected epidemiological outcomes under different vaccine uptake levels and vaccination priority targets

We compared the attack rates of infection and critical cases expected under two epidemiological scenarios. In the first one, we assumed that all the individuals aged 50

years or more have been vaccinated according to the initial priority target defined by the national vaccination program (*The World Bank. World Bank Open Data.*) but that none of the younger individuals received the vaccine. In the second one, we assumed that the corresponding number of vaccine doses have been randomly distributed throughout all ages eligible for vaccination. Both scenarios were simulated by considering the immunity acquired from natural infection before the launch of the vaccination program and the increased transmissibility associated with SARS-CoV-2 lineages emerged in 2021. Our findings suggest that, with a limited vaccine supply, the better strategy to reduce the potential burden of critical disease is to prioritize vaccination of older individuals (see Figure 4.3). On the one hand, we found that the vaccination of 100% of individuals older than 50 years of age has the potential of reducing the attack rate of critical disease to 0.13% (95%CI: 0.11-0.15%), 0.11% (95%CI: 0.09-0.13%), and 0.09% (95%CI: 0.07-0.12%) in rural, remote, and urban areas, respectively. On the other hand, we estimated that if the same number of vaccines are randomly administered to individuals older than 10 years of age, the fraction of the population developing a critical disease is expected to be 0.20% (95%CI: 0.16-0.23%) in rural, 0.15% (95%CI: 0.10-0.21%) in remote, and 0.12% (95%CI: 0.08-0.16%) in urban areas. As concerns the reduction of the number of infections, the effect of these two alternative vaccination strategies is almost equivalent, with differences in the attack rates ranging from 0.5% to 1.1% across the three geographical contexts (see Figure 4.3).

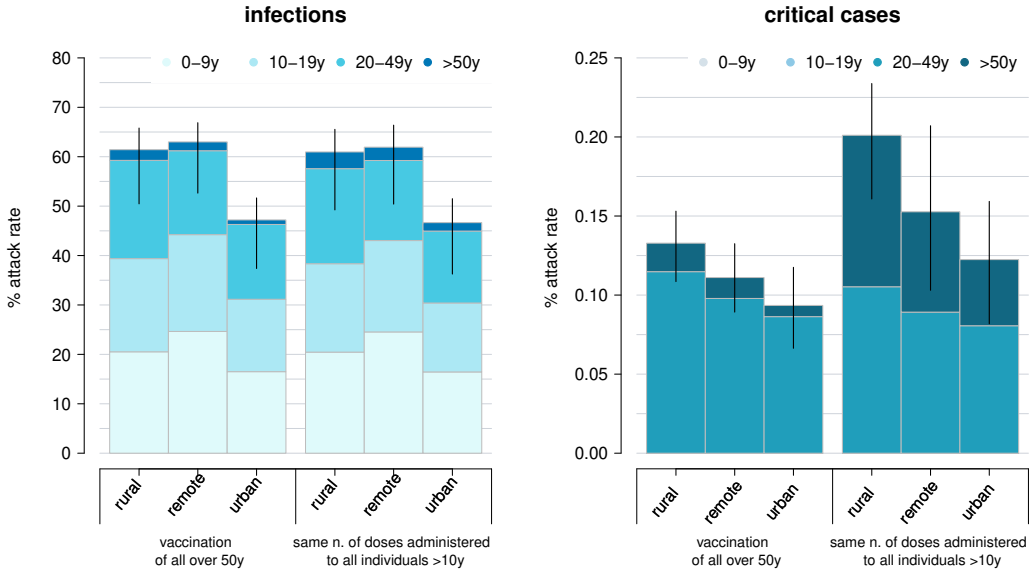


FIGURE 4.3: Estimated attack rates of infections and critical cases, in each site (rural, remote, and urban), stratified by age groups (0-9, 10-19, 20-49, >50 years), as obtained under the assumption that all the individuals over 50 years are vaccinated and under in the hypothetical scenario where the corresponding number of vaccine doses are randomly distributed throughout the population over 10 years. Coloured bars represent average estimates, stratified by the age of infected individuals; solid lines represent the 95% CI of model estimates.

We projected the potential impact of an expansion of the national vaccination campaign, under the assumption that all individuals over 50 years of age have been fully vaccinated. The effect of an enhanced vaccine uptake on the COVID-19 burden is investigated by comparing simulations where the vaccination of individuals aged 30-50 years is prioritized with model estimates obtained when an equal number of subjects is randomly targeted by vaccination among individuals aged 10-50 years (see Figure 4.4). We found that the most effective strategy to reduce the attack rate of infections is to distribute the vaccines to all individuals aged 10-50 years. However, to decrease the cumulative number of critical cases, the best strategy remains the prioritization of the older segments of the population (in this case, individuals between 30 and 50 years of age). Assuming as a reference the number of doses corresponding to 50% coverage among subjects aged 30-50 years, the beneficial impact of using these doses to randomly vaccinate individuals between 10 and 50 years of age is quantified in an additional 477, 481, and 996 averted infections per 100,000 residents in the rural, the remote, and the urban contexts, respectively. Reversely, the average number of critical patients that could be further averted by targeting only subjects aged 30-50 years of age would consist of 12, 11, and 7 per 100,000 residents in rural, remote, and urban contexts respectively.

According to our simulations, the average attack rate of infections could decrease from 47.2-63.0% obtained when only individuals aged more than 50 years are targeted by vaccination to 41.7-57.7% when 100% of coverage is reached in all the individuals above 30 years. The corresponding average attack rate of critical cases would almost halve (from 0.09-0.13% to 0.05-0.07%). If the same number of doses would be used to vaccinate all the individuals above 50 years of age and uniformly vaccinate the residual eligible subjects (i.e., 10-50 years), the expected average attack rate of SARS-CoV-2 infection and of critical disease would decrease to 39.2-56.3% and to 0.06-0.09%, respectively.

