Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Country differences in transmissibility, age distribution and case-fatality of SARS-CoV-2: a global ecological analysis

Caroline Favas a,*, Prudence Jarrett a,*, Ruwan Ratnayake a, Oliver J Watson b, Francesco Checchi b

a Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel St, Bloomsbury, London, WC1E 7HT, United Kingdom
b Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

A R T I C L E   I N F O
Article history:
Received 22 March 2021
Revised 27 September 2021
Accepted 1 November 2021

Keywords:
Coronavirus
COVID-19
transmissibility
age structure
demic infections
parasites

A B S T R A C T
Objectives The first COVID-19 pandemic waves in many low-income countries appeared milder than initially forecasted. We conducted a country-level ecological study to describe patterns in key SARS-CoV-2 outcomes by country and region and explore associations with potential explanatory factors, including population age structure and prior exposure to endemic parasitic infections.

Methods We collected publicly available data and compared them using standardisation techniques. We then explored the association between exposures and outcomes using random forest and linear regression. We adjusted for potential confounders and plausible effect modifications.

Results While mean time-varying reproduction number was highest in the European and Americas regions, median age of death was lower in the Africa region, with a broadly similar case-fatality ratio. Population age was strongly associated with mean (β = 0.01, 95% CI, 0.005, 0.011) and median age of cases (β = –0.40, 95% CI, –0.53, –0.26) and deaths (β = 0.40, 95% CI, 0.17, 0.62).

Conclusions Population age seems an important country-level factor explaining both transmissibility and age distribution of observed cases and deaths. Endemic infections seem unlikely, from this analysis, to be key drivers of the variation in observed epidemic trends. Our study was limited by the availability of outcome data and its causally uncertain ecological design.

© 2021 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Introduction
Since the end of 2019, the COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has spread rapidly worldwide, resulting in considerable morbidity and mortality (Max et al., 2020). Nevertheless, it has affected countries differently, with marked geographical disparities in the observed burden of cases and deaths. While the North American continent bears the highest burden of cases and fatalities to date, African countries seem relatively spared; by September 27, 2021, they made up 2.6% of cases globally and 3.1% of the death toll despite accounting for 14% of the global population (WHO, 2020). Indeed, the first pandemic waves in many low-income countries (LICs) appeared milder than initially forecasted (Koum Besson et al., 2020; Makoni, 2020; Pearson et al., 2020; Truelove et al., 2020; Walker et al., 2020).

Many hypotheses have been put forward to explain these differences. At the health system and public health level, weaker health systems with inequities in access and limited testing capacity may have under-ascertained cases and deaths (Watson et al., 2020). The forewarning from health systems that were quickly overwhelmed in China and Europe may have led to earlier introduction and increased stringency in SARS-CoV-2 control measures in some LICs, and thus partial suppression of community transmission (Massinga Loembé et al., 2020; Mbou et al., 2020).

In terms of population structures, in LICs a younger population age structure could have had implications for transmission as well as lowering the infection fatality ratio due to a smaller proportion of older individuals who are most vulnerable to severe disease (Davies et al., 2020; Ludvigsson, 2020; Nguimkeu and Tadadjeu, 2020). Furthermore, younger populations have a lower prevalence of comorbidities that increase the risk of death from COVID-

* Corresponding author at: Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel St, Bloomsbury, London, WC1E 7HT, United Kingdom.
E-mail address: caroline.favas@lshtm.ac.uk (C. Favas).

https://doi.org/10.1016/j.ijid.2021.11.004
1201-9712/© 2021 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Figure 1. Proposed causal framework of factors determining SARS-CoV-2 transmissibility and COVID-19 disease outcomes

Pink boxes = outcome variables; blue boxes = exposures of interest; green boxes = covariates for which we obtained data; grey boxes = covariates and intermediate outcome variables for which we did not obtain data. Dotted lines represent hypotheses explored in this study.

19 (Clark et al., 2020). In addition, population density and household size are typical drivers of person-to-person disease transmission (Campbell et al., 2014; Weiss and McMichael, 2004).

It has also been postulated that greater lifetime exposure since childhood to common infections in LIC populations may confer some immune protection from SARS-CoV-2 through a more diverse and competitive microbiome, more effective non-specific immune response and decreased likelihood of the cytokine storm seen in severe disease (Kumar and Chander, 2020; Mbou et al., 2020). Parasites such as Plasmodium spp. and soil-transmitted helminths have immunomodulatory effects (Hays et al., 2020; Ssebambulidde et al., 2020). Access to improved water and sanitation may be distally associated with exposure to parasites.

