Sensitivity of internet-based surveillance for unexplained death tend to be poor in low-income countries

Kyung-Duk Min[1], Seyoung Kim[2], Yoon Young Cho[2], Sun-Young Kim[1,2*]

[1]*Institute of Health and Environment, School of Public Health, Seoul National University, Seoul 08826, Korea*
[2]*Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul 08826, Korea*

*Corresponding author
Sun-Young Kim
Department of Public Health Sciences,
Graduate School of Public Health,
Seoul National University, Seoul 08826,
Korea
Tel: +82-2-880-2768
E-mail: sykim22@snu.ac.kr

ORCID
Kyung-Duk Min
https://orcid.org/0000-0002-1000-2187
Seyoung Kim
https://orcid.org/0000-0003-3030-1683
Yoon Young Cho
https://orcid.org/0000-0002-4096-6884
Sun-Young Kim
https://orcid.org/0000-0002-5419-4424

Abstract

Global concerns have grown regarding emerging infectious diseases (EIDs) caused by previously unknown pathogens. Considering that strengthening surveillance capacity for unknown diseases is one of the core capacities for preparedness and early response to EIDs, identifying areas with poor capacity could be beneficial to prioritize regions for the improvement of surveillance. In this regard, we aimed to develop prediction models to identify high risk areas for low surveillance capacity for unknown diseases in a global scale. Unexplained death events reported between 2015 and 2019 were collected from two internet-based surveillance systems, ProMED-mail and Global Public Health Intelligence Network. From the reports, the number of reported unexplained deaths at the first report and the time gap between death and report were extracted as measures for sensitivity and timeliness of surveillance capacity, respectively. Using geographical locations of the reports and published global scale spatial data, including demographic, socioeconomic, public health and geographical variables, we fitted two boosted regression tree models to predict regions with the low sensitivity and timeliness. The performance of prediction model for the low sensitivity showed moderate validity, but in terms of the model for timeliness, the performance was unreliable. Therefore, we provided predicted risk only for low sensitivity. The mean predicted risks of low sensitivity were, respectively, 45.2%, 37.4%, 12.5%, and 3.0% in low-income, lower middle-income, upper middle-income, and high-income countries. Enhancing surveillance capacity in low-income countries is highly required, given the predicted low level of sensitivity despite the importance of early response.

Keywords: unexplained death; undiagnosed diseases; internet-based surveillance

INTRODUCTION

Global public health concerns regarding the emergence of novel diseases have increased. The impacts of these novel diseases are often severe because the whole population is immunologically naïve to the emergent diseases, and medical treatments and vaccines are limited. In addition, the steeply increasing number of international travels between countries and continents, in the con-

text of globalization, increases the pandemic risk by the emerging diseases. The coronavirus disease (COVID-19) pandemic is a prime example that shows how an emerging infectious disease (EID) can have a substantial impact on public health [1] and the world economy [2]. To make matters worse, the incidence of an EID is not a rare event. More than 300 EIDs have been reported from 1940 to 2004 [3] and the emergence of infectious diseases such as Hendra, Nipah, Severe Acute Respiratory Syndrome (SARS), Middle East respiratory syndrome (MERS), etc., is still an ongoing challenge that needs to be tackled [4, 5].

Early response is one of the key strategies to minimize the impact of EIDs at the local and global level. Relevant evidence of the benefits of early response have been reported for Ebola [6] and Influenza [7], and it has also been accumulating for COVID-19 in various countries [8]. According to a recent study [9], a three-week delayed response could have caused 18-fold more cases, and a three-week earlier response would have reduced 95% of the cases during the initial phase of COVID-19 in Wuhan, China. In enhancing the capability of early response against EIDs, typical approaches have been the identification of the potential zoonotic pathogen in wildlife [10] or identification of high-risk areas for spill-over events (i.e., the transmission of pathogens from animals to human population) [4, 11]. Considering that about 60% of EIDs are transmitted from animals [3], these efforts could motivate local health authorities to increase preparedness for EIDs or to facilitate the processes of vaccine development before the pathogen encounters the human population.

