Observations of the global epidemiology of COVID-19 from the prepandemic period using web-based surveillance: a cross-sectional analysis

Fatimah S Dawood*, Philip Ricks*, Gibril J Njie, Michael Daugherty, William Davis, James A Fuller, Alison Winstead, Margaret McCarron, Lia C Scott, Diana Chen, Amy E Blain, Ron Moolenaar, Chaoyang Li, Adebola Popoola, Cynthia Jones, Puneet Anantharam, Natalie Olson, Barbara J Marston, Sarah D Bennett

Summary

Background Scant data are available about global patterns of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spread and global epidemiology of early confirmed cases of COVID-19 outside mainland China. We describe the global spread of SARS-CoV-2 and characteristics of COVID-19 cases and clusters before the characterisation of COVID-19 as a pandemic.

Methods Cases of COVID-19 reported between Dec 31, 2019, and March 10, 2020 (ie, the prepandemic period), were identified daily from official websites, press releases, press conference transcripts, and social media feeds of national ministries of health or other government agencies. Case characteristics, travel history, and exposures to other cases were abstracted. Countries with at least one case were classified as affected. Early cases were defined as those among the first 100 cases reported from each country. Later cases were defined as those after the first 100 cases. We analysed reported travel to affected countries among the first case reported from each country outside mainland China, demographic and exposure characteristics among cases with age or sex information, and cluster frequencies and sizes by transmission settings.

Findings Among the first case reported from each of 99 affected countries outside of mainland China, 75 (76%) had recent travel to affected countries; 60 (61%) had travelled to China, Italy, or Iran. Among 1200 cases with age or sex information, 874 (73%) were early cases. Among 762 early cases with age information, the median age was 51 years (IQR 35–63); 25 (3%) of 762 early cases occurred in children younger than 18 years. Overall, 21 (2%) of 1200 cases were in health-care workers and none were in pregnant women. 101 clusters were identified, of which the most commonly identified transmission setting was households (76 [75%]; mean 2·6 cases per cluster [range 2–7]), followed by non-health-care occupational settings (14 [14%]; mean 4·3 cases per cluster [2–14]), and community gatherings (11 [11%]; mean 14·2 cases per cluster [4–36]).

Interpretation Cases with travel links to China, Italy, or Iran accounted for almost two-thirds of the first reported COVID-19 cases from affected countries. Among cases with age information available, most were among adults aged 18 years and older. Although there were many clusters of household transmission among early cases, clusters in occupational or community settings tended to be larger, supporting a possible role for physical distancing to slow the progression of SARS-CoV-2 spread.

Funding None.

Introduction

Before 2019, novel coronaviruses had resulted in two major respiratory illness outbreaks during the 21st century: severe acute respiratory syndrome (SARS), which occurred during 2002–04; and Middle East respiratory syndrome (MERS), which began in 2012. As of July 8, 2020, infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has resulted in more than 11 million laboratory-confirmed cases of COVID-19 and 530 000 confirmed deaths.[1]

A cluster of pneumonia cases of unknown cause was reported in the city of Wuhan, China, by health officials on Dec 31, 2019.[2] By Jan 7, 2020, the China National Institute of Viral Disease Control and Prevention had confirmed the

genetic sequence of SARS-CoV-2 and that the virus was associated with the previously reported pneumonia cluster in Wuhan.[3] On Jan 20, 2020, the US Centers for Disease Control and Prevention (CDC) activated its emergency operations centre in response to the emerging public health threat of COVID-19. On Jan 30, 2020, WHO declared the COVID-19 outbreak a Public Health Emergency of International Concern, and 6 weeks later, WHO characterised the COVID-19 epidemic as a pandemic.[4]

Collection of data about cases of COVID-19 from publicly available official reports has been done by other groups for rapid analyses to inform public health response measures, an approach sometimes referred to as infodemiology.[5–8] Using data obtained from web-based

Research in context

Evidence before this study
We searched PubMed between Dec 31, 2019, and April 14, 2020, with the terms "coronavirus" AND "travel" AND "global". Our search identified 49 publications. To date, scant data are available about global patterns of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spread. Five publications described studies that modelled the effect of global travel on SARS-CoV-2 spread, but no studies used global case data to describe travel exposures among early COVID-19 cases in different countries. We also searched PubMed between Dec 31, 2019, and April 14, 2020, with the terms "coronavirus" AND "cluster". Our search identified 86 publications, of which 25 described individual clusters or small numbers of clusters (fewer than ten in each publication) from individual countries, but no studies summarised cluster characteristics from a global perspective.

Added value of this study
Our analysis uses global epidemiological data from reported cases of COVID-19 to characterise the early spread of the disease and characteristics of cases and clusters of cases. Using publicly available data from national ministries of health and other government official websites and press releases, our analysis provides insights about routes of spread, the age and sex distribution of early confirmed cases and deaths, and commonly identified early transmission patterns.

Implications of all the available evidence
Our findings show the possible contribution of travel from a few countries to the global spread of SARS-CoV-2, confirm age and sex characteristics of early cases of COVID-19 reported from mainland China and several other countries to date, and identify transmission settings that played a role in early SARS-CoV-2 spread. Our analysis also highlights the relatively late detection of SARS-CoV-2 in the WHO African region and remaining knowledge gaps, including the paucity of case data from low-income countries.

surveillance done as part of CDC's emergency response, we aimed to describe the global spread of SARS-CoV-2 and characteristics of cases and clusters in affected countries during the prepandemic period (ie, from Dec 31, 2019, to March 10, 2020).

