International Journal of Infectious Diseases 96 (2020) 519–523

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Early network properties of the COVID-19 pandemic – The Chinese scenario

Ariel L. Rivas[a], José L. Febles[b], Stephen D. Smith[c], Almira L. Hoogesteijn[b], George P. Tegos[d], Folorunso O. Fasina[e,*], James B. Hittner[f]

[a] Center for Global Health, Department of Internal Medicine, Medical School, University of New Mexico, Albuquerque, USA
[b] Department of Human Ecology, CINVESTAV-IPN, Mérida, Mexico
[c] Institute for Resource Information Science, College of Agriculture, Cornell University, Ithaca, USA
[d] Micromoria LLC, Marlborough, MA, USA
[e] Food and Agriculture Organization, Dar es Salam, Tanzania & Department of Veterinary Tropical Diseases, University of Pretoria, South Africa
[f] Department of Psychology, College of Charleston, Charleston, USA

ARTICLE INFO

Article history:
Received 28 March 2020
Received in revised form 11 May 2020
Accepted 13 May 2020

Keywords:
Network-theory
Smallworld
COVID-19
Interdisciplinary
Geo-referenced

ABSTRACT

Objectives: To control epidemics, sites more affected by mortality should be identified.
Methods: Defining epidemic nodes as areas that included both most fatalities per time unit and connections, such as highways, geo-temporal Chinese data on the COVID-19 epidemic were investigated with linear, logarithmic, power, growth, exponential, and logistic regression models. A *z*-test compared the slopes observed.
Results: Twenty provinces suspected to act as epidemic nodes were empirically investigated. Five provinces displayed synchronicity, long-distance connections, directionality and assortativity – network properties that helped discriminate epidemic nodes. The rank I node included most fatalities and was activated first. Fewer deaths were reported, later, by rank II and III nodes, while the data from rank I–III nodes exhibited slopes, the data from the remaining provinces did not. The power curve was the best fitting model for all slopes. Because all pairs (rank I vs. rank II, rank I vs. rank III, and rank II vs. rank III) of epidemic nodes differed statistically, rank I–III epidemic nodes were geo-temporally and statistically distinguishable.
Conclusions: The geo-temporal progression of epidemics seems to be highly structured. Epidemic network properties can distinguish regions that differ in mortality. This real-time geo-referenced analysis can inform both decision-makers and clinicians.

Introduction

The challenges associated with the COVID-19 pandemic may require novel approaches. Given the numerous asymptomatic infections reported in this disease, actions that focus on symptomatic individuals are prone to fail (Nishiura and Linton, 2020). Some classic concepts – e.g., 'recovered' and 'contact tracing' – may not apply: patients regarded as recovered may be test-positive and people without a travel history may be infected (Lan et al., 2020; CDC, 2020). To avoid these ambiguities, here an unambiguous metric was explored: mortality. To that end, geo-referenced data were investigated using a procedure grounded on Network Theory (Meyers, 2007).

A network may be defined as a set of lines that connects circles (nodes). Accordingly, an epidemic node could be categorized as the smallest surface that (i) includes an explicit connection, (ii) captures most infections per time point, and (iii) reports secondary deaths (Rivas et al., 2012). Infected cities that possess road, railroad, and/or air travel networks fit that definition. Networks possess several properties, including directionality, assortativity, synchronicity, and smallworld (long-distance) connections (Meyers, 2007; Rivas et al., 2012; Watts and Strogatz, 1998). Directionality refers to the temporal sequence and geographical location of outbreaks. Assortativity distinguishes the magnitude of infections reported by epidemic nodes of different influence on epidemic dispersal. Synchronicity reveals epidemic nodes that are activated at the same time and exhibit similar number of infections

* Corresponding author at: FAO ECTAD Tanzania & Department of Veterinary Tropical Diseases, University of Pretoria, South Africa.
E-mail address: Folorunso.fasina@fao.org (F.O. Fasina).

https://doi.org/10.1016/j.ijid.2020.05.049
1201-9712/

per unit of time. Smallworld connections are those that, regardless of the distance between nodes, may induce outbreaks (e.g., air travel connections).