Strengthening the surveillance capacity against novel infectious diseases or EIDs could be another approach to enhance the capability of early response. Surveillance system with high sensitivity (i.e., detecting disease events based on small number of cases reported), and with high timeliness (i.e., shortening the time gap between occurrence of the events and detection), could initiate interventions in the early phase of the events [12]. In this regard, identifying regions with poor surveillance capacity is important to prioritize areas for the improvement of surveillance system.

In this study, we aimed to provide a prediction map to suggest regions with low surveillance capacity for novel infectious diseases. Specifically, we focused on capacity of internet-based surveillance systems. Although there are several limitations (e.g., heterogeneity of internet access between regions or countries [13]) in using internet-based disease surveillance systems, such systems are gaining popularity as a tool for early detection of epidemics before an outbreak is officially recognized, which initiates epidemiological investigations [14, 15]. Indeed, the SARS outbreak in 2002 demonstrated the potentials of the early detection capacity of internet-based surveillance systems [16]. In the present study, we specifically targeted surveillance capacity for unexplained death events as a proxy for the surveillance capacity on novel infectious diseases. The unexplained death events were defined as a human death case from a suspicious infectious disease without a confirmed diagnosis in the first report. Therefore, only one death could be included. Using unexplained mortality cases rather than morbidity cases, we can minimize potential bias from characteristics of diseases (e.g., severity) when measuring regional surveillance capacity.

MATERIALS AND METHODS

A global-level study was conducted in three steps. First, reports of unexplained death events from 2015 to 2019 (5 years) were collected from the most commonly used web-based surveillance systems, and relevant information, such as geographical locations (i.e., countries, states or districts) and indicators of surveillance capacity (i.e., sensitivity and timeliness), were extracted from the reports. The two indicators for surveillance capacity were the main outcome variables. Second, potential predictor variables for the surveillance capacities for unexplained diseases were collected for all global regions, except for Antarctica. The predictor variables included demographic, socioeconomic, public health, and geographical factors. The study unit was a one by one-degree latitude-longitude grid covering the world (N = 17,666). Third, machine learning algorithm-based prediction models were developed using the extracted surveillance capacity indicators and predictor variables in each grid where the unexplained death events were reported. Subsequently, predicted risk values for lower surveillance capacity were produced for every study unit using the extracted predictor variables in the grids and the developed prediction models.

The reports for unexplained death events were collected from the two internet-based surveillance systems, ProMED-Mail [17] and the Global Public Health Intelligence Network (GPHIN) [18]. ProMED-Mail uses information from media reports, official reports and local observers. The collected reports are then reviewed by analysts or experts before being disseminated to subscribers or published on a website. GPHIN incorporates natural language processing methods to systematically collect information from news articles, media releases and incidence reports and to categorize the information based on pathogen type or hosts etc.

In order to collect reports of unknown diseases from ProMed-Mail, we designed a search string including the terms "undiagnosed OR mysterious OR mystery OR novel OR unknown." For the purpose of gathering reports from GPHIN, we developed a search protocol with the following inclusion criteria: 1) reports published in English only, 2) reports containing at least one type of infectious disease, and 3) titles containing "unknown" or "undiagnosed" or "mysterious." The exclusion criteria for both data sources were as follows: 1) non-first reports, 2) novel subtypes of diseases (e.g., novel strain of norovirus, novel influenza), 3) prion diseases (e.g., posts labeled "novel prion update"), 4) endemic diseases, 5) unknown sources (e.g., unknown source of food-borne infection), and 6) components (e.g., unknown component of drugs).

Two indicators reflecting the surveillance capacity, sensitivity and timeliness, were defined as follows: 1) sensitivity—the number of mortality cases at the first report, and 2) timeliness— the time gap (in terms of days) between incidence of mortality cases and reports (Supplementary Fig. S1). As the target outcome of this study was unknown diseases, various types of diseases could be included with different levels of symptom severity. Considering that the varying levels of symptom severity may affect the evaluation indicators independent of surveillance capacity, only reports with mortality cases were included in this study.