Methods

Web-based surveillance for COVID-19
On Jan 20, 2020, the US CDC began web-based surveillance to identify newly confirmed cases of COVID-19 outside of mainland China, to inform CDC's global and domestic emergency response. This data collection was done to monitor and assess the COVID-19 outbreak to provide situational awareness and priority setting during an ongoing pandemic. As such, this work was deemed by CDC to be non-research public health surveillance, as defined in title 45 of the electronic Code of Federal Regulations, part 46.102 (L)(2).

To identify new cases of COVID-19, a team of epidemiologists screened the following data sources on a daily basis: national ministry of health or other government websites, press releases, press conference transcripts, and social media feeds (ie, Facebook, Twitter, Instagram); situation reports from national ministries of health and CDC offices in other countries; and media reports. Reports of confirmed cases of COVID-19 from non-publicly available (ie, situation reports) or unofficial (ie, media reports) sources were verified with publicly available information from ministries of health or other government websites, and these electronic source documents were downloaded and archived locally. Data from publicly available official sources were then abstracted and entered into a standardised electronic line list, in which each row represents a case and each column represents a variable. Variables abstracted included report date, illness onset date, country, patient's age and sex, whether the case occurred in a health-care worker or a pregnant woman, travel history, exposure to a previously confirmed case of COVID-19, links to known COVID-19 clusters, and exposure setting (family or household, occupational, community gathering, transport, or other). The appendix (pp 2–3) provides additional surveillance details. A comparison of weekly case counts detected through CDC's web-based surveillance platform compared with case counts published in WHO situation reports showed high concordance (unpublished data).

Exposure and cluster definitions
Countries with at least one confirmed case of COVID-19 identified through web-based surveillance were classified as affected. Information from an infected individual about recent travel to a country with confirmed cases of COVID-19, and about any contact with a previously confirmed case of COVID-19, was abstracted from publicly available case descriptions. Cases were categorised as having no epidemiological link to an affected country or to a previously confirmed case if source documents confirmed that there were no such links; otherwise, cases were categorised as having insufficient information to make this assessment.

A cluster was defined as two or more cases with an epidemiological link based on contact history and could entail multiple generations of infection (ie, cases might not all be directly linked to the primary case in the cluster). The primary case in a cluster was defined as the individual with the earliest illness onset date. For every cluster, transmission settings were categorised as family or household, occupational, community gathering, close contact not otherwise specified, or other; categorisation was based on narrative descriptions of

cases (appendix pp 2–3). One cluster could contribute case counts to more than one transmission setting since transmission settings could vary by generation of spread.

Analytical populations and methods

For this analysis, the prepandemic period of COVID-19 was defined as Dec 31, 2019, to March 10, 2020, based on characterisation of COVID-19 as a pandemic by WHO on March 11, 2020. Analytical populations are shown in figure 1. Countries and locations were defined as WHO Member States (n=194) plus the locations Hong Kong, Liechtenstein, Macau, Palestine, and Taiwan. To characterise the global spread of SARS-CoV-2 by region during this period, we calculated the number of countries and locations reporting at least one COVID-19 case by WHO region (ie, African, Americas, Eastern Mediterranean, European, South-East Asian, and Western Pacific) and epidemiological week, with week 1 corresponding to Dec 29, 2019, to Jan 4, 2020. We also analysed the frequency of travel exposures from previously affected countries among the first case reported in each country and location.

To characterise cases of COVID-19 in affected countries, we restricted the analysis to cases with information available about the patient's age or sex, and we used availability of this information as an indicator of overall data completeness. We calculated frequencies of patients' baseline characteristics, exposure histories, and death as an outcome. Frequency of hospitalisation was not analysed because data were not available to differentiate hospitalisation for isolation alone versus clinical care. We also calculated the frequency of cases by the World Bank 2020 income classification category[9] of reporting countries.

To assess changes in the epidemiology of COVID-19 in countries as outbreaks within countries progressed, the analysis was stratified by early cases (defined as up to the first 100 cases reported from each country) and later cases (defined as cases reported after the first 100 from each country). Characteristics of early and later cases were compared using the χ^2 test for categorical variables or t test for continuous variables; a p value of less than 0·05 was considered statistically significant.

To characterise COVID-19 clusters, we calculated the frequency of clusters by transmission setting and the average number of cases associated with each cluster type. Clusters were first identified from among cases with data for age or sex. Then, to identify the total number of cases and transmission settings associated with each cluster, data were analysed from the full line list dataset, regardless of whether age, sex, or both were known. We used SAS version 9.4 (SAS Institute, Cary, NC, USA) for statistical analyses.

Role of the funding source

No external funding was received for this analysis. The corresponding author had full access to all data in the study and had final responsibility for the decision to submit for publication.
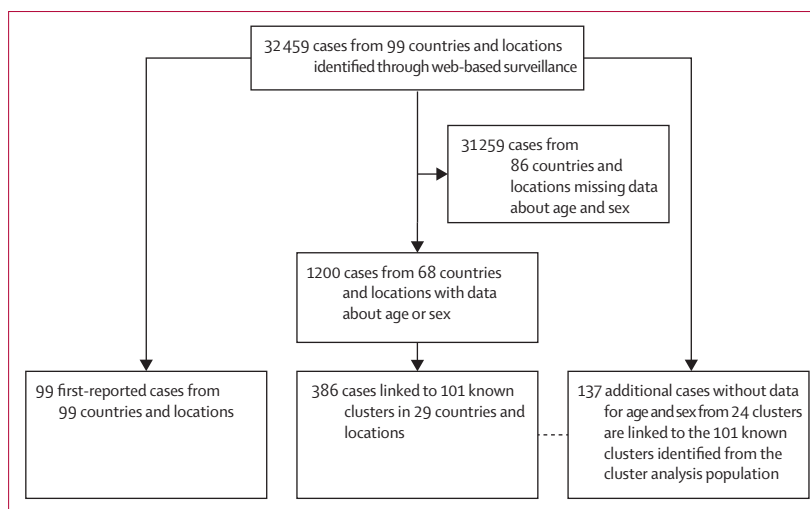