While earlier studies have documented network properties in rapidly disseminating epidemics that affect non-human species, such properties have not yet been explored in human epidemics (Rivas et al., 2012). Hence, this study pursued to: (i) elucidate whether the early COVID -19 epidemic revealed network-like properties; (ii) distinguish epidemic nodes; and (iii) using real-time assessments, confirm or reject the properties and classifications previously mentioned.

Material and methods

Data

Epidemic and georeferenced data were collected from public sources as well as ESRI Data and Maps for ArcGIS (2019) and ArcGIS Living Atlas of the World (ESRI Inc., Redlands, CA, USA) (JHU, 2020; Harvard University, 2020; NBSC, 2020; GHS, 2020). Earlier epidemic data (reported before January 21, 2020) were extracted from a published study (Huang et al., 2020). While the detection of epidemic nodes may require an operation described elsewhere (Rivas et al., 2012), the absence of high-resolution (point-based) georeferenced data prevented the use of that procedure. In light of this limitation, an empirical analysis was conducted on sites suspected to be acting as epidemic nodes, that is, to determine whether they expressed network properties. Epidemic nodes were differentiated by data patterns: those that reported more fatalities were assigned the lowest ranks and those presenting with fewer or no secondary deaths received the highest rank. Therefore, one epidemic rank could include more than one province.

Disease mapping

Maps were produced with ArcGIS Pro 2.5.0 (ESRI, Redland, CA, USA). Analyses and figures were conducted or made with commercial packages (IBM SPSS Statistics 24, IBM Corp, Armonk, NY; and Minitab 18, Minitab LLC, State College, PA, USA).

Statistical analysis

Data patterns distinguished three epidemic nodes: (1) rank I (Hubei only), (2) rank II (the sum of deaths reported in Heilongjiang and Henan provinces), and (3) rank III (the sum of deaths reported in Anhui and Chongqing provinces). To determine whether the slopes of these epidemic nodes differed, six curve-fitting regression analyses investigated linear, logarithmic, power, growth, exponential, and logistic regression models, respectively. Analyses and figures were conducted or made with commercial packages (IBM SPSS Statistics 24, IBM Corp, Armonk, NY; and Minitab 18, Minitab LLC, State College, PA, USA).

Results

Wuhan is the putative origin of the COVID-19 epidemic (Figure 1A). By February 22, 2020, neither population nor distance to Wuhan correlated with the fatalities observed in 19 Chinese provinces (both with $p > 0.05$, Supplementary Table S1 and Supplementary Figure S2). A movie summarizes the geo-dynamics of this epidemic (Supplementary Movie S3).

Not all provinces were epidemic nodes: Jiangxi and Jilin had one fatality each (Figure 1B). Thus, these two provinces did not generate secondary fatalities. In contrast, the slopes of geo-temporal data on mortality differentiated, at least, three epidemic nodes: (i) rank I (Hubei), (ii) rank II (Heilongjiang and Henan), and (iii) rank III (Anhui and Chongqing) nodes (Figure 1C).

Henan and Heilongjiang exhibited *synchronicity*. In spite of major differences (including distance to Hubei, and population size), both provinces became activated at the same time and their number of fatalities was similar, over time (Figure 1D). Heilongjiang also displayed *smallworld* (long-distance) epidemic connectivity: its capital, Harbin, is 2254 km away from Wuhan, Hubei, i.e., a ~4.5 times longer distance than the 514 km that separate Zhengzhou, Henan, from Wuhan, Hubei (Figure 1E).

Heilongjiang and Henan, as well as Anhui and Chongqing, showed a pattern compatible with *assortativity*, i.e., nodes of similar rank were associated with a similar level of mortality (Figure 1C). While it is suspected that assortativity also occurred in Hubei (press reports suggest at least two epidemic nodes developed in their hospitals and in one prison) (China Prison, 2020), the lack of point-based data prevented its identification.