Considering that quantitative investigations of associated factors of surveillance capacity are

still lacking, we assumed that a comprehensive range of potential factors, including regional, demographic, socioeconomic, public health, and geographical factors, could affect the surveillance capacity for unknown diseases. The data for the variables were acquired from various sources (Table 1).

Demographic variables were obtained from the fourth version of the Gridded Population of the World (GPW v4) [19]. GPWv4 is a raster format global-scale data with a resolution of about 1 km. As the data included estimated population sizes in years 2000, 2005, 2010, 2015, and 2020, average values for the two recent years, 2015 and 2020, were used for the analysis. Night-time light levels, regional gross domestic product (GDP), and human development index (HDI) were used to represent the regional socioeconomic status. Night-time light level data was satellite-based remote sensing data acquired from the National Aeronautics and Space Administration's Black Marble night-time light product [20] which was a raster-type data with a 500 m resolution. Higher night-time light levels were assumed to be associated with higher levels of regional economic activities [21]. GDP and HDI were acquired from a published report by Kummu et al. [22] in which the annual raster-type data for GDP and HDI were provided with a resolution of approximately 10 km for the period 1990–2015. Considering the period of this study (2015–2019), only data of 2015 was used in the prediction models. In addition, income-based country classification by the World Bank (i.e., high-income, upper-middle-income, lower-middle-income, and low-income) [23] was included as a categorical variable. National-level variables representing the level of public health system performance or surveillance capacity were also used, such as health expenditure [24] and the average of 13 International Health Regulations (IHR) core capacity scores reported by the World Health Organization [25]. Urban land use was used as a geographical factor. The land use data were acquired from Tuanmu et al. [26], which provided a consensus dataset of land use converging four global land cover products: DISCover, GLC2000, MODIS2005, and GlobCover. Among 12 land use types including forest, water and agricultural land, we considered only urban land use, which is known to be associated with accessibility to medical facilities.

The acquired predictor variables with spatial features were preprocessed to calculate values for each study unit grid. Crop and mask functions were used for clipping the raster data to fit into each grid and the getValue function was used to extract values where raster cells intersect-

Table 1. Dataset used in this study

Category	Variable	Data source
Demographic	Population	NASA SEDAC [19]
Socioeconomic	Night time light level	NASA [19]
	GDP	Kummu et al. [22]
	Human development index	Kummu et al. [22]
	Income based country classification	World Bank [2]
Public health	Health expenditure	World Bank [2]
Geographic	IHR score	World Health Organization [1]
	Urban land use	Tuanmu et al. [26]

GDP, gross domestic product; IHR, International Health Regulation.

ed with each study unit grid. The functions for the preprocessing were obtained from the raster package [27] in R v.4.0.2 [28]. The geographical distribution of the acquired variables was shown in Supplementary Fig. S2 and S3.

Boosted regression tree (BRT) [29] was applied to predict the risk of low surveillance capacity for unexplained deaths on a global scale. As a tree-based machine learning method, BRTs incorporate non-linear associations into high-order interactions between variables and usually produce better predictability than traditional generalized linear models. Previous studies have conducted global-level prediction using the BRT method, but the predictions were for other types of outcome variables (i.e., incidence of emerging zoonotic diseases or antimicrobial resistance) [4, 30]. Although BRT is limited in that it is considered a type of black-box technique and the effect of each predictor on the outcome variables cannot be quantified, relative influence of each variable can be determined. Considering that we used two indicators for measuring surveillance capacity (the number of mortality cases at the first report and the time gap between incidence of mortality cases and reports), two prediction models were fitted, one for each indicator (Model 1 for sensitivity and Model 2 for timeliness). Specifically, two binary outcome variables were used: 1) whether the number of mortality cases in the first reports were equal to ten or higher (Model 1); and 2) whether the time gap between the occurrence of mortality cases and reporting were equal to one week or longer (Model 2). Leave-one-out cross validation was used with area under the curve of the receiver-operator characteristic (AUC of the ROC) to validate predictability of the models. Because the previous global level prediction studies using BRTs produced an AUC of 0.67 [30], we did not determine the predicted risk for low surveillance capacity if our prediction model showed an AUC of less than 0.67.