Figure 1: Analytical populations from countries and locations outside of mainland China
Countries and locations defined as WHO Member States plus the locations Hong Kong, Liechtenstein, Macau, Palestine, and Taiwan.

Results

From Dec 31, 2019, to March 10, 2020 (corresponding to epidemiological weeks 1–11 of the COVID-19 outbreak), 100 (50%) of 199 countries and locations (including mainland China) reported cases of COVID-19 (table 1; figure 2). During the first 3 weeks of the COVID-19 outbreak, cases were reported from only two countries outside of mainland China: Japan and Thailand. During weeks 4 and 5, 24 additional countries reported cases, including the first confirmed cases of COVID-19 from countries in the Americas region (first affected country, the USA), the European region (first affected country, Germany), and the Eastern Mediterranean region (first affected country, United Arab Emirates). Subsequently, the number of affected countries plateaued until week 9, when the number of countries in the Eastern Mediterranean region reporting cases of COVID-19 increased from four (17% of countries in the region) to 11 (48%). During weeks 9–11, the proportion of countries and locations with confirmed cases of COVID-19 increased from 32% to 50%, and the first cases were reported from the African region (first affected country, Algeria). By March 10, 2020, just before WHO characterised the COVID-19 outbreak as a pandemic, 45 (83%) of 54 countries and locations in the European region, 16 (70%) of 23 in the Eastern Mediterranean region, and seven (64%) of 11 in the South-East Asian region had reported COVID-19 cases, whereas only 13 (37%) of 35 countries in the Americas region and six (13%) of 46 countries in the African region had reported cases. Although only 13 (43%) of 30 countries in the Western Pacific region had reported cases by week 11, most countries without cases were remote island states with relatively small population sizes.

As of March 10, 2020, 99 countries and locations outside of mainland China had reported cases of COVID-19, of which 75 (76%) identified their first-reported case in an

	Week 1 (Dec 29, 2019, to Jan 4, 2020)	Week 2 (Jan 5 to Jan 11, 2020)	Week 3 (Jan 12 to Jan 18, 2020)	Week 4 (Jan 19 to Jan 25, 2020)	Week 5 (Jan 26 to Feb 1, 2020)	Week 6 (Feb 2 to Feb 8, 2020)	Week 7 (Feb 9 to Feb 15, 2020)	Week 8 (Feb 16 to Feb 22, 2020)	Week 9 (Feb 23 to Feb 29, 2020)	Week 10 (March 1 to March 7, 2020)	Week 11* (March 8 to March 10, 2020)
African (n=46)	0	0	0	0	0	0	0	0	2 (4%)	6 (13%)	6 (13%)
Americas (n=35)	0	0	0	2 (6%)	2 (6%)	3 (9%)	3 (9%)	3 (9%)	5 (14%)	12 (34%)	13 (37%)
Eastern Mediterranean (n=23)	0	0	0	0	1 (4%)	1 (4%)	2 (9%)	4 (17%)	11 (48%)	16 (70%)	16 (70%)
European (n=54)	0	0	0	1 (2%)	8 (15%)	9 (17%)	9 (17%)	10 (19%)	29 (54%)	44 (82%)	45 (83%)
South-East Asian (n=11)	0	0	1 (9%)	2 (18%)	4 (36%)	4 (36%)	4 (36%)	4 (36%)	4 (36%)	7 (64%)	7 (64%)
Western Pacific (n=30)	1 (3%)	1 (3%)	2 (7%)	10 (33%)	12 (40%)	12 (40%)	12 (40%)	12 (40%)	13 (43%)	13 (43%)	13 (43%)
Total (n=199)	1 (1%)	1 (1%)	3 (2%)	15 (8%)	27 (14%)	29 (15%)	30 (15%)	33 (17%)	64 (32%)	98 (49%)	100 (50%)

Data are n (%) of countries with COVID-19 cases. Countries and locations defined as WHO Member States (n=194) and the following locations: Hong Kong, Liechtenstein, Macau, Palestine, and Taiwan. CDC=US Centers for Disease Control and Prevention. *Includes only cases reported as of March 10, 2020, according to CDC web-based surveillance. In addition to countries detected through CDC web-based surveillance, the following five countries reported their first cases on March 10, 2020, according to WHO situation reports: Albania, Bangladesh, Brunei, Cyprus, and Mongolia.