There is little evidence for the relative influence of these hypothesised factors on the observed heterogeneity in global epidemic trends. In Figure 1, we propose a causal framework for understanding the relationship between these potential explanatory factors and the outcomes of transmissibility, the age distribution of cases and deaths, and case-fatality at the population level.

We conducted a country-level ecological study to describe patterns in key SARS-CoV-2 outcomes by country and region and explore associations of these outcomes with potential explanatory factors.

**Methods**

**Study population and period**

We considered data available from March to October 2020 from 193 United Nations member states plus the State of Palestine, Holy See and Hong Kong Special Administrative Region. Dependent territories and other entities were excluded due to inconsistencies in reporting.

**Independent variables**

We sought publicly available data on indicators representing the domains in the causal framework (e.g., the Human Development Index was used to represent the development level; domains for which we could not identify suitable indicators are coloured in grey in Figure 1). Information on the indicators used and data sources are summarised in Table 1 and Supplementary File 1.

**Outcome variables**

**Transmissibility**

We sourced time-varying reproduction numbers by country, as estimated on a real-time basis (Imperial College 2020). These estimates are informed by the dynamics of observed COVID-19 deaths rather than cases, which are less likely to be detected and more susceptible to fluctuations in ascertainment over time due to changes in testing regimens. We averaged estimates over our analysis period (March–October), commencing from the day when 50 cumulative deaths were reported to ensure that averages were not overly influenced by the prior distribution before use to inform the Bayesian framework for estimation (itself highly dependent on observations during the first days and weeks of observed transmission).

**Age of observed cases and deaths**

We conducted a systematic search of national COVID-19 websites (e.g., Ministry of Health dashboards) or regional surveillance reports for overall and age-stratified COVID-19 case and death data. Age-specific data on cases were available from 61 countries and age-specific data on deaths from 39 countries; 35 countries reported both values. For each country, we present a “standardised median age” indicator interpretable as the median age of cases or deaths if the country’s observed age-specific cumulative incidence or death rates were applied to the world’s population age structure (World Population Prospects 2019; United Nations 2019). Further detail can be found in Supplementary File 1.

**Case-fatality ratio**

After omitting countries with <50 total observed cases, we computed a crude case-fatality ratio (CFR) for each country by dividing observed deaths by cases. For countries with available age-specific data, we computed: (i) an age-standardised CFR, derived as above by applying countries’ age-specific crude CFRs to the world’s population structure; and (ii) an incidence-standardised CFR, derived by applying each country’s age-specific CFRs to the observed
Table 1
Summary of included variables, indicators and sources of data.

| Variable measured                          | Indicator                                                                 | Year | Countries included | Data source                                                                 |
|--------------------------------------------|---------------------------------------------------------------------------|------|--------------------|-----------------------------------------------------------------------------|
| **Outcome variables**                      |                                                                           |      |                    |                                                                             |
| Transmissibility of SARS-CoV-2             | Average reproduction number estimates over the study time period, from the day when 50 cumulative deaths were reported | 2020 | 153                | Imperial College COVID-19 LMIC Reports (Imperial College, 2020)             |
| Clinical profile of cases                  | Standardised median age of cases                                          | 2020 | 61                 | NA                                                                          |
| Clinical profile of deaths                 | Standardised median age of deaths                                         | 2020 | 39                 | NA                                                                          |
| Severity of COVID-19 epidemic              | Observed case fatality ratio (CFR):                                        | 2020 | 169                | NA                                                                          |
|                                           | - Crude CFR                                                               |      |                    |                                                                             |
|                                           | - Age-standardised CFR                                                    |      |                    |                                                                             |
|                                           | - Incidence standardised CFR                                              |      |                    |                                                                             |
| **Independent variables**                  |                                                                           |      |                    |                                                                             |
| Prior exposure to infections: malaria      | The age-standardised mean predicted parasite prevalence rate for Plasmodium falciparum malaria for children 2-10 years old | 2017 | 176                | The Malaria Atlas Project database (Weiss et al., 2019)                     |
| Prior exposure to infections: malaria      | The age-standardised mean predicted all-age parasite prevalence rate for Plasmodium vivax malaria | 2017 | 163                | The Malaria Atlas Project database (Battles et al., 2019)                   |
| Prior exposure to infections: other parasites | All-age point prevalence of infection with:                               | 2017 | 186                | Global Burden of Disease Study                                             |
|                                           | - soil-transmitted helminths                                               |      |                    | (Global Burden of Disease Collaborative Network, 2018)                     |
|                                           | - schistosomiasis                                                         |      |                    |                                                                             |
|                                           | - lymphatic filariasis                                                    |      |                    |                                                                             |
| Country age structure                      | Median age (in years) of the population                                   | 2020 | 185                | (United Nations 2019; World Population Prospects 2019)                     |
| Country level of development               | Human development index                                                   | 2018 | 188                | United Nations Development Programme database (United Nations Development Programme, 2019) |
| Population density                         | Population density, as the number of persons per square kilometre        | 2020 | 196                | United Nations World Population Prospects (World Population Prospects 2019, 2019) |