RESULTS

An initial search retrieved 2,276 and 658 reports of unknown diseases from the two internet-based surveillance systems, ProMed-Mail and GPHIN, respectively. After examining relevance by screening the title and contents of each report and removing duplicates from the data sources, a total of 327 reports remained. Out of the 327 reports, 198 (60.5%) reports included human diseases with and without mortality cases. The remaining 129 reports showed mortality cases only and thus were used for analysis. Out of the 129 reports, a majority (104 reports, 80.6%) were from low-income or lower middle-income countries. Among the 129 reports, 14 did not provide detailed location information other than the country name, and 28 reports did not contain information on the time gap between the occurrence of death and reporting (Fig. 1). The geographical distribution of reported unexplained death events with an assessment of the surveillance capacity by the two indicators are shown in Supplementary Fig. S4.

Results of univariable logistic regression analysis, shown in Tables 2 and 3, provide an overview of the associations between the predictors and the two indicators of the surveillance capacity. In terms of the number of mortality cases shown in the first report, IHR score and all socioeconomic variables, including natural log of GDP per capita, national income-based country classification, natural log of HDI, and night-time light level, showed significant neg-
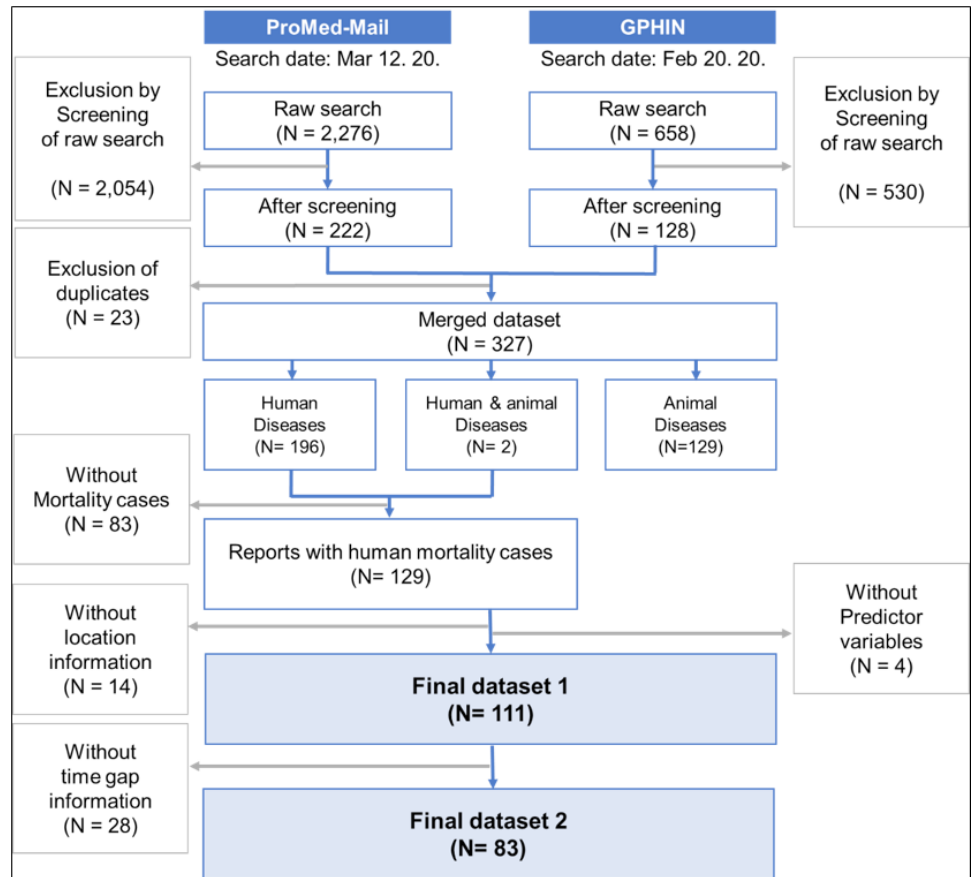
Fig. 1. A flow chart of the selection process for collecting unexplained death events. The main purpose of this study was to predict areas with low capacity of internet-based surveillance for unexplained death and two indicators, sensitivity and timeliness, were used to measure the capacity. Final dataset 1 and 2 illustrated in the figure indicates datasets for sensitivity and timeliness, respectively.