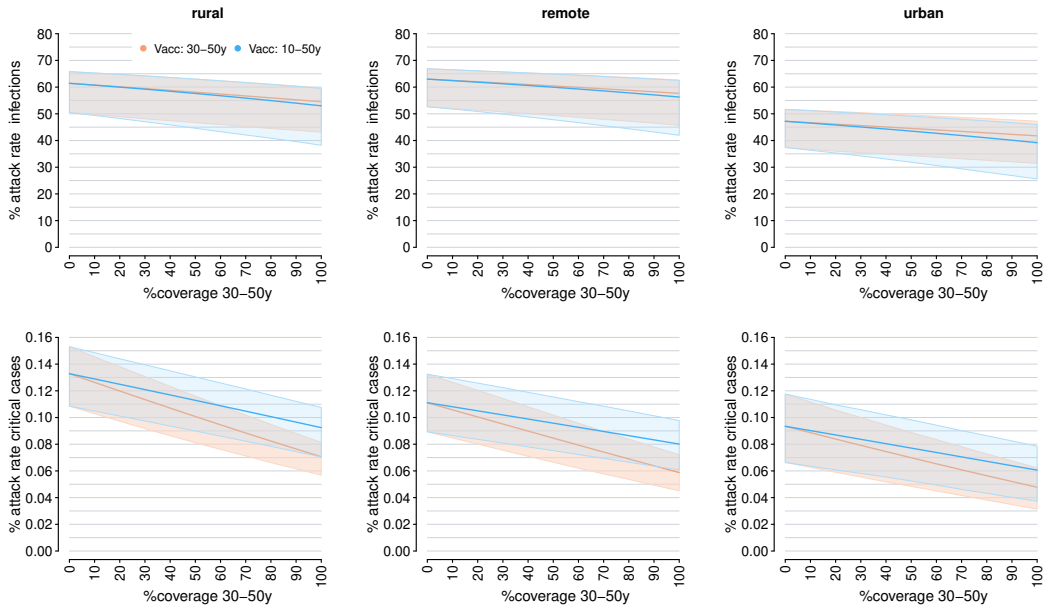


FIGURE 4.4: Estimated attack rates of infections (first row) and critical cases (second row), in each site (rural, remote, and urban), obtained under the assumption that all individuals over 50 are fully immunized and considering different scenarios in the number of available vaccine doses, which was computed exploring different coverages among subjects aged 30-50 years. We assessed the impact of two vaccination strategies, involving individuals aged 30-50 years (orange line) or the entire remaining vaccinable population (10-50 years, blue line). Lines show the mean model estimates while shaded areas represent the 95% credible interval.

To illustrate the full potential of COVID-19 vaccination, we estimated the attack rate of infections and critical cases under different combinations of vaccination coverage in the elderly ($\geq$50 years of age) and between 10-50 years of age, irrespectively to possible limits in the vaccine supply (see Figure 4.5). Obtained results corroborate that to reduce the number of infections, the vaccination of younger subjects is required. However, our estimates suggest that the vaccination with 2 doses of the entire population over 10 years would not be enough to decrease the reproduction number below the critical epidemic threshold of 1, therefore suggesting that the deployment of booster doses is required to interrupt the SARS-CoV-2 circulation in Ethiopia.

Our simulations highlight that to reduce the potential attack rate of infection under 35%, all the individuals over 50 years and at least 20% of subjects under this age should be vaccinated in remote settlements (see Figure 4.5 **A**). In rural villages, the same achievement could be reached by either vaccinating 100% of the people under 50 years of age or by reaching a 90% coverage in the people under 50 years and at least a 10% coverage among the over 50. In urban areas, 60% of coverage among individuals under 50 years may be sufficient to obtained similar results. When assuming a vaccination coverage of 100% among the elderly, the lowest attack rate of critical cases is expected to occur in urban neighborhoods where 9.3 (95%CI: 6.6-11.8)

subjects per 10,000 individuals are estimated to be exposed to COVID-19 critical disease (see Figure 4.5 **B**). To reduce the number of critical cases around such a level of incidence in remote and rural areas, the strategy minimizing the number of administered doses requires the vaccination of 90-100% of the individuals older than 50 years and at least 20-30% of younger individuals. To further reduce the attack rate of critical disease under 4 cases per 10,000 individuals in remote settlements and rural villages, a 90-100% vaccination coverage over 50 years of age should be complemented with more than 80% coverage among younger eligible subjects. In urban neighborhoods, the same target would require 100% coverage among the elderly and 60% coverage in younger individuals. Remarkably, if the maximum uptake levels achieved among the elderly would be 80%, the vaccination of at least 100%, 90%, and 70% of the individuals under 50 years would be required to reduce the attack rate of critical disease under 4 cases per 10,000 individuals in rural, remote, and urban areas, respectively. Despite a higher infection attack rate may characterize the remote settlements, under all the considered combinations of coverage levels, the highest prevalence of critical cases is expected in rural areas, where a lower proportion of the elderly might have been naturally immunized during the first pandemic year.

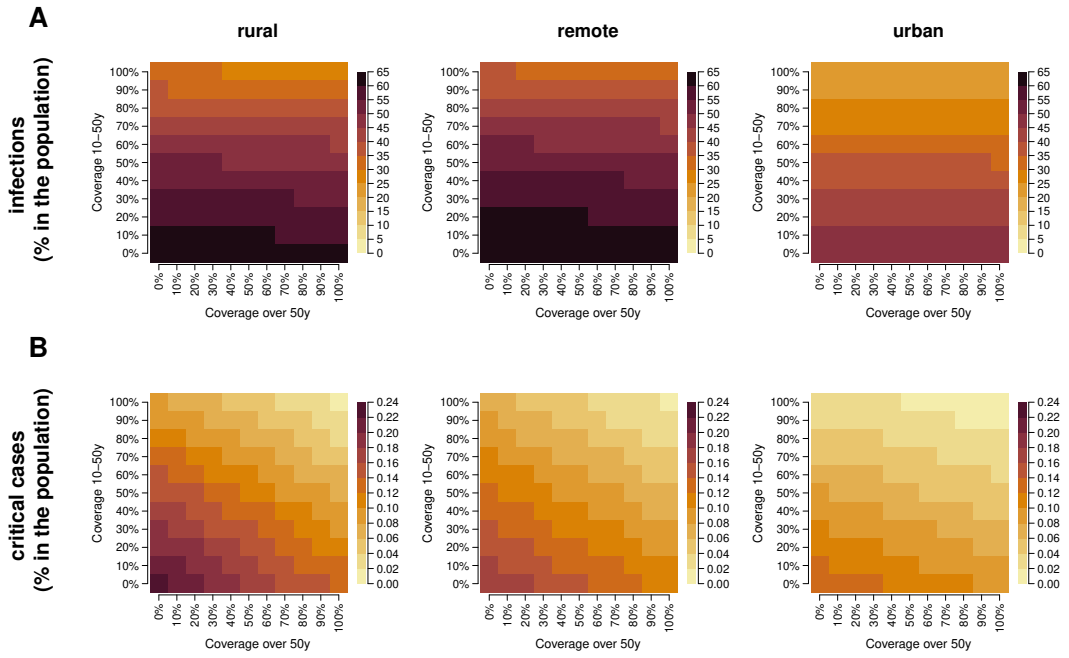


FIGURE 4.5: Attack rates of infections **A** and critical cases **B** as estimated for rural, remote, and urban areas for different combinations of coverage levels for individuals older than 50 years of age and younger individuals when assuming an $R_0$ equal to 6 and a generation time of 6.6 days. Natural immunity in rural, remote, and urban areas are set at 31%, 31%, and 45%, respectively.