| Variable measured                          | Indicator                                                                 | Year | Countries included | Data source                                                                 |
|--------------------------------------------|---------------------------------------------------------------------------|------|--------------------|-----------------------------------------------------------------------------|
| **Independent variables**                  |                                                                           |      |                    |                                                                             |
| Household size                             | The average number of usual residents (household members) per household   |      | 149                | United Nations Database on Household Size and Composition (United Nations 2019) |
| Access to WASH infrastructures             | The proportion of people using safely managed sanitation services, as a percentage of population | 2017 | 88                 | World Bank World Development Indicator database (The World Bank, 2017)     |
| Stringency of COVID-19 control measures    | Average score for stringency index from 01/01/2020 to 09/09/2020          | 2020 | 169                | Oxford COVID-19 Government Response Tracker (Thomas et al., 2020)           |
| Performance of COVID-19 testing            | Average score for testing policy indicator from 01/01/2020 to 09/09/2020  | 2020 | 169                | Oxford COVID-19 Government Response Tracker (Thomas et al., 2020)           |
| Performance of COVID-19 testing            | Testing rate over the study time period                                   | 2020 | 120                | NA                                                                          |
| Adherence to COVID-19 control measures     | The percentage net change in mobility across four categories:             |      | 130                | Google Community Mobility Reports (Google 2021)                             |
| (change in mobility)                       | 1- Retail & Recreation, 2- Grocery & Pharmacy, 3- Transit stations, 4- Workplaces |      |                    |                                                                             |
| Prevalence of comorbidities                | Age-standardised percentage of country populations at increased risk of severe COVID-19, defined as those with at least one underlying condition listed as “at increased risk” in guidelines from WHO and public health agencies in the United Kingdom and United States | 2020 | 183                | Clark et al. (Clark et al., 2020)                                           |

* Data not from a single source. Description of how these data were obtained is found in the Supplementary File 1. Abbreviations: SARS-CoV-2 = severe acute respiratory syndrome coronavirus 2; WASH = water, sanitation and hygiene; NA = not applicable; UN = United Nations

age-specific caseload in South Korea, selected as a reference due to this country’s reportedly high coverage of case detection (i.e., relatively low selection bias affecting the profile of observed cases) and standard of care (Report on the epidemiological features of coronavirus disease 2019 covid-19 outbreak in the republic of Korea from January 19 to March 2, 2020, 2020). The chosen standardisation method aims to (i) account for age differences in infection-fatality ratios while (ii) reducing bias due to incomplete testing; neither, however, accounts for the effect of age structure on incidence or entirely removes confounding.

**Statistical analyses**

We present two approaches for exploring the associations of hypothesised exposures with each of the above outcomes, while adjusting for potential confounders and accounting for plausible effect modifications chosen a priori for each outcome. For mean time-varying reproduction number (mean Rt) and crude CFR, we carried out a global analysis and an Africa-specific analysis (data were too sparse for other outcomes to perform region-specific analyses); All analyses were conducted on R software (R Foundation for Statistical Computing 2020).
Random forest regression

