An Epidemic Simulation with a Delayed Stochastic SIR Model Based on International Socioeconomic-Technological Databases

Aki-Hiro Sato
Department of Applied Mathematics and Physics
Graduate School of Informatics, Kyoto University
Honmachi, Yoshida, Sakyo-ku, Kyoto 606-8501 Japan
Email: aki@i.kyoto-u.ac.jp

Hidefumi Sawai
Universal Communication Research Institute
National Institute of Information and Communications Technology
910 North Building,
3-1 Ofuka-cho, Kita-ku, Osaka 530-0011 Japan
Email: sawai@nict.go.jp

Isao Ito
Graduate School of Medicine, Kyoto University,
Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan
Email:isaoito@kuhp.kyoto-u.ac.jp

Kentaro Iwata
Graduate School of Medicine, Kobe University,
7-5-2 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan
Email:kiwata@med.kobe-u.ac.jp

***Abstract*—This study proposes an epidemic model for Ebola virus disease (EVD) based on a combination between a delayed stochastic SIR model and a metapopulation network model. Our proposed model consists of a set of stochastic differential equations of state variables, such as the number of susceptible persons, the number of infected patients both inside and outside of isolation wards, and the number of recovered persons both inside and outside of isolation wards, the number of deaths. We collected socioeconomic, technological-environmental data, such as OAG aviation timetable data, grid statistics on world population estimates provided by the Socioeconomic Data and Applications Center (SEDAC), the number of cases and deaths by EVD reported in 2014 by the World Health Organization (WHO) and economic statistics provided from the World DataBank by the World Bank. Linking these databases, we calibrated model parameters. We then conducted a numerical simulation by using aviation timetables and calculated the potential numbers of infectious persons and deaths. We found that the pandemic would not be completely prevented in industrialized countries by medical efforts alone. We conducted sensitivity analysis for the numbers of cases and deaths in terms of possible scenarios for medical, socioeconomic, and transport dimensions. We conclude that the medical efforts in industrialized countries can control the pandemic rate and that international aviation transport can reduce the number of passengers from places where epidemic outbreaks occur, and delay the beginning of the pandemic. The numerical simulation model can be extended to epidemic diseases other than EVD.**

***Keywords*-global air transportation, epidemic spread, data-driven numerical simulation, grid statistics data**

I. Introduction

The possibility of epidemic outbreaks, originating from technological advantages such as the international air transportation system, needs to be considered. Severe outbreaks result from person-to-person transmission of a disease with a high fatality rate. Natural geographical barriers against infectious diseases can now be easily overcome through trade and travel. This problem has already been recognized from the epidemic outbreaks of Severe Acute Respiratory Syndrome (SARS) (2002–2003) [1] and H1N1 swine-originated pandemic influenza (2009) [2]. To quantify risks of pandemic scenario, we must use Big Data and conduct numerical simulation linking different types of Big Data. This is a typical challenge of Big Data simulation. This paper studies how Big Data analytics contributes to conducting epidemiological simulation.

Severe epidemic spread of Ebola virus disease (EVD) has been recognized as a considerable threat to global security since December 2013. The World Health Organization (WHO) declared the Ebola epidemic an international health emergency on August 8, 2014 [3]. Furthermore, on September 18, 2014, the UN Security Council declared the Ebola outbreak in West Africa a threat to international peace and security.

As of June 2015, the WHO said it would hold an emergency meeting to consider South Korea's outbreak of Middle East Respiratory Syndrome (MERS), which it described as "large and complex." The virus has infected 138 people in South Korea and killed 14 of them since it was first diagnosed on May 20 in a businessman who had returned from a trip to the Middle East. Authorities have sealed off at least two hospitals and about 4,000 people are in quarantine, either at home or in medical facilities in South Korea. Experts predict more cases will emerge until the incubation period, which is believed to be up to two weeks for everyone exposed to an infected patient, is over [4].

The outbreak is the largest outside Saudi Arabia, where the disease was first identified in humans in 2012, and has stirred fears in Asia of a repeat of a 2002 and 2003 scare, when SARS killed about 800 people worldwide [5]. MERS is caused by a coronavirus from the same family as the one

that causes SARS. As of May 24, 2015, worldwide, a total of 1,134 cases and 427 deaths (case fatality rate 37.7%) have been reported. MERS is more deadly than SARS because an overall of case fatality of SARS was 14% to 15% [1]. There is no cure or vaccine at least for now.

Cases like the recent outbreaks of EVD and MERS have been seen in the modern era, in which we face on an unprecedented risk of pandemic by the severe infectious disease. Moreover, we must recognize its negative impacts on the world economy. Both human and economic losses due to severe epidemics may reach a global level. However, since both monetary and human costs for preventing an epidemic spread are expensive, we must demand cooperation among sectors. To do so, we need to characterize costs and risks of epidemic spreads, and share this information with various types of stakeholders. Data-driven numerical simulation is useful for this purpose.

In this article, we attempt to develop a numerical simulation model of the epidemic spread, based on international socioeconomic-technological data on population, aviation systems, and geography. From 2014 to 2015, we computed time series for the probable numbers of cases and deaths in each city having an airport, and simulated potential scenarios where the epidemics transited into a pandemic.

This article is organized as follows. Section II shows fundamental properties of EVD and existing study on EVD. Section III introduces a stochastic model of epidemics with regard to human mobility through international aviation networks. Section IV mentions characteristics of the global air transport system and its descriptive statistics. Section V shows results of numerical simulation. Section VI discusses several scenarios and conditions to prevent the EVD pandemic. Section VII is devoted to conclusions.