Table 2. Univariable logistic regression analysis for low sensitivity of internet-based surveillance

Variable	Odds ratio	95% Confidence interval	*p*-value	N missing
Log (Population size)	0.843	0.669–1.059	0.140	0
Log (GDP per capita)	0.758	0.604–0.918	0.010	0
Income classification[1]	0.076	0.004–0.384	0.013	0
Log (HDI)	0.038	0.005–0.236	0.001	0
Night-time light level[2]	0.973	0.945–0.995	0.038	0
Health expenditure (2015 US$)	0.937	0.799–1.063	0.358	1
IHR score	0.970	0.951–0.989	0.003	3
Urban land use	0.779	0.435–1.041	0.251	0

Low sensitivity was defined as the number of unexplained death cases in the first reports were equal to ten or higher.
[1] Income-based country classification by World Bank (4 categories; high income, upper middle income, lower income, low income). The odds ratio was for high and upper middle-income levels compared to low and lower middle-income levels as a reference.
[2] Night-time light level values range from 0 to 255, and higher value indicates higher brightness form artificial light.
GDP, gross domestic product; HDI, human development index; IHR, International Health Regulation.

ative associations (Odds Ratios [95% confidence intervals] were 0.758 [0.604–0.918], 0.076 [0.004–0.384], 0.038 [0.005–0.236], 0.973 [0.945–0.995], and 0.970 [0.951–0.989] for natural log of GDP, income-based country classification, natural log of HDI, night-time light level

Table 3. Univariable logistic regression analysis for poor timeliness of internet-based surveillance

Variable	Odds ratio	95% Confidence interval	*p*-value	N missing
Log (Population size)	1.045	0.816–1.351	0.726	29
Log (GDP per capita)	1.009	0.816–1.249	0.936	29
Income classification[1]	0.638	0.214–1.788	0.401	29
Log (HDI)	0.645	0.134–2.951	0.574	29
Night-time light level[2]	0.999	0.983–1.015	0.912	29
Health expenditure (2015 US$)	1.026	0.902–1.171	0.689	29
IHR score	0.999	0.979–1.019	0.890	32
Urban land use	1.054	0.892–1.302	0.541	29

Poor timeliness was defined as the time gap between the occurrence of mortality cases and reporting were equal to one week or longer.

[1] Income-based country classification by World Bank (4 categories; high income, upper middle income, lower income, low income). The odds ratio was for high and upper middle-income levels compared to low and lower middle-income levels as a reference.

[2] Night-time light level values range from 0 to 255, and higher value indicates higher brightness form artificial light.

GDP, gross domestic product; HDI, human development index; IHR, International Health Regulation.

and IHR, respectively). However, none of the associations were significant for the time gap between the occurrence of mortality cases and reporting.