### 4.3.4 Sensitivity analyses

We performed a set of sensitivities analyses to evaluate the impact on our findings of the model assumptions we made regarding the SARS-CoV-2 basic reproduction

number ($R_0$), the overall level of natural immunity acquired during the first year of the pandemic, and the vaccine efficacy against infection and critical disease (see figure 4.6). The conduction of these sensitivities aimed at exploring the potential impact on our estimates of the emergence of hyper transmissible lineages, possible changes in restrictions implemented by the government, alternative infection rates experienced during the first pandemic phase, different values of the vaccine efficacy possibly led by the waning of natural immunity, by the deployment of alternative vaccine products, or by immune escape phenomena associated with SARS-CoV-2 variants emerged in late 2021 (Viana et al., 2022; Wolter et al., 2022; Gozzi et al., 2022; Abdullah et al., 2021).
We found that, in the case that 100% of individuals older than 50 years of age get vaccinated, a 15% increase in the transmissibility of the virus could result in an increase in the overall attack rate of critical cases by 3.6-6.5%. For the same scenario, 15% lower transmissibility would result in a reduction in the estimated attack of critical disease of 6.3-9.7%. When assuming a higher initial natural immunity (38% in rural and in remote, 53% in urban), the expected attack rate of critical disease decreases by 24.8-26.1% across the different settings. The same quantity is expected to increase by 26.3-29.7% when considering lower initial immunity levels (22% in rural and in remote, 40% in urban). Finally, while a more effective vaccine could reduce the fraction of individuals developing critical outcomes by 3.2-6.0%, possible immune escape or the progressive waning of natural immunity could increase this fraction by 10.8-17.3%.

Results from the sensitivity analysis under alternative vaccination scenarios are presented in the Figures 4.9.

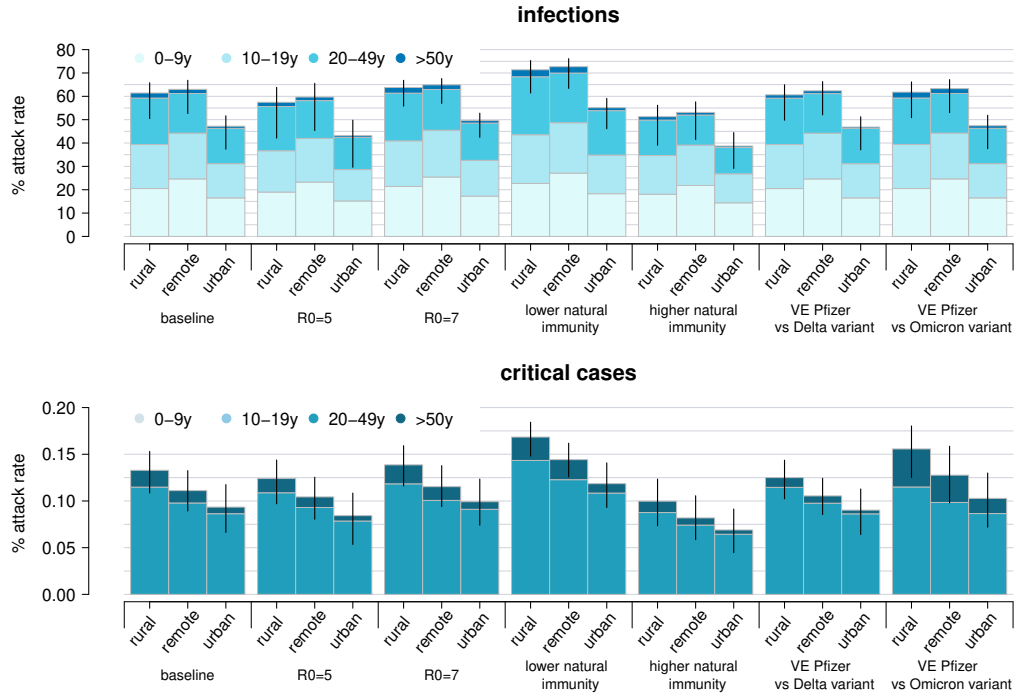


FIGURE 4.6: Comparison of the attack rate of infections and critical cases in different geographical contexts of the SWSZ when vaccines are administered to all subjects older than 50 years as obtained under our baseline assumptions and in the different sensitivity analyses carried out. Bars represent average model estimates. Black lines represent the corresponding 95%CI. Different colors are used to highlight the average fraction of cases expected across different age bands.

## 4.4 Discussion

In this study, we adopted an age-stratified modeling approach to estimate the potential transmission of SARS-CoV-2 between different ages over time in rural, remote and urban areas of the South West Shewa Zone of Ethiopia. Obtained results suggest that before the launch of the national vaccination program, the infection transmission might have been mainly assortative, i.e. involving social interactions between individuals of similar age. In particular, we estimated that, in remote settlements, around half of the infections among the individuals over 60 years was caused by interactions with subjects of similar age. More in general, we found that, on average, the elderly might have been responsible of 35.7%, 9.5% , and 9.1% of all critical cases occuring across all ages in remote, rural and urban settings, respectively.On the other hand, we found that a pivotal role in the spread of SARS-CoV-2 was played by subjects under 30 years, who might have been responsible for about half of the infections in all the considered areas. Our simulations also suggest that during the first months after the launch of the vaccination, the natural immunity acquired in the first pandemic phase and the reopening of schools significantly decreased the contribution of the elderly to the transmission of infection, increasing the proportion of critical cases caused by younger infections.

Beyond possible phenomena of vaccine hesitancy, a limited vaccine supply should be considered when exploring the impact of vaccination strategies against COVID-19 in sub-Saharan African countries. Our estimates highlighted that - after almost two years in the pandemic - prioritizing the vaccination of the older segments of the population remains the best strategy to minimize the burden of critical illness in low-income settings. These results emerged irrespectively to the overall number of available doses and despite the estimated contribution played by young individuals in the spread of the disease. Our findings therefore confirmed the results of analyses conducted in early 2021 across different countries under the assumption of an unlimited vaccine supply (e.g., (Marziano et al., 2021b; "Grad, and Daniel B. Larremore. 2021."Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serostatus.""; Yang et al., 2021).
Presented results should be carefully interpreted because of the following limitations. First, in our model school closure is the only intervention considered to estimate the age-specific immunity before the launch of the national vaccination program. This means that changes in the transmission intensity caused by variations of the social distancing measures and restrictions policies adopted until March 2021 were not considered (including an initial suspension of nonessential productive activities in 2020 (Trentini et al., 2021) and the erratic re-opening of schools for short time periods (Scott et al., 2021). Although our estimates well compare with the age distribution of infections ascertained until March 2021, the circulation of SARS-CoV-2 after this date and the waning of natural immunity have likely altered the current immunity profiles in the South West Shewa Zone. To better highlight the overall potential of different vaccination strategies, SARS-CoV-2 transmission was simulated under the hypothetical scenario of an unmitigated COVID-19 epidemic. Thus, our estimates of the expected number of infections and critical cases might be considered as illustrative worst-case scenarios. In fact, the absolute burden of COVID-19 in the considered settings would strongly depend on the level of restrictions that will be adopted during the rollout of vaccination. Finally, because of the lack of direct data from Africa, the relative susceptibility, the age-specific risks of developing a critical disease, and the potential increased transmissibility and immune escape associated with novel SARS-CoV-2 variants were assumed from evidence gathered in other countries (Hu et al., 2021; Zardini et al., 2021; Liu et al., 2021).