Temporal *directionality* was documented: epidemic nodes of higher influence on epidemic dispersal (lower rank) grew in number of fatalities before higher rank nodes did, e.g., the blue line of data points was growing before the green line, the green line was growing before the red line, and the red line was growing before the orange line (vertical lines, Figure 1F). Therefore, the data provided graphic evidence on the geographical and temporal location of epidemic nodes as well as Network properties, such as synchronicity and directionality.

To elucidate whether the geo-temporal series depicted in Figure 1B and C were similar, a curve-fitting regression algorithm tested the slopes of rank I–III nodes. Data patterns distinguished three epidemic nodes: (1) rank I (Hubei only), (2) rank II (the sum of deaths reported in Heilongjiang and Henan provinces), and (3) rank III (the sum of deaths reported in Anhui and Chongqing provinces). To determine whether the slopes of these epidemic nodes differed, six curve-fitting regression analyses investigated linear, logarithmic, power, growth, exponential, and logistic regression models, respectively.

The time series correlated with the mortality series. The power curve was the best fitting model for all three slopes. Using the unstandardized slope regression coefficients and standard errors of the coefficients for the power curve models, z-tests compared pairs (rank I vs. rank II, rank I vs. rank III, and rank II vs. rank III) of regression coefficients (Paternoster et al., 1998). z-test values ranged between 4.73 and 11.88. All three z-test values exceeded the z-critical value for alpha = 0.001 two-tailed test, which is $z = 3.30$. Therefore, rank I-III nodes were statistically significantly different from one another.

Discussion

Findings supported the view that, once the epidemic structure is consolidated (after a brief phase reveals a linear growth in fatalities), the epidemic progression may display network properties. While anticipated 22 years ago (Watts and Strogatz, 1998), this is the first demonstration of network properties – including smallworld connections – in human epidemics. Supporting the potential epidemic role of long-distance (smallworld), rapidly connecting structures, the cumulative number of fatalities did not correlate with either Euclidian distance or population. In contrast, a network-based, geo-temporal analysis detected and differentiated three epidemic nodes. Because classic epidemiological concepts –including Euclidian distances, geographical assessments of population density, and the ratio of secondary infections generated per primary infections (also known as the basic reproductive number or R_0) – are sensitive to heterogeneous geographical structures, findings provide a real-time, directly measurable alternative to measure and monitor epidemic