Table 1: Total number of countries and locations with confirmed COVID-19 cases by WHO region and epidemiological week, including mainland China, Dec 29, 2019, to March 10, 2020
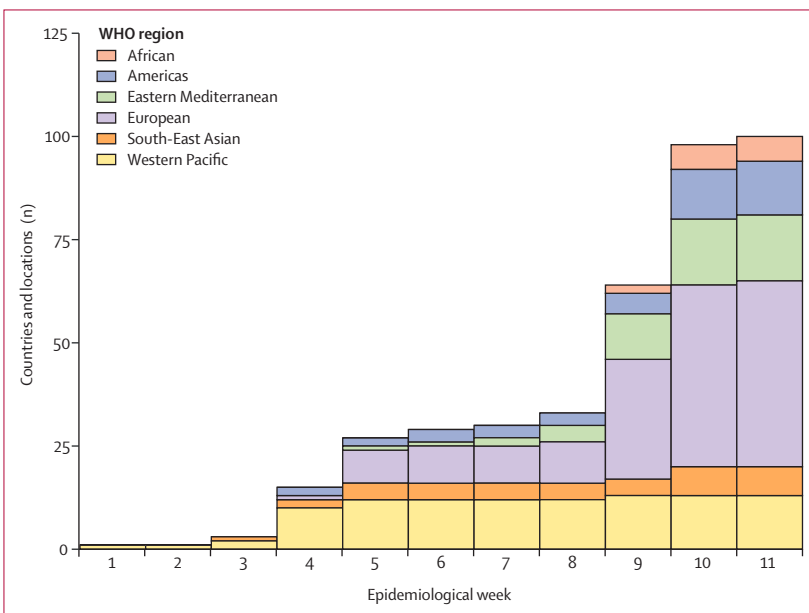

Figure 2: Countries and locations with confirmed cases of COVID-19, by WHO region and epidemiological week, from Dec 31, 2019, to March 10, 2020
Countries and locations defined as WHO Member States plus the locations Hong Kong, Liechtenstein, Macau, Palestine, and Taiwan. Epidemiological weeks are defined in table 1. Data for epidemiological week 11 are up to March 10, 2020.

first-reported cases in affected countries in the Western Pacific region (83%; n=10) and in the South-East Asian region (57%; n=4), cases with travel to Italy accounted for a larger proportion of first-reported cases in the African region (50%; n=3), the European region (36%; n=16), and the Americas region (38%; n=5). Seven (44%) first-reported cases in the Eastern Mediterranean region had a history of travel to Iran, whereas six (37%) had no reported travel history. First-reported cases with reported travel history to countries other than mainland China, Italy, and Iran were India (n=1) from the South-East Asian region; Germany (n=2), Japan (n=1), Latvia (n=1), Spain (n=1), and Switzerland (n=1) from the European region; France (n=2) from the African region; and Ecuador (n=1), Singapore (n=1), and Spain (n=3) from the Americas region. One individual from the African region had a travel history to multiple countries.

From Dec 31, 2019, to March 10, 2020, CDC's web-based surveillance identified 32 459 cases from 99 countries and locations outside of mainland China, of whom 1200 (4%) individuals from 68 countries and locations had data available for sex or age (figure 1). The 68 countries and locations with cases with information on age and sex were Algeria, Andorra, Argentina, Australia, Austria, Azerbaijan, Bahrain, Bhutan, Brazil, Bulgaria, Cambodia, Cameroon, Canada, Chile, Croatia, Czech Republic, Denmark, Dominican Republic, Egypt, Estonia, Finland, France, Georgia, Germany, Greece, Hong Kong, Iceland, India, Indonesia, Iraq, Ireland, Israel, Italy, Japan, Kuwait, Lebanon, Lithuania, Macau, Malaysia, Mexico, Moldova, Morocco, Nepal, Netherlands, New Zealand, Oman, Peru, Philippines, Poland, Portugal, Romania, San Marino, Senegal, Serbia, Singapore, South Africa, South Korea, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Thailand, Togo, Tunisia, Ukraine, United Arab Emirates, and Vietnam. The 31 countries and locations with cases for which no case had information about age or sex were Afghanistan, Armenia, Belarus, Belgium, Bosnia and Herzegovina, Colombia, Costa Rica, Ecuador, Hungary, Iran, Jordan, Latvia, Liechtenstein, Luxembourg, Maldives,

individual with history of travel to an affected country (22 [22%] with travel to China, 11 [11%] to Iran, 27 [27%] to Italy, and 15 [15%] to another country). Of these 75 cases, 34 (45%) had information available confirming that travel took place during the 14 days before their symptom onset. 24 (24%) first-reported cases had no reported travel history in the 14 days before illness onset.

Travel exposures of the first-reported cases varied by WHO region (figure 3). The proportion of cases attributed to travel to mainland China by WHO region ranged from 0% (n=0) in the Eastern Mediterranean and African regions to 83% (n=10) in the Western Pacific region. Although cases attributed to travel to mainland China accounted for most

Malta, Monaco, Nigeria, North Macedonia, Norway, Pakistan, Palestine, Panama, Paraguay, Qatar, Russia, Saudi Arabia, Slovakia, Slovenia, the UK, and the USA. Compared with 31259 cases (from 86 countries and locations) without age and sex information, a higher proportion of the 1200 cases with data for age or sex were reported from countries that reported this information for at least two cases (51% [15816 of 31259] vs 94% [1132 of 1200]; p<0·0001) and from low-income or middle-income countries (1% [172 of 31259] vs 5% [55 of 1200]; p<0·0001). Of the 31259 cases without age or sex information, 25358 (78%) were from three countries with the highest global case counts, with 7904 cases without age or sex information from Iran, 9169 from Italy, and 7285 from South Korea. The proportion of cases with information missing for age and sex increased over time (appendix p 4).