Despite these limitations, the major strengths of the presented analysis rely on the investigation of the relative impact of prioritizing different ages for vaccination under a limited vaccine supply and on the provided estimates of the contribution of different ages groups in causing severe COVID-19 cases over different pandemic phases. In this regard, our results highlight that social distancing measures should focus on reducing contacts between the elderly (representing the most vulnerable individuals) and individuals younger than 30 years of age (representing the most common infectors in the spread of the disease).

## 4.5 Supplementary Figures

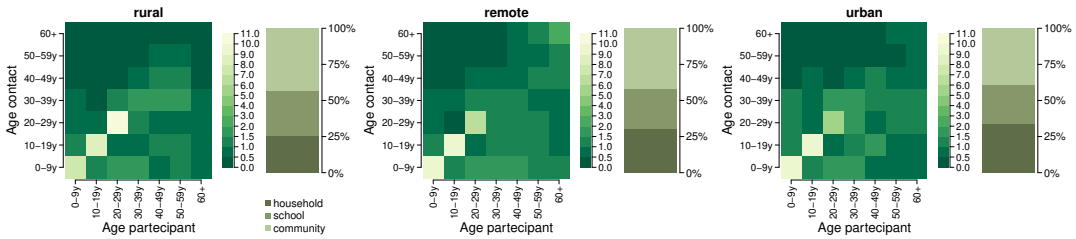


FIGURE 4.7: Contact matrices representing the mean number of daily contacts reported by a participant in the age group i with individuals in the age group j in each site (rural, remote, and urban). The bar plots show the percentage of contacts that occurred in each setting (household, school, and community).

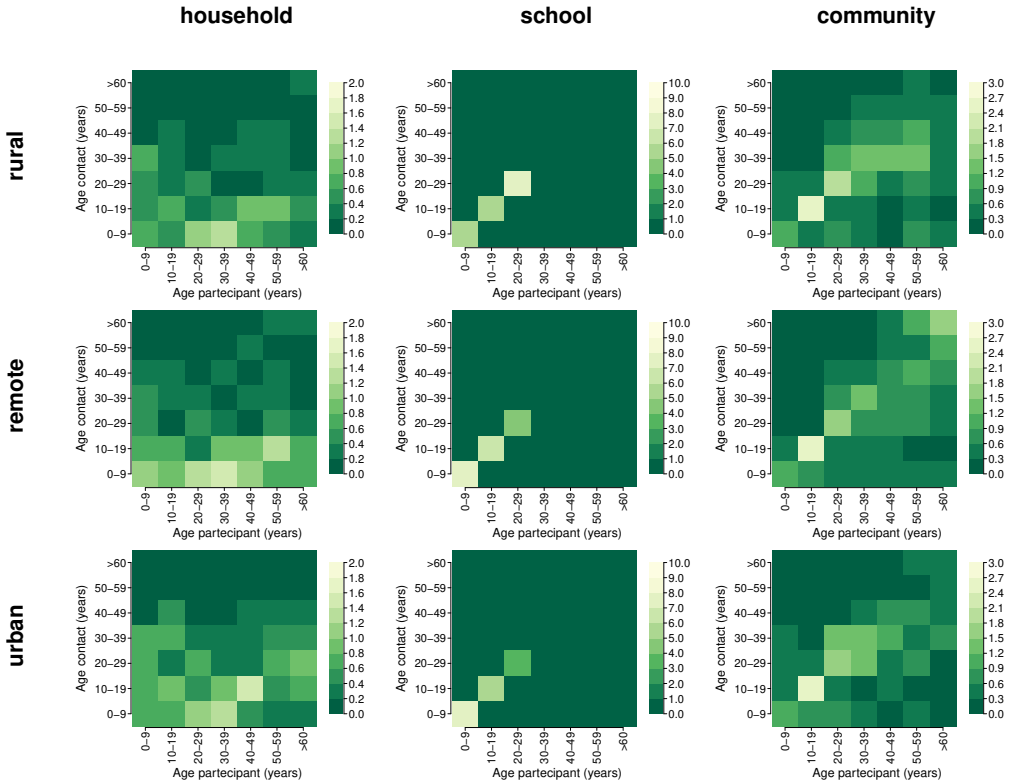


FIGURE 4.8: Contact matrices representing the mean number of daily contacts reported by a participant in the age group i with individuals in the age group j in each setting (household, school, and community) and site (rural, remote, and urban).

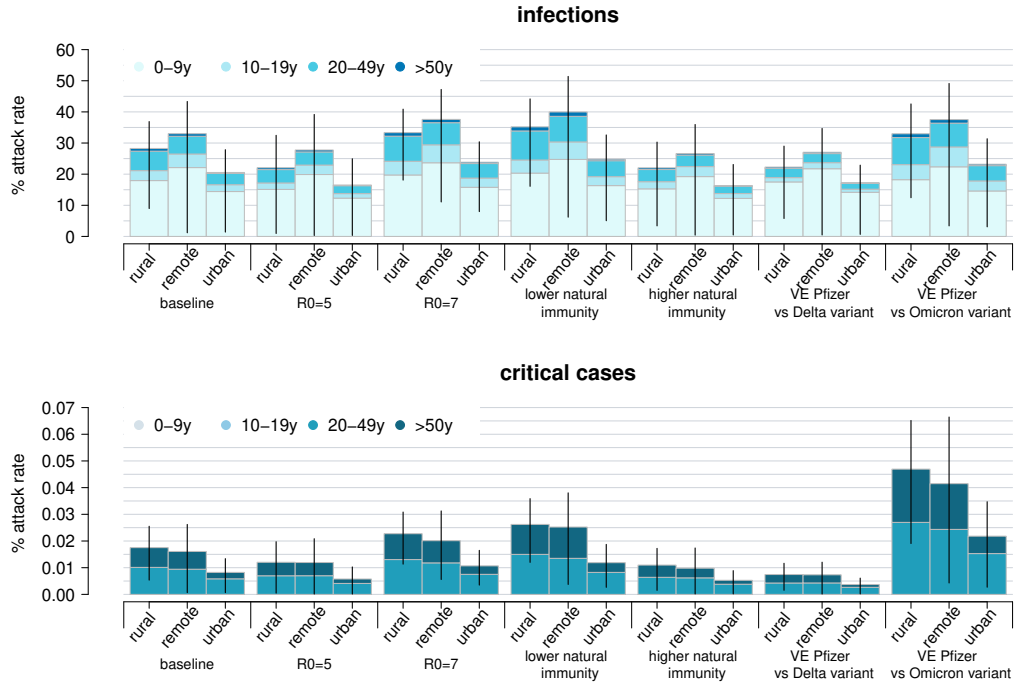


FIGURE 4.9: Comparison of the estimated overall percentage of infections and critical cases in different geographical contexts of the SWSZ in the baseline scenario (vaccines are administered to all subjects older than 10 years) and sensitivity analyses. Black lines represent the 95% credible intervals.

# Chapter 5

# Conclusion

The design of appropriate public health policies in response to epidemic phenomena requires to be informed with reliable estimates of the potential impact of the infection on the individual health and the robust evaluation about the possible effect of alternative preventive and control measures. The latter could be assisted by the analysis of model simulations based on data-driven approaches which investigate the potential epidemic trajectories expected under different scenarios reflecting time varying epidemiological conditions (e.g., the emergence of new viral variants), public restrictions (e.g. school closure and lockdowns) and vaccinations policies.