The results of the two prediction models using BRT are as follows: The LOOCV AUC was 0.70 (indicating a moderate validity) in Model 1, but 0.58 (indicating a rather low validity) in Model 2. Socioeconomic or public health-related predictors (night-time light level, health expenditure, HDI, IHR score, and GDP per capita) showed a higher relative influence than the others (Supplementary Fig. S5). As low AUC values were obtained in Model 2, we predicted risk of low surveillance capacity by using Model 1 only (Fig. 2). The averages of predicted risk of low surveillance capacity were 45.2%, 37.4%, 12.5%, and 3.0% in low-income, lower middle-income, upper middle-income, and high-income countries, respectively. In terms of geographical classification, the sub-Saharan African countries showed the highest average predicted risk (43.0%) and North America showed the lowest average predicted risk (2.7%).
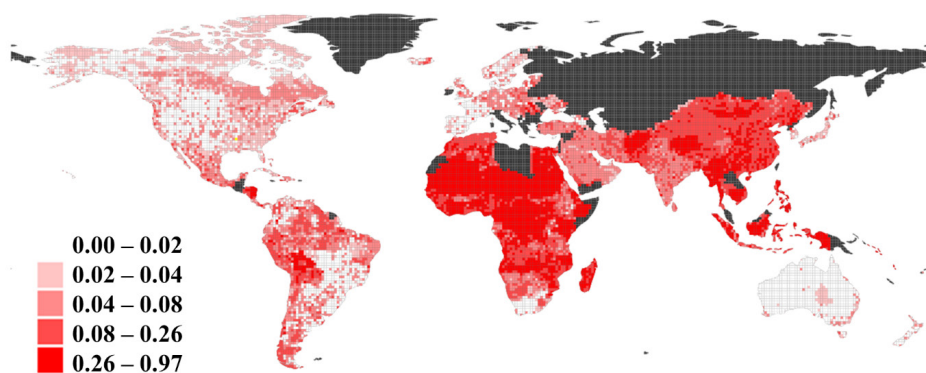

Legend:
- 0.00 – 0.02
- 0.02 – 0.04
- 0.04 – 0.08
- 0.08 – 0.26
- 0.26 – 0.97

Fig. 2. Predicted risk for low sensitivity of internet-based surveillance for unexplained deaths. Low sensitivity was defined as the number of unexplained death cases in the first reports were equal to ten or higher. Boosted regression tree method was employed for the prediction and various published spatial data including demographic, socioeconomic, public health and geographical variables were used as predictors. Areas with grey color showed lack of predictor variables.

DISCUSSION

The purpose of this study was to develop a prediction map representing the risk of low surveillance capacity for unexplained deaths in a global scale, in order to prioritize regions for strengthening surveillance capacity. To this end, we acquired reports of unexplained death events from internet-based surveillance systems and various predictor variables including demographic, socioeconomic, public health, and geographic factors. Surveillance capacity was measured by two indicators; the number of mortality cases during the first report (sensitivity) and the time gap between occurrence of mortality cases and reporting (timeliness). Two prediction models were fitted, one for each indicator, but only the model for predicting sensitivity showed reasonable validity, revealing a high risk of low sensitivity in low income countries and the sub-Saharan region.

The findings of this study, based on the outcomes from the logistic regression models and BRT results using sensitivity measurements, suggest that socioeconomic and public health-related factors can explain the risk of lower sensitivity for unexplained death events detection. The clear differences between regional socioeconomic statuses could be attributed to the availability of different levels of human and financial resources for surveillance by the regions. High disease burden with higher and more frequent mortality events in resource-poor regions could also explain the results. Considering that the risk of novel disease emergence is also high in the low-income and lower-middle-income countries [4], the results imply the urgent demands for improving early detection and response capability in the risk area.

On the other hand, the lack of predictability of Model 2, which was fitted to timeliness measurements, indicates that the current predictor factors are insufficient to explain the timeliness capacity of the surveillance. Latent predictors, which are associated with timeliness but were not used in this study, or a lower sample size than that used for Model 1 could contribute to the lack of predictability. However, the results could also indicate that internet-based syndromic surveillance systems function appropriately in low socio-economic areas, compared to the others, in terms of timeliness. A previous study that evaluated timeliness capacity for reporting of EIDs showed that the time gap between disease onset and report tend to be high in African countries [14]. However, the findings of our study suggest that contrasting results could be derived, presumably, due to the following differences between the studies. First, the previous study used symptom onset to measure timeliness, not incidences of death as used in our study. Second, the time period of the previous study was between 1996 and 2009, but we used data between 2015 and 2019. Third, the target outcome was a WHO-verified outbreak in the previous study, but we employed unexplained death as the target outcome.

In addition, we found that only 18 out of 129 unexplained death events were reported to both ProMed-Mail and GPHIN, suggesting low agreement between the two data sources. The low level of agreement may imply that multiple data sources should be incorporated for practical implementation of internet-based surveillance systems for unknown diseases.