Of 1200 cases with age or sex information, 874 (73%) were early cases and 326 (27%) were later cases. Among the 874 early cases, 690 (79%) were from high-income countries, whereas two (<1%) were from low-income countries (table 2). The median age of 762 early cases with age information was 51 years (IQR 35–63); 25 (3%) of 762 cases with age information were in children younger than 18 years. 460 (56%) of 826 early cases with sex information were in males. 375 (43%) of 874 early cases reported travel from an affected country, 244 (28%) reported contact with a previously confirmed COVID-19 case, and 130 (15%) had no reported epidemiological link to travel to an affected country or a previously confirmed case. Compared with 874 early cases, the 326 later cases were more likely to have no reported epidemiological link to travel or a previously confirmed case or insufficient information about exposure history (255 [29%] vs 282 [87%]; p<0·0001). Most cases in children with exposure history (20 of 23 [87%]), regardless of whether they were early or later cases, were associated with travel from an affected country, contact with a previously confirmed case, or both.

Overall, 52 deaths were identified among the 1200 cases with data available for age or sex. Cases with a reported outcome of death were older (median age 72 years [IQR 62–82]) than were those who did not die (55 years [36–65]; p<0·0001); the age range of cases who died was 35–94 years, and no deaths were reported in children. Dates of symptom onset and death were available for 12 cases; among these, the median time from symptom onset to death was 18 days (IQR 10–23).

Among the 1200 cases with data for age or sex, 21 (2%) were reported among health-care workers. Of these 21 cases, seven (33%) had clear reports of exposure to other cases among health-care workers or patients, six (29%) had other exposures that were not related to health-care settings, including three that had a history of recent travel to an affected country, and eight (38%) had inadequate information about exposures to ascertain exposure setting. Although 184 (15%) of 1200 cases were reported in women aged 18–49 years, none of these women were known to be pregnant.
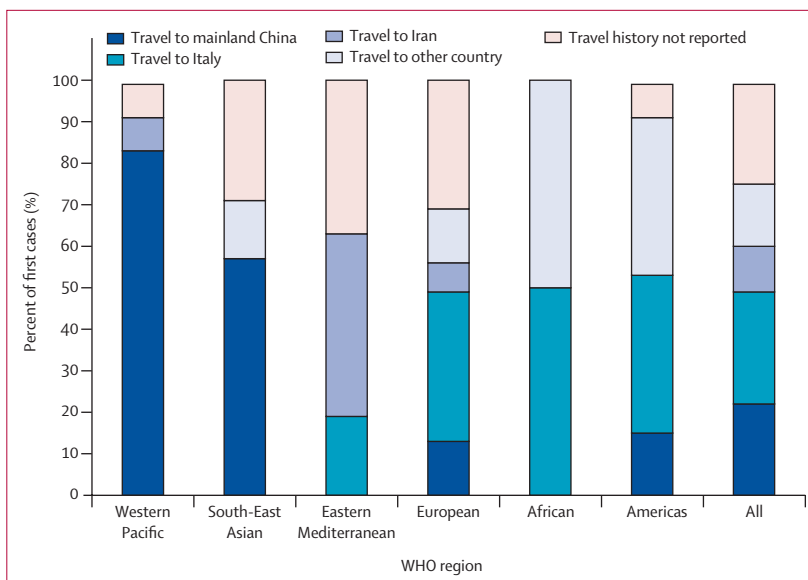

Figure 3: **Travel exposure of first cases reported from 99 countries and locations outside of mainland China, by WHO region**
First case in each country was defined as the case with the earliest report date. Countries and locations defined as WHO Member States plus the locations Hong Kong, Liechtenstein, Macau, Palestine, and Taiwan.

386 (32%) of 1200 cases with data for age or sex were linked to 101 known clusters from 29 countries (Algeria, Australia, Austria, Bulgaria, Canada, Chile, Croatia, Czech Republic, France, Germany, Greece, Hong Kong, India, Israel, Italy, Japan, Lebanon, Macau, Malaysia, Philippines, Romania, Senegal, Singapore, South Korea, Spain, Taiwan, Thailand, United Arab Emirates, and Vietnam; table 3; figure 2). All identified primary cases were in adults aged 18 years and older. 44 (44%) of 101 clusters were described as occurring among close contacts without additional information about transmission setting. The most commonly identified transmission setting among clusters was households (76 [75%]), followed by non-health-care occupational settings (14 [14%]), and community gatherings (11 [11%]). Although household transmission was noted in three-quarters of clusters, fewer cases per cluster were recorded (mean 2·6 [range 2–7]) than in other transmission settings. By contrast, community gatherings accounted for roughly a tenth of clusters but were associated with the highest mean number of cases per cluster (14·2, range 4–36). Clusters associated with community gatherings included transmission among tour groups (n=4), faith-based events (n=4), and dinner parties (n=3). Nine clusters involved transmission associated with transport, including taxis (n=4), flights (n=3), a train (n=1), and a cruise ship (n=1).