Epidemiological data collected from disease cases identified in the population from passive surveillance mainly consist of records associated with individuals seeking care. As such, the myopic observation of surveillance data may lead to an erroneous assessment of the circulation of an infectious pathogen in the community, preventing the identification of the real risks associated with the infection and hampering a correct estimation of the potential impact of the viral spread on public health systems. The problem is particularly relevant for infectious diseases characterized by a larger share of asymptomatic infections, as it is the case for COVID-19. To overcome this issue and to better understanding of the effect of the infection spread at the population level, a careful selection of unbiased sample data of the infected individuals is needed.

For many viral diseases, including COVID-19, vaccination represents the key tool to reduce the risk of severe outcomes caused by the infection transmission and the consequent pressure on health care systems, avoiding the side effects led by strict restriction regimes on the social-economic system of modern societies. Unfortunately, the ongoing pandemic has highlighted a marked geographical heterogeneity in the access to resources required to face epidemic threats. In fact, while the implementation of prolonged lockdowns and the suspension of working activities is not sustainable in countries with vulnerable economies, so far, vaccination in Africa has reached coverage levels well below initial expectations. As a consequence, beyond possible phenomena of vaccine hesitancy in the population, a limited vaccine supply should be considered when exploring the potential effectiveness of public policies to counter COVID-19 disease spread in sub-Saharan Africa. An illustrative example is provided by what has occurred in Ethiopia. Although a national campaign for COVID-19 vaccination was launched in March 2021 (*World Health Organization. Ethiopia introduces COVID-19 vaccine in a national launching ceremony.*), the vaccine uptake level in this country is one of the lowest of Africa, with only 1.35% of the citizens being fully immunized as of January 7, 2022 (*CovidVax*; *Our world in data. COVID-19 Data Explorer*).

On the one hand, this thesis deals with the estimation of parameters necessary to evaluate the clinical course of individuals after the infection with SARS-CoV-2 (the viral pathogen responsible of COVID-19). Estimated parameters could be used to inform model analysis on the potential burden of COVID-19 across different geographical settings. On the other hand, the presented work aims at highlighting how social mixing and demographic factors can influence the transmission of SARS-CoV-2 and the emergence of severe disease cases across different demographic and economic contexts.

The research activity presented in Chapter 2 highlighted that the severity of the SARS-CoV-2 virus is strongly correlated with age. Specifically, provided estimates suggest that being older than 60 years of age is associated with about 40% likelihood of developing respiratory symptoms or fever $\geq$37.5C after SARS-CoV-2 infection, that about 1% of the infections occurring in this age group may either require intensive care or may be at risk of death. The carried out analysis also highlighted that during the first Italian wave of infection, the median length of stay in hospital and in intensive care units were 10 and 11 days, respectively. The main innovation represented by the presented work is related to the approach adopted to derive such estimates, which allowed the minimization of the risks of bias in the identification of infections and the provision of a comprehensive quantitative assessment of all the main epidemiological parameters essential to model COVID-19 burden.

The analysis presented in Chapter 3 relied on the development and simulation of transmission model informed with realistic data on social contact patterns characterizing different settings of Ethiopia. Presented results suggest that the lower COVID-19 burden observed in sub-Saharan Africa during the first year of the pandemic compared to that estimated for high-income countries may be strongly related to a combination of younger population age-structures and the prolonged adoption between spring and autumn 2020 of school closures to counter the infection spread. Remarkably, the carried-out analysis also highlights that socio-demographic factors can determine marked geographical heterogeneities in the expected disease burden within the same region.

Estimates provided in Chapter 4 suggest that the highest fraction of SARS-CoV-2 infections arise from the interaction of schoolchildren and young adults with individuals of similar age. After estimating the age-specific immunity profile that could have characterized different populations at the launch of the national vaccination program, the study highlights that vaccination of the elderly is the best strategy to reduce the number of critical patients led by SARS-CoV-2 transmission. Presented results also show that prioritizing younger ages may represent the best approach to reduce the infection spread but that vaccination based on 2 doses only may be insufficient to interrupt the transmission in the population.

# Bibliography


Abdullah, F et al. (2021). "Decreased severity of disease during the first global omicron variant covid-19 outbreak in a large hospital in tshwane, south africa". In: *International Journal of Infectious Diseases*.

Andrews, N et al. (2021). "Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant." In: *The New England Journal of Medicine* 385.7, pp. 585–594.

Bhatraju, Pavan K et al. (2020). "Covid-19 in critically ill patients in the Seattle region—case series". In: *New England Journal of Medicine* 382.21, pp. 2012–2022.

Bi, Qifang et al. (2020). "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study". In: *The Lancet Infectious Diseases* 20.8, pp. 911–919.

Biggerstaff, Matthew et al. (2020). "Early insights from statistical and mathematical modeling of key epidemiologic parameters of COVID-19". In: *Emerging infectious diseases* 26.11.

Bonelli, Fabrizio et al. (2020). "Clinical and analytical performance of an automated serological test that identifies S1/S2-neutralizing IgG in COVID-19 patients semiquantitatively". In: *Journal of Clinical Microbiology* 58.9, e01224–20.

Brand, Samuel PC et al. (2020). "Forecasting the scale of the COVID-19 epidemic in Kenya". In: *MedRxiv*.

Brisson, M et al. (2000). "Modelling the impact of immunization on the epidemiology of varicella zoster virus". In: *Epidemiology & Infection* 125.3, pp. 651–669.

Bubar, Kate M et al. "Grad, and Daniel B. Larremore. 2021."Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serostatus."" In: *Science. https://doi.org/10.1126/science. abe6959* ().

Buitrago-Garcia, Diana et al. (2020). "Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis". In: *PLoS medicine* 17.9, e1003346.

Burki, Talha Khan (2021). "Undetected COVID-19 cases in Africa". In: *The Lancet Respiratory Medicine* 9.12, e121.

Byambasuren, Oyungerel et al. (2020). "Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis". In: *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada* 5.4, pp. 223–234.

Cereda, Diletta et al. (2020). "The early phase of the COVID-19 outbreak in Lombardy, Italy". In: *arXiv preprint arXiv:2003.09320*.

Chinazzi, Matteo et al. (2020). "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak". In: *Science* 368.6489, pp. 395–400.

Cohen, Andrew N and Bruce Kessel (2020). "False positives in reverse transcription PCR testing for SARS-CoV-2". In: *MedRxiv*.

Cohen, Jacob (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Corman, Victor M et al. (2020). "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR". In: *Eurosurveillance* 25.3, p. 2000045.

*CovidVax*. URL: https://covidvax.live/continent/africa.

Dare, Anna J et al. (2015). "Deaths from acute abdominal conditions and geographical access to surgical care in India: a nationally representative spatial analysis". In: *The Lancet Global Health* 3.10, e646–e653.