Random forest (RF) regression is a machine-learning approach that may be used to efficiently explore the importance of predictor variables, and possible effect modifications, for a given outcome. RF imposes minimal statistical assumptions on data and copes well with collinearity (Breiman, 2001). It consists of generating a large number of regression trees (i.e., partitions of the independent variables, in varying order, with each variable generating a node or ‘split’) and averaging over these based on their accuracy in predicting the outcomes. We implemented two RF approaches for each outcome, using the randomForest R package, with 1000 trees grown: (i) using non-missing independent variable data only; (ii) imputing missing independent variable data through the rfitmpute proximity method (Breiman, 2003) (only variables with at least 60% completeness were subjected to imputation; remaining variables were excluded altogether). As the two approaches yielded similar results, for brevity, we only present the latter. We then computed various metrics of variable importance, among which we present, and consider most informative: the mean minimal depth (MMD: a low value indicates that the variable is generally close to the root of the grown trees, i.e., a large proportion of the data are meaningfully partitioned on the basis of this variable); the mean squared error (MSE) increase (i.e., by how much model error increases if the variable is omitted); and the number of trees (out of 1000) in which the variable is the first node based on which the data are split (the higher, the more fundamental the variable may be).

Linear regression

As each outcome was continuous and not structured hierarchically, we applied ordinary least-squares fixed-effects linear models (LM) to explore associations guided by our a priori causal framework. We imputed missing data for the independent variables using the mice package (van Buuren and Groothuis-Oudshoorn, 2011); as with the RF models, only variables with at least 60% completeness were subjected to imputation; variables were otherwise excluded. For each outcome, we first observed collinearity among independent variables through scatterplots and Pearson correlation coefficients (see Supplementary File 2). We screened potential confounding variables through univariate analysis (retaining variables with a P-value <0.20). We fitted models through stepwise backward variable selection, retaining variables that improved goodness of fit (adjusted and F-statistic P-value testing whether the model fits data better than the null model) or influenced the effect of potential exposures on the outcomes. We tried alternative collinear variables and tested plausible two-way interactions. We verified model assumptions, including normality and the homoscedasticity of residuals. For each outcome, we present two models: one with all exposures retained (Supplementary File 3); and a “reduced” model with only significant (P <0.05) and/or model-influential exposures retained (Table 2).

Results

Observed country patterns

Figure 2 summarises trends in each of the outcomes by World Health Organisation (WHO) region, as available. Mean reproduction number was highest in the European regional office (EURO) and Pan American health organisation (PAHO) regions (range 0.92 to 1.77 and 0.73 to 1.73, respectively) and lowest in the African regional office (AFRO) region (0.96 to 1.45). Even when standard-ised for differences in age structure, the median age of observed cases was higher in the AFRO and Eastern Mediterranean regional office (EMRO) regions. In contrast, ascertained deaths occurred at younger ages in those regions compared with EURO and Western Pacific regional office (WPRO) regions. While the crude CFR did not vary widely across regions, higher CFR was found in the EMRO, AFRO, and PAHO regions when CFR was standardised for age and incidence. Figure 3 compares the age-standardised median age of cases and deaths for each country. Countries in the EURO and PAHO regions are clustered in the upper left quadrant (i.e., median age of cases <40, median age of deaths >70). Most countries in the AFRO region are clustered in the lower right quadrants (i.e., median age of cases >40, median age of death <70). Supplementary Files 1 and 2 provide data completeness for each outcome, results by country and graphical explorations of the correlation between independent variables and outcomes.

Statistical associations

Table 2 summarises key results from the two multivariate regression models (RF and the LM reduced version) of imputed predictors for mean, age-standardised median age of observed cases and deaths, and age-standardised CFR. Models for crude and incidence-standardised CFR were excluded as their fit was poor. Supplementary File 3 presents detailed results (all exposures fitted) for each outcome globally and transmissibility and crude CFR for the AFRO region.

In the RF model for mean, household size, prevalence of filariasis and median population age were the three most important variables when considering the different metrics of importance. Mean mobility change, population density and prevalence of Plasmodium falciparum were also important. Testing rate and testing policy were effect modifiers for the association between median population age and mean. In the LM model, filariasis prevalence, median population age, mean stringency of COVID-19 control measures and population density showed significant associations (P <0.05). The association with P. falciparum prevalence was non-significant in the reduced model. When considering only the AFRO region, population age, population density and mean stringency did not remain important, but the importance of prevalence of filariasis increased in the RF model and remained significantly associated in the reduced LM model (P <0.001).

In the RF models for median age of observed cases and deaths, median population age was the most important variable along with proportion at increased COVID-19 risk, testing rate and prevalence of helminths. Mobility change was also important for median age of observed cases, whereas stringency index was important for median age of observed deaths. Mean mobility change, mean stringency index and proportion at increased risk were effect modifiers. In the LM model, median population age was positively associated with median age of cases (P <0.0001) and negatively associated with median age of deaths (P <0.01). The prevalence of helminths was not significantly associated with either outcome.