This study has several limitations. First, our study only evaluated the sensitivity and timeliness of internet-based syndromic surveillance systems. We recommend follow-up studies to

assess other attributes such as data quality, cost-effectiveness, predictive value positive, etc. [31] and to evaluate other types of syndromic surveillance systems, which use hospital-based clinical data usually obtained from emergency departments [32]. Second, our analysis could not incorporate any temporal variations; previous studies have suggested that surveillance capacity could be enhanced over time [14]. Third, we included only two internet-based surveillance system in spite of the existence of many other data sources such as Healthmap, Medisys, etc. Incorporation of more data sources of internet-based systems will improve the data quality for this research.

Despite the limitations, to the best of the authors' knowledge, this study is the first to conduct a global-level prediction for low surveillance capacity specifically targeting unexplained death. Our results suggest that enhancing surveillance capacity is particularly important and needed in sub-Saharan Africa and in low-income countries. Recently, the World Bank initiated the West Africa Regional Disease Surveillance Systems Enhancement Project, which aims to strengthen capacity of infectious disease surveillance in West Africa [33]. However, the early detection capacity still needs improvement [34], especially for surveillance sensitivity, as revealed in this study.

SUPPLEMENTARY MATERIALS

Supplementary materials are only available online from: https://doi.org/10.12729/jbtr.2022.23.4.191.

REFERENCES

1. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report – 88. Geneva: World Health Organization; 2020.

2. Maliszewska M, Mattoo A, van der Mensbrugghe D. The potential impact of COVID-19 on GDP and trade: a preliminary assessment. Washington: World Bank; 2020.

3. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. Nature 2008;451:990-993.

4. Allen T, Murray KA, Zambrana-Torrelio C, Morse SS, Rondinini C, Di Marco M, Breit N, Olival KJ, Daszak P. Global hotspots and correlates of emerging zoonotic diseases. Nat Commun 2017;8:1124.

5. Rohr JR, Barrett CB, Civitello DJ, Craft ME, Delius B, DeLeo GA, Hudson PJ, Jouanard N, Nguyen KH, Ostfeld RS, Remais JV, Riveau G, Sokolow SH, Tilman D. Emerging human infectious diseases and the links to global food production. Nat Sustain 2019;2:445-456.

6. Kellerborg K, Brouwer W, van Baal P. Costs and benefits of early response in the Ebola virus disease outbreak in Sierra Leone. Cost Eff Resour Alloc 2020;18:13.

7. Bootsma MCJ, Ferguson NM. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. Proc Natl Acad Sci USA 2007;104:7588-7593.

8. Walker PG, Whittaker C, Watson O, Baguelin M, Ainslie KEC, Bhatia S, Boonyasiri A, Boyd

O, Cattarino L, Cucunubá Z, Cuomo-Dannenburg G, Dighe A, Donnelly CA, Dorigatti I, van Elsland S, FitzJohn R, Flaxman S, Fu H, Gaythorpe K, Geidelberg L, Grassly N, Green W, Hamlet A, Hauck K, Haw D, Hayes S, Hinsley W, Imai N, Jorgensen D, Knock E, Laydon D, Mishra S, Nedjati-Gilani G, Okell LC, Riley S, Thompson H, Unwin J, Verity R, Vollmer M, Walters C, Wang HW, Wang Y, Winskill P, Xi X, Ferguson NM, Ghani AC. The global impact of COVID-19 and strategies for mitigation and suppression. London: Imperial College London; 2020.

9. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, Wesolowski A, Santillana M, Zhang C, Du X, Yu H, Tatem AJ. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 2020;585:410-413.

10. Karesh WB, Mazet JAK, Clements A. Upstream surveillance: scanning for zoonoses in wildlife. Int J Infect Dis 2014;21:28.

11. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. Nature 2017;546:646-650.

12. Cameron AR, Meyer A, Faverjon C, Mackenzie C. Quantification of the sensitivity of early detection surveillance. Transbound Emerg Dis 2020;67:2532-2543.