Davies, Nicholas G et al. (2020). "Age-dependent effects in the transmission and control of COVID-19 epidemics". In: *Nature medicine* 26.8, pp. 1205–1211.

Davies, Nicholas G et al. (2021). "Increased mortality in community-tested cases of SARS-CoV-2 lineage B. 1.1. 7". In: *Nature* 593.7858, pp. 270–274.

Di Domenico, Laura et al. (2020). "Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies". In: *BMC medicine* 18.1, pp. 1–13.

Diekmann, Odo, Johan Andre Peter Heesterbeek, and Johan AJ Metz (1990). "On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations". In: *Journal of mathematical biology* 28.4, pp. 365–382.

Dowd, Jennifer Beam et al. (2020). "Demographic science aids in understanding the spread and fatality rates of COVID-19". In: *Proceedings of the National Academy of Sciences* 117.18, pp. 9696–9698.

Emery, Jon C et al. (2020). "The contribution of asymptomatic SARS-CoV-2 infections to transmission on the Diamond Princess cruise ship". In: *Elife* 9, e58699.

Endris, Neima, Henok Asefa, and Lamessa Dube (2017). "Prevalence of malnutrition and associated factors among children in rural Ethiopia". In: *BioMed research international* 2017.

*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 37 weekly bulletin.* URL: https://www.ephi.gov.et/images/novel_coronavirus/EPHI_PHEOC_COVID-19_Weekly_Bulletin_37_English_01192021.pdf.

*Ethiopian Institute of Public Health. COVID-19 pandemic preparedness and response in Ethiopia - 6 weekly bulletin.* URL: https://www.ephi.gov.et/images/novel_coronavirus/EPHI_-PHEOC_COVID-19_Weeklybulletin_6_English_06082020.pdf.

*Eurostat. Population Structure and Ageing.* URL: https://ec.europa.eu/eurostat/statisticsexplained/index.php/Population_structure_and_ageing#The_share_of_elderly_people_continues_to_increase.

Faes, Christel et al. (2020). "Time between symptom onset, hospitalisation and recovery or death: statistical analysis of Belgian COVID-19 patients". In: *International journal of environmental research and public health* 17.20, p. 7560.

Falsey, Ann R et al. (2021). "Phase 3 safety and efficacy of AZD1222 (ChAdOx1 nCoV-19) Covid-19 vaccine". In: *New England Journal of Medicine* 385.25, pp. 2348–2360.

Ferguson, Neil et al. (2020). "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand". In: URL: https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19.

Fu, Leiwen et al. (2020). "Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: a systematic review and meta-analysis". In: *Journal of Infection* 80.6, pp. 656–665.

*Genomic epidemiology of novel coronavirus - Africa-focused subsampling*. URL: https://nextstrain.org/ncov/gisaid/africa.

Ghisolfi, Selene et al. (2020). "Predicted COVID-19 fatality rates based on age, sex, comorbidities and health system capacity". In: *BMJ global health* 5.9, e003094.

Gilbert, Marius et al. (2020). "Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study". In: *The Lancet* 395.10227, pp. 871–877.

Gozzi, Nicolò et al. (2022). "Preliminary modeling estimates of the relative transmissibility and immune escape of the Omicron SARS-CoV-2 variant of concern in South Africa". In: *medRxiv*.

Grasselli, Giacomo et al. (2020). "Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy". In: *Jama* 323.16, pp. 1574–1581.

Guan, Wei-jie et al. (2020). "Clinical characteristics of coronavirus disease 2019 in China". In: *New England journal of medicine* 382.18, pp. 1708–1720.

Guanghong, Ding et al. (2004). "SARS epidemical forecast research in mathematical model". In: *Chinese Science Bulletin* 49.21, pp. 2332–2338.

Gudina, Esayas Kebede et al. (2021a). "COVID-19 in Oromia Region of Ethiopia: a review of the first 6 months' surveillance data". In: *BMJ open* 11.3, e046764.

Gudina, Esayas Kebede et al. (2021b). "Seroepidemiology and model-based prediction of SARS-CoV-2 in Ethiopia: longitudinal cohort study among front-line hospital workers and communities". In: *The Lancet Global Health* 9.11, e1517–e1527.

Gupta, Sunetra, Neil Ferguson, and Roy Anderson (1998). "Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents". In: *Science* 280.5365, pp. 912–915.

Guzzetta, Giorgio et al. (2020). "Potential short-term outcome of an uncontrolled COVID-19 epidemic in Lombardy, Italy, February to March 2020". In: *Eurosurveillance* 25.12, p. 2000293.

Guzzetta, Giorgio et al. (2021). "Impact of a nationwide lockdown on SARS-CoV-2 transmissibility, Italy". In: *Emerging infectious diseases* 27.1, p. 267.

Haileamlak, Abraham (2018). "How can Ethiopia mitigate the health workforce gap to meet universal health coverage?" In: *Ethiopian journal of health sciences* 28.3, p. 249.

Hamer, William Heaton (1906). *Epidemic disease in England: the evidence of variability and of persistency of type*. Bedford Press.

Harris, Ross J et al. (2021). "Effect of Vaccination on Household Transmission of SARS-CoV-2 in England". In: *New England Journal of Medicine*.

Hellewell, Joel et al. (2020). "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts". In: *The Lancet Global Health* 8.4, e488–e496.

Hilton, Joe and Matt J Keeling (2020). "Estimation of country-level basic reproductive ratios for novel Coronavirus (SARS-CoV-2/COVID-19) using synthetic contact matrices". In: *PLoS computational biology* 16.7, e1008031.

Horby, Peter et al. (2011). "Social contact patterns in Vietnam and implications for the control of infectious diseases". In: *PloS one* 6.2, e16965.

Hu, Shixiong et al. (2021). "Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China". In: *Nature communications* 12.1, pp. 1–11.

*International Monetary Fund. Policy responses to Covid-19. 2020.* URL: https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19.

*Istituto Superiore di Sanità, 2021*. URL: https://www.epicentro.iss.it/coronavirus/open-data/covid_19-iss.xlsx.AccessedApril29,2021.

*Italian National Institute of Statistics. Demographic indicators.* URL: Availablefrom: http://dati.istat.it/Index.aspx?DataSetCode=DCIS_INDDEMOG1&Lang=en.AccessedMarch22,2021.

*Italian National Institute of Statistics2020. Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV-2*. URL: Available:https://www.istat.it/it/files/2020/08/ReportPrimiRisultatiIndagineSiero.pdf.

Keeling, Matt J and Pejman Rohani (2011). *Modeling infectious diseases in humans and animals*. Princeton university press.

Kermack, William Ogilvy and Anderson G McKendrick (1927). "A contribution to the mathematical theory of epidemics". In: *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115.772, pp. 700–721.

Kiem, Cécile et al. (2020). "Evaluation des stratégies vaccinales COVID-19 avec un modèle mathématique populationnel". PhD thesis. Haute Autorité de Santé; Institut Pasteur Paris; Santé publique France.