Lastly, RF suggested that median population age was also an important predictor of age-standardised CFR, but this was not borne out in the LM. Both models had a poor fit.

Discussion

We aimed to identify factors at national level which may explain the global heterogeneity of SARS-CoV-2 epidemics. We found that median population age may explain variability in transmissibility and the age of observed cases and deaths, with a significant association remaining even after age-standardisation. Potential associations between endemic infections and COVID-19 appear unlikely, based on this analysis, to be key drivers in the variation in observed COVID-19 trends. However, the association with filariasis prevalence at global and AFRO levels is intriguing. The observed age distribution amongst reported cases and deaths (after
Table 2
Summary of key associations between independent variables and the outcomes. Exposures of interest are in italics.

| Most important variables | Random forest regression | Multivariate linear regression (reduced model) |
|-------------------------|--------------------------|-----------------------------------------------|
|                         | MMD          | MSE increase  | Times a root | Coef.   | 95% CI | p-value |
| **Outcome: mean time-varying reproductive number (RT)** | | | | | | |
| Median population age (years) | 2.824 | 0.0056 | 62 | 0.0080 | 0.0049 to 0.0112 | <0.0001 |
| Prevalence of lymphatic filariasis (%) | 2.556 | 0.0064 | 131 | -1.9256 | -3.1422 to -0.6915 | 0.0024 |
| Prevalence of P. falciparum (%) | 3.122 | 0.0229 | 66 | - | - | - |
| Mean household size (persons) | 2.202 | 0.0093 | 118 | - | - | - |
| Mean mobility change (%) | 2.358 | 0.0025 | 32 | - | - | - |
| Population density (persons per square kilometre) | 2.996 | 0.0115 | 2 | -0.000036 | -0.000060 to 0.0041 | -0.00012 |
| Mean stringency index (score) | 3.472 | 0.0013 | 0 | 0.0023 | 0.0003 to 0.0043 | 0.0228 |
| **Main effect modifications** | **MMD** | **Occurrences** | Coef. % | 95% CI | p-value |
| Median population age x testing policy | 2.195 | 323 | - | - | - | - |
| Median population age x testing rate | 3.130 | 226 | - | - | - | - |
| (Adjusted)\(^a\) R-squared (F-test; p-value) | 0.38 | - | 0.31 (F = 16.88; p < 0.0001) | |
| **Outcome: age-standardised median age of observed cases** | | | | | | |
| Median population age (years) | 2.144 | 6.2533 | 106 | -0.3961 | -0.5276 to -0.2645 | <0.0001 |
| Prevalence of STH (%) | 3.179 | 2.2431 | 80 | - | - | - |
| Mean mobility change (%) | 2.537 | 1.4986 | 28 | - | - | - |
| -29 to -20 | - | - | - | 1.6983 | -0.9826 to 4.3793 | 0.2097 |
| -19 to -10 | - | - | - | 0.7165 | -1.9888 to 3.4218 | 0.5978 |
| -10 or less | - | - | - | - | - | - |
| Mean testing rate per population (per 1,000) | 2.434 | 2.0958 | 76 | - | - | - |
| Population at increased risk (%) | 2.169 | 4.0191 | 65 | - | - | - |
| **Main effect modifications** | **MMD** | **Occurrences** | Coef. % | 95% CI | p-value |
| Median population age x mean mobility change | 2.210 | 206 | - | - | - | - |
| Median population age x mean stringency index | 1.980 | 201 | - | - | - | - |
| Median population age x proportion at increased risk | 1.376 | 251 | - | - | - | - |
| (Adjusted)\(^a\) R-squared (F-test; p-value) | 0.25 | - | 0.44 (F = 12.58; p < 0.0001) | - |
| **Outcome: age-standardised median age of observed deaths** | | | | | | |
| Median population age (years) | 1.768 | 9.8945 | 119 | 0.3974 | 0.1702 to 0.6245 | 0.0011 |
| Prevalence of STH (%) | 3.208 | 5.5241 | 87 | - | - | - |
| Mean stringency index (score) | 1.848 | 8.9399 | 92 | -0.2044 | -0.3579 to -0.0509 | 0.0105 |
| Mean testing rate per population (per 1,000) | 1.946 | 6.0048 | 71 | - | - | - |
| Population at increased risk (%) | 1.727 | 6.5484 | 68 | -0.4517 | -0.8482 to -0.0553 | 0.0267 |
| **Main effect modifications** | **MMD** | **Occurrences** | Coef. % | 95% CI | p-value |
| Median population age x mean mobility change | 2.005 | 105 | - | - | - | - |
| Median population age x mean stringency index | 1.496 | 171 | - | - | - | - |
| Median population age x proportion at increased risk | 1.923 | 162 | - | - | - | - |
| (Adjusted)\(^a\) R-squared (F-test; p-value) | 0.63 | - | 0.65 (F = 24.53; p < 0.0001) | - |
| **Outcome: age-standardised CFR** | | | | | | |
| Median population age (years) | 1.851 | 0.0000 | 105 | - | - | - |
| Population at increased risk (%) | 2.403 | 0.0000 | 68 | 0.0814 | 0.0159 to 0.1470 | 0.0167 |
| **Main effect modifications** | **MMD** | **Occurrences** | Coef. % | 95% CI | p-value |
| Median population age x mean stringency index | 0.930 | 128 | - | - | - | - |
| Median population age x proportion at increased risk | - | - | - | -0.0019 | -0.0041 to 0.0002 | 0.0787 |
| (Adjusted)\(^a\) R-squared (F-test; p-value) | -0.21 | - | 0.19 (F = 4.63 p < 0.05) | - |