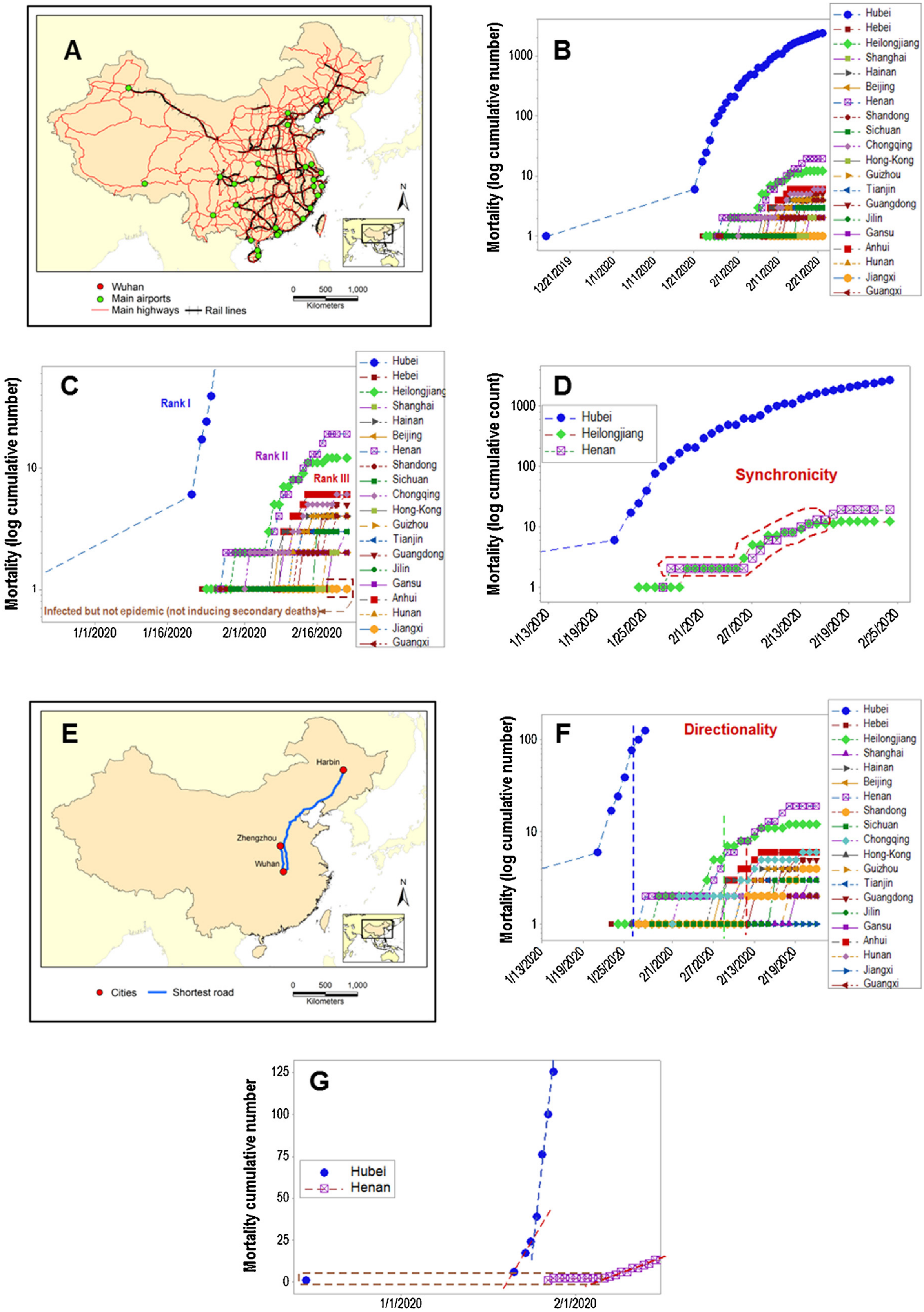

Figure 1. A. Location of the epidemic network. Centered on Wuhan, the province of Hubei shows three (road, railroad and air) networks. In addition, Wuhan has river-mediated connections (not shown). B. Temporal progression of covid-19 fatalities in twenty Chinese areas. C. Differentiation of epidemic nodes. To facilitate visualization, the same plot shown in A is displayed with truncated data. At least four groups of data patterns are observed: (i) the rank I node (composed only by Hubei data), which includes most fatalities at all times, (ii) the rank II node (composed of Heilongjiang and Henan data), which reported the second highest number of deaths; (iii) the rank III node (composed of Anhui and Chongqing data), which reported fewer deaths and they were observed after those of rank I and II nodes; and (iv) the remaining provinces, which did not display a slope and generated only one or no secondary fatalities. D. Synchronicity. A truncated set displays the data of rank I and II nodes. It is shown that Heilongjiang and

progression (Meyers, 2007; Rivas et al., 2012; Li et al., 2011). Epidemic networks are not limited to large countries, such as China: they can also be observed in small regions (Rivas al et al., 2012).

To provide a context to these findings, cultural, demographic and biological perspectives are discussed. The Chinese New Year, in 2020, was celebrated in January 25. It is included within the 40-day long Spring Festival (January 10 to February 18), when most Chinese citizens take their annual vacations. Hence, every year, billions of trips are conducted during this period of time (Tian et al., 2020; Chen et al., 2020). Press agencies have estimated that up to 5 million inhabitants of Wuhan left the city prior to the lockdown imposed on January 23, 2020; of whom about 70% visited other places within the Hubei province (Associated Press, 2020). Therefore, the hypothesis that the road, railroad, as well as air and river connecting networks facilitated viral dispersal is not rejected.