Discussion

During the first 11 weeks of the COVID-19 outbreak outside mainland China, cases were detected in half of all countries and locations globally, with an acceleration in case detection during weeks 9–11 of the outbreak. Although most countries and locations in the WHO European, Eastern

	Early cases* in each country (n=874)	Later cases† in each country (n=326)	p value
Country income level‡			
High income	690 (79%)	326 (100%)	<0·0001
Upper middle income	129 (15%)	0	..
Lower middle income	53 (6%)	0	..
Low income	2 (<1%)	0	..
Age, years§	51 (35–63)	55 (45–65)	<0·0001
<18	25/762 (3%)	11/324 (4%)	<0·0001
18–49	339/762 (44%)	99/324 (31%)	..
≥50	398/762 (52%)	214/324 (66%)	..
Sex	..	..	0·21
Male	460/826 (56%)	145/282 (51%)	..
Female	366/826 (44%)	137/282 (49%)	..
Special populations¶	..	..	..
Health-care workers	16 (2%)	5 (2%)	0·73
Pregnant women	0	0	..
Exposure history	..	..	..
Travel from affected countries	375 (43%)	17 (5%)	<0·0001
Contact with previously confirmed case	244 (28%)	27 (8%)	..
No travel history or contact with a previously confirmed case	130 (15%)	59 (18%)	..
Insufficient information	125 (14%)	223 (68%)	..
Deaths‖	48 (5%)	4 (1%)	0·0012

Data are n or n/N (%) or median (IQR). p values are for the comparison between early and late cases. Analyses were restricted to cases with available data about patient age or sex and exclude cases associated with an outbreak among passengers on the *Diamond Princess* cruise ship in Yokohama Bay, Japan. *Defined as cases among the first 100 reported from each country and location. †Defined as cases reported after the first 100 in each country and location. ‡Defined based on World Bank gross national income per capita classification.[9] §Age data were available for 762 of 874 cases among the first 100 cases reported from countries and 324 of 326 cases among later cases reported from countries. ¶Health-care worker status and pregnancy status might have been underascertained because of under-reporting. ‖Deaths might have been underascertained because of incomplete follow-up reporting.

Table 2: Demographic characteristics of early COVID-19 cases compared with later cases in each affected country outside mainland China

	Clusters	Cluster size	Total cases*
All clusters	101 (100%)	5·7 (2–36)	386 (100%)
Transmission setting			
Household	76 (75%)	2·6 (2–7)	266 (69%)
Occupational (non-health-care setting)	14 (14%)	4·3 (2–14)	95 (25%)
Occupational (health-care setting)	6 (6%)	4·2 (2–9)	27 (7%)
Community gathering	11 (11%)	14·2 (4–36)	90 (23%)
Transport	9 (9%)	3·3 (2–10)	35 (9%)
Close contacts†	44 (44%)	3·8 (2–26)	157 (41%)

Data are n (%) or mean (range). A cluster was defined as two or more cases with an epidemiological link. *Total cases from among 1200 cases with information about age or sex that were associated with clusters and transmission settings. An additional 137 cases associated with the 101 clusters shown here were identified from among cases without information about age or sex and are included in cluster size counts. †Case narrative stated that cases in the cluster had close contact but did not provide additional information about transmission setting. Some clusters involving close contact might have occurred in the other transmission settings listed in this table.

Table 3: Transmission setting of COVID-19 clusters

Mediterranean, and South-East Asian regions reported confirmed cases by the time WHO characterised the outbreak as a pandemic, only a third of countries in the Americas and African WHO regions had reported cases, suggesting delayed introduction, delayed detection, or both. Most small island nations in the Western Pacific WHO region also had not reported cases. By the end of epidemiological week 11, more than 50% of countries in the Americas WHO region had reported cases.[10] Almost two-thirds of first cases in affected countries were in people reported to have recent travel from only three affected countries (China, Iran, and Italy), showing how travel from a few countries with substantial SARS-CoV-2 transmission might have seeded additional outbreaks around the world. Consistent with case characteristics reported from mainland China and the USA,[11,12] almost all cases in this analysis occurred among adults (aged ≥18 years), with only 3% occurring among children (aged <18 years). Death related to COVID-19 was associated with older age, although the role of underlying conditions could not be assessed, and data were too sparse to stratify cases and deaths by country to account for variation in population age structure. Although

there were many clusters of household transmission among early cases, clusters in occupational or community settings tended to be larger, supporting a possible role for physical distancing to slow the spread of SARS-CoV-2.

SARS-CoV-2 is the first coronavirus outbreak to be characterised as a pandemic by WHO. Compared with the coronaviruses that caused SARS (SARS-CoV) and MERS (MERS-CoV), SARS-CoV-2 spread more rapidly in the early weeks of the global outbreak. For example, by approximately 11 weeks after SARS-CoV and MERS-CoV were detected outside China and Saudi Arabia, respectively, 27 countries had reported 6500 cases of SARS and three countries had reported nine cases of MERS.[13,14] By comparison, at the comparable timepoint in the SARS-CoV-2 outbreak, 99 countries had collectively reported more than 100 000 cases of COVID-19. The factors contributing to the more rapid spread of SARS-CoV-2 compared with other novel coronaviruses are unclear. Estimates of the basic reproductive number (R_0) of SARS-CoV-2 have varied widely, but a pooled estimate using data from multiple models (R_0 2·9, 95% CI 2·1–4·5)[15] was similar to the estimated R_0 of SARS-CoV (consensus estimate, R_0 of 3)[16] but lower than some estimates for MERS-CoV (estimated range, R_0 of 2–7).[17,18] However, some case series and preliminary transmission modelling studies suggest a larger role for asymptomatic or presymptomatic transmission for SARS-CoV-2 compared with SARS-CoV,[19] and perhaps MERS-CoV.[20] Moreover, although many early affected countries implemented quarantine, contact tracing, and isolation efforts to contain the spread of SARS-CoV-2, variation in these efforts and early case underdetection[6,21] might have facilitated the early spread of SARS-CoV-2. Case detection in the WHO African region occurred late in the prepandemic period, possibly reflecting differences in global case finding, surveillance,

and testing practices. Delayed detection and transmission in the WHO African region was also noted during the 2009 H1N1 influenza pandemic.[22,23]