\(^a\) for LM only; Abbreviations: CFR—case fatality ratio; Coef.—coefficient; MMD—mean minimal depth; P. falciparum—Plasmodium falciparum; STH—soil-transmitted helminths; yo—years old

age-standardisation) suggests key differences in surveillance and testing capacity between countries and regions, affecting the representativeness of reported cases and deaths.

**Population age structure**

While we emphasise caution over causal inference due to hidden confounding and incomplete data, we find that population age structure presents a consistent association suggesting that its full impact on country-specific epidemics warrants further research. Similar to what has been observed with severe acute respiratory syndrome (SARS) and Middle East Respiratory Syndrome coronaviruses (Zimmermann and Curtis, 2020), most studies of COVID-19 suggest that children are less susceptible to infection and less infectious (Madewell et al., 2020; Maltezou et al., 2021; Viner et al., 2020). Our findings show that this may also play out at the population level such that countries with a younger population age structure have a smaller susceptible population, less transmission and milder epidemics, reflecting observed epidemic trends in Sub-Saharan Africa. However, epidemiological studies on the role of children have often relied on passive case detection and thus are likely to miss the majority of pauci- or asymptomatic cases in these age groups (Flasche and Edmunds, 2020). Evidence regarding transmission from asymptomatic individuals is contradictory. Some studies suggest that asymptomatic individuals account for a significant share of all transmission (Johansson et al., 2021; Ravindra et al., 2020), whereas others found that the secondary household attack rate from asymptomatic index cases was less than 1% and that COVID-19 spread is mainly driven by symptomatic individuals (Cao et al., 2020; Madewell et al., 2020). In addition, outbreaks among children and adolescents have been important in introducing transmission into households in the United Kingdom (Children’s Task and Finish Group, 2020). Although refuted by some (ECDC, 2020; Ludvigsson, 2020), the role of secondary school-aged children (age 11–18 years) is considered an important driver of transmission (Flasche and Edmunds, 2020). Het-
Figure 2. Analysis outcomes, by World Health Organization region. All boxplots indicate the median and inter-quartile range (boxes), 95% percentile intervals (whiskers) and outliers (dots). CFR = case-fatality ratio.
AFRO= African regional office; EMRO= Eastern Mediterranean regional office; EURO= European regional office; PAHO= Pan American health organisation; SEARO= South-East Asia regional office; WPRO= Western Pacific regional office

Heterogeneity in social contact patterns across age and locations may also influence the role population age structure plays in the transmission of SARS-CoV-2 (Mossong et al., 2008; Prem et al., 2017). In our AFRO-specific analysis, population age structure did not remain predictive of transmission, and testing variables were less important, likely reflecting increased homogeneity in age structure and lower testing capacity across African countries, reducing our ability to detect significant associations.