Kiem, Cécile Tran et al. (2021). "A modelling study investigating short and medium-term challenges for COVID-19 vaccination: From prioritisation to the relaxation of measures". In: *EClinicalMedicine* 38, p. 101001.

Kiti, Moses Chapa et al. (2014). "Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya". In: *PloS one* 9.8, e104786.

Kucharski, Adam J et al. (2020). "Early dynamics of transmission and control of COVID-19: a mathematical modelling study". In: *The lancet infectious diseases* 20.5, pp. 553–558.

Lau, EHY et al. (2020). "Temporal dynamics in viral shedding and transmissibility of COVID-19". In: *Annual Scientific Meeting 2020, Hong Kong College of Community Medicine, Hong Kong*.

Lavezzo, Enrico et al. (2020). "Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'". In: *Nature* 584.7821, pp. 425–429.

Li, Long-quan et al. (2020). "COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis". In: *Journal of medical virology* 92.6, pp. 577–583.

Li, Sheng et al. (2017). "Demographic transition and the dynamics of measles in six provinces in China: A modeling study". In: *PLoS medicine* 14.4, e1002255.

Liu, Hengcong et al. (2021). "Herd immunity induced by COVID-19 vaccination programs to suppress epidemics caused by SARS-CoV-2 wild type and variants in China". In: *medRxiv*.

Loembé, Marguerite Massinga and John N Nkengasong (2021). "COVID-19 vaccine access in Africa: Global distribution, vaccine platforms, and challenges ahead". In: *Immunity* 54.7, pp. 1353–1362.

Loembé, Marguerite Massinga et al. (2020). "COVID-19 in Africa: the spread and response". In: *Nature Medicine* 26.7, pp. 999–1003.

Longini Jr, Ira M and M Elizabeth Halloran (1996). "A frailty mixture model for estimating vaccine efficacy". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 45.2, pp. 165–173.

Ma, Shujuan et al. (2020). "Epidemiological parameters of COVID-19: case series study". In: *Journal of medical Internet research* 22.10, e19994.

Makoni, Munyaradzi (2020). "COVID-19 in Africa: half a year later". In: *The Lancet Infectious Diseases* 20.10, p. 1127.

Marziano, Valentina et al. (2021a). "Retrospective analysis of the Italian exit strategy from COVID-19 lockdown". In: *Proceedings of the National Academy of Sciences* 118.4.

Marziano, Valentina et al. (2021b). "The effect of COVID-19 vaccination in Italy and perspectives for living with the virus". In: *Nature communications* 12.1, pp. 1–8.

Mbow, Moustapha et al. (2020). "COVID-19 in Africa: Dampening the storm?" In: *Science* 369.6504, pp. 624–626.

McCombs, Audrey and Claus Kadelka (2020). "A model-based evaluation of the efficacy of COVID-19 social distancing, testing and hospital triage policies". In: *PLoS computational biology* 16.10, e1008388.

Melegaro, Alessia et al. (2017). "Social contact structures and time use patterns in the Manicaland Province of Zimbabwe". In: *PloS one* 12.1, e0170459.

Mohammed, Hussen et al. (2020). "Containment of COVID-19 in Ethiopia and implications for tuberculosis care and research". In: *Infectious Diseases of Poverty* 9.1, pp. 1–8.

Munayco, César V et al. (2020). "Early transmission dynamics of COVID-19 in a southern hemisphere setting: Lima-Peru: February 29th–March 30th, 2020". In: *Infectious Disease Modelling* 5, pp. 338–345.

Muniz-Rodriguez, Kamalich et al. (2020). "Severe acute respiratory syndrome coronavirus 2 transmission potential, Iran, 2020". In: *Emerging infectious diseases* 26.8, p. 1915.

Murthy, Srinivas, Aleksandra Leligdowicz, and Neill KJ Adhikari (2015). "Intensive care unit capacity in low-income countries: a systematic review". In: *PloS one* 10.1, e0116949.

Mwananyanda, Lawrence et al. (2021). "Covid-19 deaths in Africa: prospective systematic postmortem surveillance study". In: *bmj* 372.

Nachega, Jean B et al. (2021). "Addressing challenges to rolling out COVID-19 vaccines in African countries". In: *The Lancet Global Health* 9.6, e746–e748.

Nikolai, Lea A et al. (2020). "Asymptomatic SARS Coronavirus 2 infection: Invisible yet invincible". In: *International Journal of Infectious Diseases*.

Ofotokun, Igho and Anandi N Sheth (2021). "Africa's COVID-19 Experience—A Window of Opportunity to Act". In: *JAMA Network Open* 4.9, e2124556–e2124556.

Okereke, Melody (2021). "Spread of the delta coronavirus variant: Africa must be on watch". In: *Public Health in Practice (Oxford, England)* 2, p. 100209.

Onder, Graziano, Giovanni Rezza, and Silvio Brusaferro (2020). "Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy". In: *Jama* 323.18, pp. 1775–1776.

Oran, Daniel P and Eric J Topol (2020). "Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review". In: *Annals of internal medicine* 173.5, pp. 362–367.

*Our world in data. COVID-19 Data Explorer*. URL: `https://ourworldindata.org/explorers/coronavirus-data-explorer`.

O'Driscoll, Megan et al. (2021). "Age-specific mortality and immunity patterns of SARS-CoV-2". In: *Nature* 590.7844, pp. 140–145.

Park, Minah et al. (2020). "A systematic review of COVID-19 epidemiology based on current evidence". In: *Journal of clinical medicine* 9.4, p. 967.

Peiris, Joseph Sriyal Malik et al. (2003). "Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study". In: *The Lancet* 361.9371, pp. 1767–1772.

Perez-Saez, Javier et al. (2021). "Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland". In: *The Lancet Infectious Diseases* 21.4, e69–e70.

Poletti, Piero et al. (2018). "The hidden burden of measles in Ethiopia: how distance to hospital shapes the disease mortality rate". In: *BMC medicine* 16.1, pp. 1–12.

Poletti, Piero et al. (2020a). "Age-specific SARS-CoV-2 infection fatality ratio and associated risk factors, Italy, February to April 2020". In: *Eurosurveillance* 25.31, p. 2001383.

Poletti, Piero et al. (2020b). "Probability of symptoms and critical disease after SARS-CoV-2 infection". In: *arXiv preprint arXiv:2006.08471*.

Poletti, Piero et al. (2021). "Association of age with likelihood of developing symptoms and critical disease among close contacts exposed to patients with confirmed sars-cov-2 infection in italy". In: *JAMA network open* 4.3, e211085–e211085.

Pollán, Marina et al. (2020). "Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study". In: *The Lancet* 396.10250, pp. 535–544.

*Population of Africa in 2020, by age group*. URL: `https://www.statista.com/statistics/1226211/population-of-africa-by-age-group/`.

Pouwels, Koen B et al. (2021). "Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK". In: *Nature medicine* 27.12, pp. 2127–2135.

Prem, Kiesha, Alex R Cook, and Mark Jit (2017). "Projecting social contact matrices in 152 countries using contact surveys and demographic data". In: *PLoS computational biology* 13.9, e1005697.