It is suggested that the fact that epidemic networks show highly structured (non-random) and distinct patterns can foster novel research opportunities in basic science. Because specific geographical locations are both fragmented (heterogeneous) and dynamic – at least seasons and human mobility differ over time), they cannot be assumed or hypothesized. Yet, they can be explicitly measured. These precisions matter when responses, to be rapidly deployed and effective, have to be geographically specific. That is so because the Critical Response Time (time available to choose and deploy a policy expected to be successful) may be extremely short in a rapidly disseminating epidemic (Rivas et al., 2003).

Here connectivity – not contact tracing – was emphasized. Two reasons explain this priority: (i) in epidemics with asymptomatic patients (as clearly shown in the 'Diamond Princess' cruise case, where 86% of test-positive individuals were asymptomatic) (CNBC, 2020), control measures that depend on detection of symptomatic patients will likely fail; and (ii) to estimate where the epidemic is going, geo-referenced information on connectivity is needed.

While this study lacked high-resolution georeferenced data and – given the uncertainties associated with asymptomatic cases – focused on mortality, it may apply when massive testing is conducted and geo-referenced point-based data are available. Findings suggest that it is possible to design responses that, instead of bringing patients to hospitals, *bring hospitals to the patients.*

For example, let us assume that a fatality and/or a test-positive individual was reported in a specific factory/neighborhood/school of a medium-size city. Using point-based data, it may be

found that there is a bridge connecting the affected area with the rest of the city – a point that, if disassembled, can prevent epidemic dispersal, provided that, in addition and immediately, policy-makers send a mobile, emergency hospital to that area (including medical personnel) and two isolation perimeters are established. The purpose of the outer perimeter is to create a parking lot to be used by vehicles that come to the isolated area (providing food and medical supplies), which cannot return to the city unless they are disinfected and remain in quarantine in the outer perimeter. Instead of quarantining people, the *outer perimeter would quarantine used vehicles and/or equipment*. As currently demonstrated in South Korea, this proposal is feasible, less disruptive and, potentially, more effective than delayed and generic (non-georeferenced) policies (Lessons from SK, 2020).

Furthermore, real-time statistical testing of data gathered from sites suspected to be epidemic nodes may rapidly confirm or reject that hypothesis, as described here. While forecasts may be erroneous when geographical features are not evaluated (Jewell et al., 2020; Marchant et al., 2020), real-time assessments analyze and describe facts (not assumptions) as they are, as soon as they occur. For instance, analyses that can be conducted within a few minutes can show where and when interventions are likely to be successful (e.g. Hubei province, in December, 2019; or Henan, in early February, 2020, Figure 1G) and confirm or reject the hypothesis that a given site is an influential epidemic node.

The fact that bio-geo-temporal interactions may follow Network properties complements and, probably, may expand current research efforts on infectious diseases. One area of paramount relevance is antimicrobial resistance (AMR). While usually described with emphasis on bacterial pathogens and their ability to reproduce in the presence of antibiotics, the survival of viral pathogens to antiviral drugs is also included in the study of AMR. While geo-referenced data on bacterial and viral strains, as well as data on resistance against antibiotics and antiviral drugs are available, they are necessary but not necessarily sufficient to understand and/or predict how and/or when infectious diseases will spread and where they are more likely to affect specific subpopulations (Okeke and Edelman, 2001; Lauderdale et al., 2004; Schaumburg et al., 2014; Tacconelli et al., 2018). The exploration and use of geographically explicit network properties may add a methodological tool to the study of AMR.