Older population age structure was associated with a lower median age of cases, after standardising for age and adjusting for confounding, contradicting what is known about age-dependent risk of symptomatic COVID-19. Control measures or behaviour change strategies targeting older people in countries with younger populations may explain this observation. Neither stringency index nor change in mobility are disaggregated by age. Alternatively, lower-income countries might have prioritised older and at-risk people for testing, or reserved testing for travellers, due to lower testing
capacity. Similar patterns were observed in higher-income countries earlier in the pandemic when testing was not widely used and focussed on diagnosing severe infections and infections in key workers to assist with quarantine efforts (United Kingdom Department of Health and Social Care 2021). Testing rate and policy appear to modify the effect of age, supporting this explanation. However, age structure retained its importance after including these effect modifications.

Conversely, countries with older populations had a higher median age of observed deaths from COVID-19, even after age-standardisation. As these same countries have younger cases overall, these findings may reflect better clinical care in countries with older populations so that younger people are more likely to survive severe disease. Moreover, outbreaks in long-term care settings have accounted for a large proportion of deaths in high-income countries and disproportionately affected older people (Comas-Herrera et al., 2021). Notably, increasing prevalence of comorbidities was associated with younger age-standardised age of death, which may reflect a comparatively higher prevalence of diabetes, cardiovascular disease and other chronic conditions occurring at a younger age due to life-course risk factors.

Prior exposure to endemic infections

For transmissibility, prevalence of filariasis ranked highly in the RF model and showed a strong negative association in the LM model. Country prevalence of filariasis may actually be a proxy measure for an unknown factor. An alternative, albeit tenuous, explanation relates to the fact that individuals with prior microfilarial infection appear to have a lower proinflammatory response induced by Th1-type cytokines (Sahu et al., 2008). In SARS-CoV-2 infections, the immunological response involving T-cells seems to be skewed towards these Th1 cells, especially in patients with severe disease (Poland et al., 2020). Therefore, prior exposure to filariasis may reduce the probability of individuals infected by SARS-CoV-2 becoming symptomatic, which may lower their infectiousness and thus population-level transmissibility.

Prevalence of P. falciparum was also a variable of moderate importance in predicting transmissibility in the RF model. However, the association was not significant in the reduced LM model. Although results from the RF model do not indicate a direction of association, one hypothesis that emerged from the literature was that exposure to malaria triggers the production of poly-specific antibodies capable of interacting with multiple antigens, which may confer some protection against SARS-CoV-2 infection (Panda et al., 2020). It is plausible that comparison at national level is insufficiently granular to detect any potential effect of endemic infections. Within-country variations in non-specific immunity are not reflected in our analysis, and sub-national data were not available for most countries. Finally, we emphasize that these ecological associations between filariasis and P. falciparum are useful for generating hypotheses for future research but by themselves do not provide a basis for causal inference.

Case-fatality

Our descriptive analysis does not indicate a relatively lower CFR in the AFRO region, contradicting any narrative that the virus is less lethal in this region. The low number of country observations for the age- and incidence-standardised CFR models is a limitation, and findings related to CFR are likely subject to confounding by poor case ascertainment. Generally, the CFR models fit poorly. The impact of a large number of undiagnosed cases on the CFR, and the limitations of its use in an ongoing epidemic, are well-known (Ritchie and Roser, 2021). CFR is a dynamic value that changes according to disease incidence (i.e., high incidence could lead to more severe cases, a larger hospital burden, and reduced capacity for life-saving care). In addition, studies suggest that the provision of early ambulatory treatment of COVID-19 can substantially reduce hospitalization and death, and hence might be an important determinant of survival (McCullough et al., 2020; Procter et al., 2021). We had no country-level information on prehospital and hospital treatments, which could have affected case fatality statistics. While it is logical that population age is an im-

Figure 3. Scatter plot diagram of the age-standardised median age of deaths and cases (in years) for 35 countries for which both could be computed.
portant determining factor, we cannot conclude this on the basis of our results.

Limitations

An important limitation of this study is the incomplete ascertainment of cases and deaths due to surveillance quality, testing capacity, and cause-of-death ascertainment, which is likely to be highly variable among countries. We could not adjust for all confounders. To ensure coherence with the WHO’s compilation of global COVID-19 data, we relied on publicly-available COVID-19 surveillance data available on national websites without favouring other sources. Data quality is a critical limitation that comprises inconsistent testing information, inconsistent reporting of cases and deaths, and missing data over time, making interpretation of data quality difficult without substantial investigation in data collection practices and biases for each country. In addition, due to COVID-19 cases and death statistics being a crucial input for evaluating governmental pandemic responses, this information is politically sensitive and may thus lead to political influences over how COVID-19 reporting occurs.