Among cases with information about age, only 3% were children; household transmission studies from China have reported mixed findings about whether children are as likely as adults to be infected with SARS-CoV-2.[24,25] Multiple factors could have contributed to the early predominance of adults among confirmed cases, including early transmission linked to travel and occupational exposures, and that case finding and testing efforts focused on individuals meeting early case definitions for COVID-19, which included a limited range of symptoms, focused on travel exposures, or both of these. Multiple early reports have suggested that asymptomatic or mild illness might be more common among children.[26,27] Studies that use prospective systematic case identification, testing, and data collection are needed to estimate the prevalence of infection in children, to ascertain how prevalence could evolve with time and with varying community mitigation measures (eg, school closures) and to characterise the full spectrum of SARS-CoV-2 infection and outcomes among children.

Transmission settings of early COVID-19 clusters could provide insights into potential drivers of community transmission and can inform ongoing community mitigation efforts. In our analysis, SARS-CoV-2 infection was transmitted in various settings. Multiple large clusters in our analysis, and large outbreaks reported elsewhere,[28] have been associated with transmission in faith-based settings, highlighting the need to partner with faith-based organisations when designing and implementing community mitigation efforts. Health-care-associated clusters were also identified from early case reports, highlighting the need for early recognition of suspected cases, strict infection prevention and control practices, and health-care worker surveillance for illness detection. Consistent with data from other respiratory disease outbreaks including SARS, MERS, and previous influenza pandemics,[29–31] household transmission of SARS-CoV-2 was common. In the early period of the outbreak, household introduction occurred through adult household members rather than through children, who have been identified as important drivers of transmission in previous influenza pandemics.[32] Household transmission patterns might change as more widespread community transmission occurs and community mitigation strategies (eg, school closures) change over time.

Our analysis was possible because of publicly available data shared by national ministries of health early in the SARS-CoV-2 outbreak. However, several limitations should be considered when interpreting findings. First, the analysis of case characteristics was limited to 4% of global confirmed cases that had publicly available information at least about a case's age or sex. Public reporting of case characteristics was inversely related to case reporting burden, as shown by the fact that more than three-quarters of cases without information about age or sex were from

the three countries with the highest case counts (ie, Iran, Italy, and South Korea). Second, publicly available data varied in completeness, which could have resulted in underascertainment of some case characteristics, and typically did not include information about clinical complications. Thus, frequencies of selected variables (eg, health-care worker, pregnancy status, and death) and clusters should be interpreted as minimum estimates. Further, the proportion of patients who were reported to have died should not be interpreted as a case-fatality ratio, because early reporting of deaths could be biased because of delays from symptom onset to death, increased probability of detection of more severe cases, or both of these.[33] Although the extent of ascertainment bias in our analysis is not known, the age and sex distribution of cases is similar to reports of early confirmed cases from mainland China and elsewhere.[11,12,34,35] Third, the first confirmed case from each country might not have been the first true case of infection in some countries, since early global case detection efforts varied substantially. Finally, almost all cases in our analysis were reported from middle-income and high-income countries from Asia and Europe, driven primarily by late detection in other regions and in part by differences in the level of detail reported for confirmed cases from countries in other regions. The epidemiology of COVID-19 in low-income countries and in Africa could differ, as reported in previous influenza pandemics,[22] and data representative of these settings will be needed to assess the full global effect of the COVID-19 pandemic.

To validate findings from our analysis, we compared case characteristics from our dataset with those reported up to March 9, 2020, from the Open COVID-19 Data Working Group dataset,[36] an aggregation of data from official government sources, peer-reviewed scientific papers, and media websites. Among 3039 cases (9% of 31765 cases) with available data for age or sex through the Open COVID-19 dataset (as of June 18, 2020), the country income level, age, and sex distribution of cases was generally consistent with findings from our analysis, except that the proportion of cases among adults aged 50 years or older was higher in our analysis (appendix p 5). Fewer deaths (n=24) were reported through the Open COVID-19 dataset, and data for special populations or clusters were not available. The proportion of cases with information about age or sex could be higher in the Open COVID-19 Data Working Group dataset compared with our dataset, because the Open COVID-19 dataset is continuously updated and includes data from unofficial sources (eg, media reports).

Much remains unknown about COVID-19, including how its epidemiology will evolve in the context of more widespread community transmission, the effect of community mitigation measures on spread, risk factors for severe disease, how risk factors and severe outcomes might vary between low-resource and high-resource settings, the effect of COVID-19 on vulnerable populations, and the economic and social effects of the virus. Additional studies with detailed epidemiological and clinical data,

and ideally with systematic testing of suspected cases, including among special populations (eg, health-care workers, children, and pregnant women), could further our understanding of COVID-19 and inform preparedness and response measures for the current pandemic.

Editorial note: *The Lancet* Group takes a neutral position with respect to territorial claims in published text, figures, and tables.

References

1 WHO. Coronavirus disease (COVID-19): situation report—170. July 8, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200708-covid-19-sitrep-170.pdf?sfvrsn=bca86036_2 (accessed July 10, 2020).

2 WHO. Emergencies preparedness, response: pneumonia of unknown cause—China. Jan 5, 2020. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/ (accessed June 26, 2020).

3 WHO. Novel coronavirus (2019-nCoV): situation report—1. Jan 21, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf (accessed June 26, 2020).

4 WHO. WHO Director-General's opening remarks at the media briefing on COVID-19. March 11, 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (accessed June 26, 2020).