To materialize these possibilities, interdisciplinary teams are required. While sometimes viewed as synonymous, multi- and inter-disciplinarity are quite different: while multidisciplinary teams rarely create new knowledge (they tend to use knowledge already available, which does not necessarily apply to a new problem, such as COVID-19), problem- and/or site-specific problem-solving requires interdisciplinary research –which includes but exceeds the perspectives of any one discipline (Hittner et al., 2019). To study and control COVID-19, as well as other epidemics, interdisciplinary teams could include, at least, biomedical, cartographic, behavioral, logistical, computational, educational and mathematical expertise. Such an approach, it is argued, may facilitate urgently needed training on preparedness.

Author contributions

ALR designed the study. FOF and ALH extracted the data and reviewed the literature. JLF and SDS generated cartographic information. JBH conducted the statistical analysis. All authors contributed to writing of the report.

Funding source

None declared.

Henan (the two provinces included in the rank II node) became activated at the same time and exhibited a similar increase in the number of fatalities over 18 consecutive days. E. Smallworld (long-distance) connections. A map that includes the provincial capitals of Heilongjiang (Harbin) and Henan (Zhengzhou), as well as the shortest link between these cities and Wuhan, Hubei, supports the notion that both short- and long-distance connections took place. Heilongjiang demonstrated smallworld epidemic connections (and synchronicity with Henan) even though the population of Heilongjiang is $\sim$2.5 smaller than that of Henan. F. Directionality. A truncated set of data demonstrates a non-random pattern in the temporal sequence of events: rank I node displayed a linear or exponential growth in the number of fatalities before rank II node deaths began to increase; similarly, the mortality of the rank II node were increasing while the number of rank III node deaths was unchanged. These patterns suggest a 'from-high-to-low rank' directionality in the number of fatalities. G. Earliest data patterns. A truncated plot displays the earliest data patterns of Hubei and Henan fatalities. The first phase does not show a structured or distinct pattern. Instead, both the data collected in Hubei and observations gathered in its northern neighbor, Henan, reveal a horizontal pattern (dark brown rectangle). Only later (about a month later in Hubei, one week later in Henan), a stage characterized by a linear growth in the number of fatalities is observed (red lines), which later becomes quadratic or exponential (blue line). Therefore, to be both effective and less costly, interventions should occur during the earliest (horizontal)stage. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Conflict of interest

None of the authors has any conflict of interest that should prevent the review or publication of this manuscript.

Acknowledgments

The authors appreciate the resources and comments provided by Dr. Steve Strogatz (Cornell University, USA).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ijid.2020.05.049.