We attempted to control for confounding at the aggregate level with respect to testing. We controlled for testing performance by adjusting for the population testing rate and testing policy but acknowledge that these measures are themselves subject to bias. Low levels of testing in low-income countries persist and may mask the true epidemic scale and lead to under-ascertainment of deaths (Watson et al., 2020, n.d.). Furthermore, available testing data may overrepresent older and sicker patients. Testing data disaggregated by age, sex, socioeconomic status and geographic location were not widely available, so it is not possible to estimate the extent to which case ascertainment reflects bias. We note that countries with lower human development index (HDI) and economic indicators generally have a higher prevalence of endemic infections. Countries within WHO regions may also be heterogeneous in terms of health indicators (WHO, 2021) and the wider public health and socio-economic context.

Our study is based on a causal framework describing an evolving and incompletely understood pandemic and reflects the scientific understanding at this time. Our conceptual model may not consider all factors that influence SARS-CoV-2 and residual confounding due to these unknown factors may exist.

In general, ecological studies generate hypotheses but do not provide a basis for causal inference. For example, stringency of control measures appears to be associated with a higher mean; however, this may reflect reverse causality whereby countries with high observed transmission would have maintained strict measures for longer.

We averaged values for those variables that change over time, which may obscure the temporal relationships between them and means we cannot draw conclusions about variations in epidemic trends over time. A longitudinal study based on the same variables would be a useful next step.

Lastly, for many independent variables, our data are derived from modelled estimates (e.g., prevalence of soil-transmitted helminths) based on limited national-level data and therefore may not reflect the true measure. Unsystematic error in explanatory data would have biased coefficients towards the null and thus masked potential associations.

Conclusions and further work

Population age structure appears to be an important factor associated with the transmissibility of SARS-CoV-2 and age distribution of COVID-19 cases and deaths at the national level, even after such outcomes are age-standardised. Our findings do not conclusively support an effect of exposure to endemic parasitic infections on either transmissibility or age distribution of cases and deaths. Research at subnational or individual country level should be conducted to investigate these hypotheses further. Where possible, analysis considering the sociodemographic characteristics of those tested will be useful in understanding the general role of lifelong exposures to infection in the observed patterns of disease. Further, studying social contact patterns in a broader range of countries and the role of urbanization could provide useful insights. This work may be important not only for SARS-CoV-2 but could also inform preparedness and response to future pandemic threats.

Declaration of competing interest

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no support from any organisation other than those listed in the “Funding” section for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

We are grateful to colleagues at the Imperial College of London for supporting our work and for providing useful feedback on the draft of this study report, in particular Charlie Whittaker. We also acknowledge the contribution of John Ackers and Helena Helmy, who provided helpful insights on parasitic infections and their immune modulatory effects.

Funding

This work was supported by the UK Foreign, Commonwealth and Development Office and Wellcome Trust [grant number 221303/Z/20/Z, ‘Epidemic Preparedness - Coronavirus research programme’]; the UK Research and Innovation as part of the Global Challenges Research Fund [grant number ES/P010873/1], and the UK Foreign Commonwealth and Development Office. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethical Approval

This study was based on publicly available data only and did not require ethical approval.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jijd.2021.11.004.

References

Battle KE, Lucas TCD, Nguyen M, Howes RE, Nandi AK, Twohig KA, et al. Articles Mapping the global endemcity and clinical burden of Plasmodium vivax, 2000-17: a spatial and temporal modelling study. Lancet 2019;394:332–43. doi:10.1016/S0140-6736(19)31096-7.
Breiman L. Manual for Setting Up, Using, and Understanding Random Forest V4.0. 2003.
Breiman L. Random forests. Mach Learn 2001;45:5–32. doi:10.1023/A: 0109932404324.
van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;45(3):1–67. doi:10.18637/jss.v045.i03.
Campbell SJ, Savage GB, Gray DJ, Atkinson JAM, Soares Magalhães RJ, Nery SV, et al. Water, Sanitation, and Hygiene (WASH): A Critical Component for Sustainable Soil-Transmitted Helminth and Schistosomiasis Control. PloS Negl Trop Dis 2014;8. doi:10.1371/journal.pntd.0002651.