Quaife, Matthew et al. (2020). "The impact of COVID-19 control measures on social contacts and transmission in Kenyan informal settlements". In: *BMC medicine* 18.1, pp. 1–11.

Riccardo, Flavia et al. (2020). "Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020". In: *Eurosurveillance* 25.49, p. 2000790.

Saad-Roy, Chadi M et al. (2020). "Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years". In: *Science* 370.6518, pp. 811–818.

Salje H Tran Kiem C, Lefrancq N Courtejoie N Bosetti P Paireau J Andronico A Hozé N Richet J Dubost CL Le Strat Y Lessler J Levy-Bruhl D Fontanet A Opatowski L Boelle PY Cauchemez S (2020). "Estimating the burden of SARS-CoV-2 in France". In: *Science* 368.6498.

Scott, Nick et al. (2021). "Modelling the impact of relaxing COVID-19 control measures during a period of low viral transmission". In: *Medical Journal of Australia* 214.2, pp. 79–83.

Sheikh, Aziz et al. (2021). "SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness". In: *The Lancet*.

Stefanelli, Paola et al. (2021). "Prevalence of SARS-CoV-2 IgG antibodies in an area of northeastern Italy with a high incidence of COVID-19 cases: a population-based study". In: *Clinical Microbiology and Infection* 27.4, 633–e1.

Subbarao, Sathyavani et al. (2021). "Vaccine Effectiveness Against Infection and Death Due to SARS-CoV-2, Following One and Two Doses of the BNT162b2 and ChADox-1 in Residents of Long-Term Care Facilities in England, Using a Time-Varying Proportional Hazards Model". In: *Following One and Two Doses of the BNT162b2 and ChADox-1 in Residents of Long-Term Care Facilities in England, Using a Time-Varying Proportional Hazards Model*.

*The World Bank. World Bank Open Data.* URL: https://data.worldbank.org/indicator/SP.DYN.CDRT.IN?locations=ET.

Thiruvengadam, Ramachandran et al. (2021). "Effectiveness of ChAdOx1 nCoV-19 vaccine against SARS-CoV-2 infection during the delta (B. 1.617. 2) variant surge in India: a test-negative, case-control study and a mechanistic study of post-vaccination immune responses". In: *The Lancet Infectious Diseases*.

Trentini, Filippo et al. (2020). "Healthcare strain and intensive care during the COVID-19 outbreak in the Lombardy region: A retrospective observational study on 43,538 hospitalized patients". In: *medRxiv*.

Trentini, Filippo et al. (2021). "Modeling the interplay between demography, social contact patterns, and SARS-CoV-2 transmission in the South West Shewa Zone of Oromia Region, Ethiopia". In: *BMC medicine* 19.1, pp. 1–13.

Tshangela, A et al. (2020). "COVID-19 in Africa: the spread and response." In: *Nature Medicine*.

*UNICEF. Ethiopia COVID-19 Situation Report No. 3. 2020.* URL: https://www.unicef.org/ethiopia/sites/unicef.org.ethiopia/files/2020-04/UNICEF%20Ethiopia%20COVID-19%20Situation%20Report%20No.%203%20-%2027%20March-3%20April%202020.pdf.

*United Nations Department of Economic and Social Affairs. 2019 UN World Population Prospects.* URL: https://population.un.org/wpp/DataQuery/.

*United States Department of Agriculture. Economic Research Service.* URL: https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/.

Van Zandvoort, Kevin et al. (2020). "Response strategies for COVID-19 epidemics in African settings: a mathematical modelling study". In: *BMC medicine* 18.1, pp. 1–19.

Verity, Robert et al. (2020). "Estimates of the severity of coronavirus disease 2019: a model-based analysis". In: *The Lancet infectious diseases* 20.6, pp. 669–677.

Vespignani, Alessandro et al. (2020). "Modelling covid-19". In: *Nature Reviews Physics* 2.6, pp. 279–281.

Viana, Raquel et al. (2022). "Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa". In: *Nature*, pp. 1–10.

Viner, Russell M et al. (2021). "Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis". In: *JAMA pediatrics* 175.2, pp. 143–156.

Volz, Erik et al. (2021). "Assessing transmissibility of SARS-CoV-2 lineage B. 1.1. 7 in England". In: *Nature* 593.7858, pp. 266–269.

Walker, Patrick GT et al. (2020). "The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries". In: *Science* 369.6502, pp. 413–422.

Wallinga, Jacco and Marc Lipsitch (2007). "How generation intervals shape the relationship between growth rates and reproductive numbers". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1609, pp. 599–604.

Wallinga, Jacco and Peter Teunis (2004). "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures". In: *American Journal of epidemiology* 160.6, pp. 509–516.

Waroux, O le Polain de et al. (2018). "Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact: a survey in Southwest Uganda". In: *BMC infectious diseases* 18.1, pp. 1–12.

Wolter, Nicole et al. (2022). "Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study". In: *The Lancet*.

Wood, Simon N et al. (2021). "COVID-19 and the difficulty of inferring epidemiological parameters from clinical data". In: *The Lancet Infectious Diseases* 21.1, pp. 27–28.

*World Health Organization. Contact tracing in the context of COVID-19: interim guidance, 10 May 2020.* URL: https://apps.who.int/iris/handle/10665/332049.

*World Health Organization. Covid-19 Response Bullettin Ethiopia. 2020.* URL: https://www.afro.who.int/sites/default/files/2020-04/ETHIOPIA_COVID-19%20response%20bulletin_10APR2020%20%28002%29.pdf.

*World Health Organization. Ethiopia introduces COVID-19 vaccine in a national launching ceremony.* URL: https://www.afro.who.int/news/ethiopia-introduces-covid-19-vaccine-national-launching-ceremony.

*World Health organization. Health Workforce Requirements for Universal Health Coverage and the Sustainable Development Goals. Human Resource for Health Observers Series No. 17.* URL: https://apps.who.int/iris/bitstream/handle/10665/250330/9789241511407-eng.pdf.

*World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. 2020.* URL: https://covid19.who.int/WHO-COVID-19-global-data.csv.

Wu, Joseph T, Kathy Leung, and Gabriel M Leung (2020). "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study". In: *The Lancet* 395.10225, pp. 689–697.

Wu, Joseph T et al. (2020). "Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China". In: *Nature medicine* 26.4, pp. 506–510.

Yang, Juan et al. (2020). "Disease burden and clinical severity of the first pandemic wave of COVID-19 in Wuhan, China". In: *Nature communications* 11.1, pp. 1–10.

Yang, Juan et al. (2021). "Despite vaccination, China needs non-pharmaceutical interventions to prevent widespread outbreaks of COVID-19 in 2021". In: *Nature Human Behaviour*, pp. 1–12.

Zardini, Agnese et al. (2021). "A quantitative assessment of epidemiological parameters required to investigate COVID-19 burden". In: *Epidemics* 37, p. 100530.

Zhang, Juanjuan et al. (2020). "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China". In: *Science* 368.6498, pp. 1481–1486.

Zhou, Fei et al. (2020). "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study". In: *The lancet* 395.10229, pp. 1054–1062.