References

Associated Press. https://apnews.com/c42eabe1b1e1ba9fcb2ce201cd3abb72 [Accessed 25 February 2020].
CDC confirms possible instance of community spread of COVID-19 in U.S. https://www.cdc.gov/media/releases/2020/s0226-Covid-19-spread.html [Accessed 29 February 2020].
Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. Lancet 2020;395(10226):764–6, doi:http://dx.doi.org/10.1016/S0140-6736(20)30421-9.
China prisons report 500 cases, as virus spreads in South Korea – as it happened. The Guardian; 2020.
6:40pm: Japan says 79 more people have tested positive for coronavirus on Diamond Princess cruise ship. CNBC; 2020 https://www.cnbc.com/2020/02/19/coronavirus-live-updates-china-hubei-deaths.html [Accessed 24 February 2020].
The Government of Hong-Kong Special Administrative Region. https://www.censtatd.gov.hk/hkstat/sub/so20.jsp [Accessed 05 March 2020].
Harvard University. https://worldmap.harvard.edu/chinamap/ [Accessed 25 February 2020].
Hittner JB, Hoogesteijn AL, Fair JM, van Regenmortel MHV, Rivas AL. The Third Cognitive Revolution: the consequences and possibilities for biomedical research. EMBO Rep 2019;20:e47647, doi:http://dx.doi.org/10.15252/embr.201847647.
Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506, doi:http://dx.doi.org/10.1016/S0140-6736(20)30183-5.
Jewell NP, Lewnard JAPhD, Jewell BL. Caution warranted: using the Institute for Health Metrics and Evaluation Model for predicting the course of the COVID-19 pandemic. An Int Med 2020;173:, doi:http://dx.doi.org/10.7326/M20-1565 [Accessed 05 May 2020].
Johns Hopkins University – Center for Systems Science and Engineering. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [Accessed 09 February to 01 March 2020].
Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, et al. Positive RT-PCR test results in patients recovered from COVID-19. JAMA 2020;, doi:http://dx.doi.org/10.1001/jama.2020.2783.
Lauderdale T-L, McDonald LC, Shiau Y-R, Chen P-C, Wang H-Y, Lai J-F, et al. The status of antimicrobial resistance in Taiwan among gram-negative pathogens: the Taiwan Surveillance of Antimicrobial Resistance (TSAR) program, 2000. Diagn Microbiol Infect Dis 2004;48:211–9, doi:http://dx.doi.org/10.1016/j.diagmicrobio.2003.10.005.
Lessons from South Korea's COVID-19 outbreak: the good, bad, and ugly. The Diplomat. https://thediplomat.com/2020/03/lessons-from-south-koreas-covid-19-outbreak-the-good-bad-and-ugly/ [Accessed 11 March 2020].
Li J, Blakeley D, Smith? RJ. The failure of R0. Comput Math Methods Med 2011;527610, doi:http://dx.doi.org/10.1155/2011/527610.
Marchant R, Samia NI, Rosen O, Tanner MA, Cripps S. Learning as we go-an examination of the statistical accuracy of COVID19 daily death count predictions. , doi:http://dx.doi.org/10.1101/2020.04.11.20062257.
Meyers LA. Contact network epidemiology: bond percolation applied to infectious disease prediction and control. Bull Am Math Soc 2007;44:63–86.
National Bureau of Statistics of China. http://data.stats.gov.cn/english/easyquery.htm?cn=E0103 (Accessed 05 March 2020].
Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel 1 coronavirus (COVID-19) infections. medRxiv 2020;, doi:http://dx.doi.org/10.1101/2020.02.03.20019497.
Okeke N, Edelman R. Dissemination of antibiotic-resistant bacteria across geographic borders. Clin Infect Dis 2001;33:364–9, doi:http://dx.doi.org/10.1086/321877.
Paternoster R, Brame R, Mazerolle P, Piquero A. Using the correct statistical test for the equality of regression coefficients. Criminology 1998;36:859–66.
Rivas AL, Fasina FO, Hammond JM, Smith SD, Hoogesteijn AL, Febles AL, et al. Epidemic protection zones: centred on cases or based on connectivity?. Transb Emerg Dis 2012;59:464–9, doi:http://dx.doi.org/10.1111/j.1865-1682.2011.01301.x.
Rivas AL, Tennenbaum SE, Aparicio JP, Hoogesteyn AL, Mohammed H, Castillo-Chávez C<ET-AL. Critical Response Time (time available to implement effective measures for epidemic control): model building and evaluation. Can J Vet Res 2003;67:307–15.
Rivas AL, Fasina FO, Hoogesteyn AL, Konah SN, Febles JL, Perkins DJ, et al. Connecting network properties of rapidly disseminating epizoonotics. PLoS ONE 2012;7:e39778, doi:http://dx.doi.org/10.1371/journal.pone.0039778.
Schaumburg F, Alabi AS, Peters G, Becker K. New epidemiology of *Staphylococcus aureus* infection in Africa. Clin Microbiol Infect 2014;20:589–96, doi:http://dx.doi.org/10.1111/1469-0691.12690.
Tacconelli E, Sifakis F, Harbarth S, Schrijver R, van Mourik M, Voss A, et al. Surveillance for control of antimicrobial resistance. Lancet Infect Dis 2018;18:e99–e106, doi:http://dx.doi.org/10.1016/S1473-3099(17)30485-1.
Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020;eabb6105, doi:http://dx.doi.org/10.1126/science.abb6105.
Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998;393:440–2.