THE LANCET
Infectious Diseases

Supplementary appendix

The Global Epidemiology of COVID-19: Observations from the Pre-Pandemic Period using Web-based Surveillance, A Cross-Sectional Anaysis

Appendix

Supplemental Methods

Web-based surveillance for COVID-19

Each newly identified case was assigned a case identification code based on the reporting country and the order in which the case was reported from the country. Many Ministry of Health and other government websites and press releases provided sequential updates about confirmed COVID-19 cases as additional information became available. Therefore, efforts were made to review the same publicly available official sources for updates, and data were updated accordingly. In most cases, countries used sequential numbering to identify cases so updates were easily linked to initial case reports; in instances where country case numbering was not provided, patient age and sex were reported and used to link updates with initial case reports.

An absence of reported healthcare worker or pregnancy status was categorized as 'no' for the purposes of this analysis. An absence of information about travel history or exposure to a previously confirmed COVID-19 case was categorized as 'unknown.'

Exposure and cluster definitions

Community gatherings were defined as social gatherings outside the home, such as for a community meal or faith-based event. Occupational transmission was defined as occurring in the workplace and was further stratified as occurring while providing healthcare services versus other types of occupational transmission. When a cluster involved more than one transmission setting, cases were counted towards the specific transmission setting in which they were thought to have transmitted or acquired infection (e.g. cases were counted towards more than one transmission setting if they acquired and transmitted infection in different settings).

Cases reported as confirmed during testing aboard a cruise ship in Yokohama Bay, Japan[1] as part of an outbreak investigation were excluded from the cluster analysis as these cases represented a unique outbreak in a closed setting in which case demographics might differ from other early cases (Cluster IDs: CL-HK-014 and CL-JPN-080). In addition, cases reported from two large outbreaks from South Korea among members of a church community (Cluster ID: CL-KOR-031) and psychiatric ward (Cluster ID: CL-KOR-HXX) were

excluded because individual case level data were not available for all cases and the outbreaks extended beyond

March 11, 2020; each outbreak was reported to have >100 associated cases.[2]

1.   Kakimoto K, Kamiya H, Yamagishi T, Matsui T, Suzuki M, Wakita T. Initial Investigation of Transmission of COVID-19 Among Crew Members During Quarantine of a Cruise Ship - Yokohama, Japan, February 2020. MMWR Morb Mortal Wkly Rep 2020;69:312-3.
2.   The updates on COVID-19 in Korea as of 5 March. (Accessed at
https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&tag=&act=view&list_no=366456.)

Table 1 Proportion of confirmed Coronavirus Disease 2019 cases with missing age and sex information by epidemiologic week, N=32,459 cases

	No. (%) of COVID-19 Cases with Missing Age and Sex Information by Epi Week										
	Week 1 12/29/2020-1/4/2020	Week 2 1/5/2020-1/11/2020	Week 3 1/12/2020-1/18/2020	Week 4 1/19/2020-1/25/2020	Week 5 1/26/2020-2/1/2020	Week 6 2/2/2020-2/8/2020	Week 7 2/9/2020-2/15/2020	Week 8 2/16/2020-2/22/2020	Week 9 2/23/2020-2/29/2020	Week 10 3/1/2020-3/7/2020	Week 11[a] 3/8/2020-3/10/2020
Number of Cases with Missing Age & Sex Information	0	0	0	2	24	31	16	490	5024	16985	8687
% of Cases with Missing Age & Sex Information	0	0	0	6	17	25	14	75	96	98	99
Total Cases	**0**	**0**	**4**	**36**	**141**	**122**	**113**	**651**	**5243**	**17346**	**8803**

[a]Includes only cases reported as of March 10, 2020 according to CDC web-based surveillance.

Table 2 Demographic Characteristics of Early Coronavirus Disease Cases Compared to Later Cases in Each Affected Country outside Mainland China using data from the Open COVID-19 Data Working Group dataset downloaded June 18, 2020*, N=3039 Cases** from 83 Countries and Locations[a].

Characteristics	Early Cases[a] in Each Country, n (%) (n=1467)	Later Cases[a] in Each Country, n (%) (n=1572)	p-value
Country income level			
High income	1087 (74)	1572 (100)	<0.0001
Upper middle income	237 (16)	0	
Lower middle income	138 (9)	0	
Low income	5 (<1)	0	
Age[c]			
<18 years	57/1186 (5)	46/1478 (3)	<0.0001
18-49 years	658/1186 (55)	598/1478 (40)	
≥50 years	471/1186 (40)	834/1478 (56)	
Male	784/1409 (56)	818/1567 (52)	0.214
Deaths[e]	20 (1)	4 (<1)	0.0012

* Data accessed on June 18, 2020 from https://github.com/beoutbreakprepared/nCoV2019/tree/master/latest_data.

**Analysis restricted to cases with information about case age or sex. If case age was provided as an age range, the midpoint of the age range was used if the age range was ≤10 years and age data were considered missing if the age range was >10 years. Case ages from the Philippines were excluded from the analysis because all cases had an age of 0 years entered suggesting a possible error in data entry.

[a] Early cases were defined as those among the first 100 cases reported from each country or location. Later cases were defined as those reported after the first 100 cases in each country or location.

[b] Country income defined based on World Bank gross national income per capita classification available at https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.

[c] Age data were available for 1186/1467 cases among the first 100 cases reported from countries and 1478/1572 cases among later cases reported from countries. Data about case sex were available for 1409/1467 cases among the first 100 cases reported from countries and 1567/1572 cases among later cases reported from countries.