5 Xu B, Kraemer MUG. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis* 2020; **20:** 534.

6 Tuite AR, Bogoch II, Sherbo R, Watts A, Fisman D, Khan K. Estimation of coronavirus disease 2019 (COVID-19) burden and potential for international dissemination of infection from Iran. *Ann Intern Med* 2020; **172:** 699–701.

7 Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020; **172:** 577–82.

8 Eysenbach G. Infodemiology: the epidemiology of (mis)information. *Am J Med* 2002; **113:** 763–65.

9 The World Bank. World Bank country and lending groups. 2020. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (accessed June 26, 2020).

10 WHO. Coronavirus disease 2019 (COVID-19): situation report—54. March 14, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200314-sitrep-54-covid-19.pdf?sfvrsn=dcd46351_8 (accessed June 26, 2020).

11 Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020; **41:** 145–51 (in Chinese).

12 CDC COVID-19 Response Team. Coronavirus disease 2019 in children—United States, February 12–April 2, 2020. *MMWR Morb Mortal Wkly Rep* 2020; **69:** 422–26.

13 WHO. Update 47: studies of SARS virus survival, situation in China. May 5, 2003. https://www.who.int/csr/sars/archive/2003_05_05/en/ (accessed June 26, 2020).

14 WHO. Emergencies preparedness, response: novel coronavirus infection—update. Nov 30, 2012. https://www.who.int/csr/don/2012_11_30/en/ (accessed June 26, 2020).

15 Park SW, Bolker BM, Champredon D, et al. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. *medRxiv* 2020; published online Feb 28. https://doi.org/10.1101/2020.01.30.20019877 (preprint).

16 Bauch CT, Lloyd-Smith JO, Coffee MP, Galvani AP. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. *Epidemiology* 2005; **16:** 791–801.

17 Majumder MS, Rivers C, Lofgren E, Fisman D. Estimation of MERS-coronavirus reproductive number and case fatality rate for the spring 2014 Saudi Arabia outbreak: insights from publicly available data. *PLoS Curr* 2014; published online Dec 18. https://doi.org/10.1371/currents.outbreaks.98d2f8f3382d84f390736cd5f5fe133c.

18 Hsieh YH. 2015 Middle East respiratory syndrome coronavirus (MERS-CoV) nosocomial outbreak in South Korea: insights from modeling. *PeerJ* 2015; **3:** e1505.

19 Peiris JS, Yuen KY, Osterhaus AD, Stohr K. The severe acute respiratory syndrome. *N Engl J Med* 2003; **349:** 2431–41.

20 Al-Tawfiq JA, Memish ZA. Drivers of MERS-CoV transmission: what do we know? *Expert Rev Respir Med* 2016; **10:** 331–38.

21 Niehus R, De Salazar PM, Taylor A, Lipsitch M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. *medRxiv* 2020; published online Feb 18. https://doi.org/10.1101/2020.02.13.20022707 (preprint)

22 Dawood FS, Iuliano AD, Reed C, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet Infect Dis* 2012; **12:** 687–95.

23 Nzussouo NT, Michalove J, Diop OM, et al. Delayed 2009 pandemic influenza A virus subtype H1N1 circulation in West Africa, May 2009–April 2010. *J Infect Dis* 2012; **206** (suppl 1): S101–07.

24 Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis* 2020; published online April 27. https://doi.org/10.1016/S1473-3099(20)30287-5.

25 Li W, Zhang B, Lu J, et al. The characteristics of household transmission of COVID-19. *Clin Infect Dis* 2020; published online April 17. https://doi.org/10.1093/cid/ciaa450.

26 Dong Y, Mo X, Hu Y, et al. Epidemiology of COVID-19 among children in China. *Pediatrics* 2020; **145:** e20200702.

27 Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta Paediatr* 2020; **109:** 1088–95.

28 Republic of Korea CDC. The updates of COVID-19 in Republic of Korea. March 5, 2020. https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&tag=&act=view&list_no=366456 (accessed June 26, 2020).

29 Wilson-Clark SD, Deeks SL, Gournis E, et al. Household transmission of SARS, 2003. *CMAJ* 2006; **175:** 1219–23.

30 Hui DS, Azhar EI, Kim YJ, Memish ZA, Oh MD, Zumla A. Middle East respiratory syndrome coronavirus: risk factors and determinants of primary, household, and nosocomial transmission. *Lancet Infect Dis* 2018; **18:** e217–27.

31 Lau LL, Nishiura H, Kelly H, Ip DK, Leung GM, Cowling BJ. Household transmission of 2009 pandemic influenza A (H1N1): a systematic review and meta-analysis. *Epidemiology* 2012; **23:** 531–42.

32 Monto AS, Davenport FM, Napier JA, Francis T Jr. Effect of vaccination of a school-age population upon the course of an A2-Hong Kong influenza epidemic. *Bull World Health Organ* 1969; **41:** 537–42.

33 Spychalski P, Blazynska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. *Lancet Infect Dis* 2020; **20:** 774–75.

34 Ng Y, Li Z, Chua YX, Chaw WL, et al. Evaluation of the effectiveness of surveillance and containment measures for the first 100 patients with COVID-19 in Singapore: January 2–February 29, 2020. *MMWR Morb Mortal Wkly Rep* 2020; **69:** 307–11.

35 Zhang J, Litvinova M, Wang W, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect Dis* 2020; **20:** 793–802.

36 Xu B, Gutierrez B, Mekaru S, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. *Sci Data* 2020; **7:** 106.