13. Choi J, Cho Y, Shim E, Woo H. Web-based infectious disease surveillance systems and public health perspectives: a systematic review. BMC Public Health 2016;16:1238.

14. Chan EH, Brewer TF, Madoff LC, Pollack MP, Sonricker AL, Keller M, Freifeld CC, Blench M, Mawudeku A, Brownstein JS. Global capacity for emerging infectious disease detection. Proc Natl Acad Sci USA 2010;107:21701-21706.

15. Magid A, Gesser-Edelsburg A, Green MS. The role of informal digital surveillance systems before, during and after infectious disease outbreaks: a critical analysis. In: Radosavljevic V, Banjari I, Belojevic G (eds.). Defence against bioterrorism. Dordrecht: Springer; 2018. p. 189-201.

16. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. Can Med Assoc J 2009;180:829-831.

17. Yu VL, Madoff LC. ProMED-mail: an early warning system for emerging diseases. Clin Infect Dis 2004;39:227-232.

18. Mawudeku A, Blench M. Global public health intelligence network (GPHIN). Proceedings of 7th Conference of the Association for Machine Translation in the Americas; 2002; Cambridge, MA.

19. NASA Socioeconomic Data and Applications Center [SEDAC]. Gridded Population of the World (GPW), v4 [Internet]. 2017 [cited 2022 Sep 15]. Available from: https://beta.sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev10

20. Román MO, Wang Z, Sun Q, Kalb V, Miller SD, Molthan A, Schultz L, Bell J, Stokes EC, Pandey B, Seto KC, Hall D, Oda T, Wolfe RE, Lin G, Golpayegani N, Devadiga S, Davidson C, Sarkar S, Praderas C, Masuoko EJ. NASA's black marble nighttime lights product suite. Remote Sens Environ 2018;210:113-143.

21. Mellander C, Lobo J, Stolarick K, Matheson Z. Night-time light data: a good proxy measure for economic activity? PLOS ONE 2015;10:e0139779.

22. Kummu M, Taka M, Guillaume JHA. Gridded global datasets for gross domestic product and human development index over 1990–2015. Sci Data 2018;5:180004.

23. World Bank. World bank country and lending groups [Internet]. 2020 [cited 2022 May 30]. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

24. World Bank. World bank open data [Internet]. 2020 [cited 2022 May 30]. Available from: https://data.worldbank.org/

25. World Health Organization. Average of 13 international health regulations core capacity scores, 1st version of the questionnaire [Internet]. 2020 [cited 2022 May 30]. Available from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/-average-of-13-international-health-regulations-core-capacity-scores-1st-version-of-the-questionnaire

26. Tuanmu MN, Jetz W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. Glob Ecol Biogeogr 2014;23:1031-1045.

27. Hijmans RJ, van Etten J, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, Lamigueiro OP, Bevan A, Racine EB, Shortridge A. Package "raster" [Internet]. 2015 [cited 2022 Sep 15]. Available from: https://mran.microsoft.com/snapshot/2015-08-14/web/packages/raster/raster.pdf

28. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020.

29. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol 2008;77:802-813.

30. Van Boeckel TP, Pires J, Silvester R, Zhao C, Song J, Criscuolo NG, Gilbert M, Bonhoeffer S, Laxminarayan R. Global trends in antimicrobial resistance in animals in low- and middle-income countries. Science 2019;365:eaaw1944.

31. Groseclose SL, Buckeridge DL. Public health surveillance systems: recent advances in their use and evaluation. Annu Rev Public Health 2017;38:57-79.

32. Henning KJ. What is syndromic surveillance? Morb Mortal Wkly Rep 2004;53:7-11.

33. The World Bank. Regional disease surveillance systems enhancement (REDISSE) [Internet]. 2016 [cited 2022 July 2]. Available from: https://projects.worldbank.org/en/projects-operations/project-detail/P154807

34. Ravi SJ, Snyder MR, Rivers C. Review of international efforts to strengthen the global outbreak response system since the 2014–16 West Africa ebola epidemic. Health Policy Plan 2019;34:47-54.