II. Ebola virus disease

27,443 cases of Ebola including 11,207 deaths have been reported as of June 24, 2015 since the Ebola outbreak began in March, 2014 [6]. The Centers for Disease Control and Prevention (CDC) made an estimation in June, 2014, extrapolating trends to January 20, 2015. Without additional interventions or changes in community behavior, the model estimated that countries where the outbreaks of EVD are raging would have approximately 550,000 cases (1.4 million when corrected for underreporting) [7]. The current situation of Ebola is not worse than the worst estimation by CDC in June, 2014. However, the cumulative numbers of reported cases and deaths are still increasing.

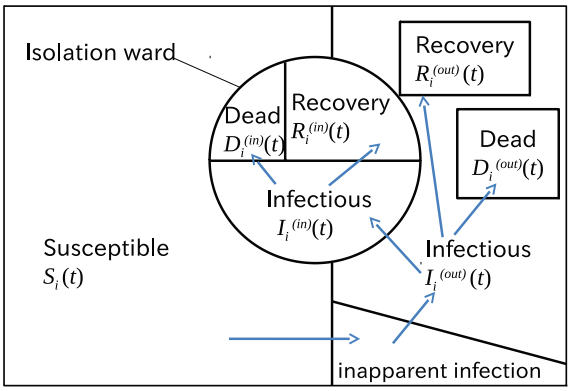
Because of the development and spread of commercial jet airplanes, we can travel from one place to another across our planet within 48 hours. This implies that an infectious disease can easily spread from endemic countries. Consequently, many industrialized countries, which had been free of many epidemics, have suffered from the severe spread of infectious diseases in recent years.

The strain of Ebola virus raging in the latest severe outbreak seems different from other strains of Ebola viruses that have been found in the past. The case fatality rate of the past EVD outbreak is estimated to be as high as 90%, with an incubation period of typically 2 to 10 days [8]. However, the case fatality rate of the latest EVD outbreaks is less than those of previous outbreaks, with a case fatality rate estimated to be 70.8% in the report by the WHO Ebola Response Team [9]. For Sierra Leone, the case fatality rate was almost the same as that estimated in the WHO report with (74%) [10]. With adequate treatment for EVD patients, the case fatality rate could be reduced to as low as 43% [11]. This treatment has been applied since October, 2014, in three West African countries.

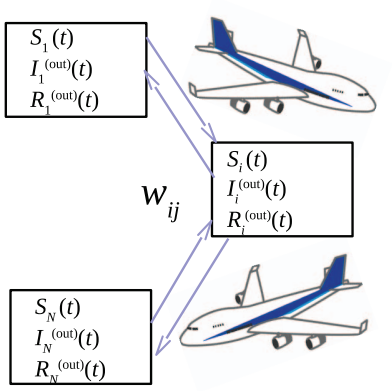
Furthermore, the incubation period for the current outbreak was longer than for past documented cases of EVD. It had been believed that the longest incubation period of EVD was 21 days, and that usually symptoms appeared after 2 to 10 days [8]. In the latest outbreaks, however, the incubation period has been found to be 6 to 12 days [10], with an estimated average of 11.4 days [9]. Normally, in an uncontrollable outbreak, the number of cases and the number of deaths show exponential increase in the initial stage. If the number of cases reaches a level which cannot be handled in an economical or operational sense, then a disaster will occur, impacting the whole planet. Therefore, it is important to understand and infer the numbers of infections and deaths for EVD epidemics from a global point of view.

III. Model

Our numerical simulation is based on a combination of a delayed stochastic SIR model [12], [13] and a metapopulation network model [2]. The susceptible, infectious, recovered or removed (SIR) model was developed by Kermack and McKendrick [14]. One might assume four kinds of states; susceptible (S), infectious (I), recovered (R), and dead (D). The susceptible population transits into an infectious population with a transition probability based on contact with an infectious person. An infectious person moves to either a recovered state with a transition probability $(1-\rho)\beta$ or a dead state with a transition probability $\rho\beta$, where β is a transition probability for an infectious person to move to other states and a fatality rate represents ρ ranging from 0 to 1. Since medical efforts and personal-protection gear for health-care workers can decrease the infection rate, a value of α in industrialized countries should be less than α for developing countries. Moreover, medical efforts applied to patients can decrease fatality rates, so a value of ρ in industrialized countries should also be less than the value of ρ for developing countries. Since an effort to detect infectious persons outside isolation wards contributes to preventing epidemic spread, artificial removal rate γ depends on areas. Since transition rate β and incubation period τ



(a)



(b)

Figure 1. Conceptual illustration of the model. (a) Causal chain of state transitions at city i. (b) Metapopulation network model among N cities with an airport.

are properties of a specific strain of Ebola virus, they are uncontrollable and take unique values independent of place.

Therefore, the difference between outbreaks in industrialized and developing countries is said to be infection, fatality and artificial removal rates.

As shown in Figure 1, epidemic spread occurs when an infectious person comes in contact with a susceptible person, causing the susceptible person to become an infectious person. We denote the number of susceptible persons in city i $(i = 1, \ldots, M)$ at time t as $S_i(t)$, the number of infected patients outside isolation wards as $I_i^{(out)}(t)$, the number of recovered persons outside isolation wards as $R_i^{(out)}(t)$, the number of deaths outside isolation wards as $D_i^{(out)}(t)$, the number of infected patients inside isolation wards as $I_i^{(in)}(t)$, the number of recovered persons inside of isolation wards as $R_i^{(in)}(t)$, and the number of deaths inside isolation wards as $D_i^{(in)}(t)$.

We assumed six types of parameters to characterize an epidemic disease: α_i, β, ρ_i (treatment), ρ^* (non-treatment), γ_i, and τ, which represent, respectively, infection; transition; fatality (treatment and non-treatment); artificial removal rates, and an incubation period at city i. Therefore, the total number of the population outside the isolation wards can be calculated as $N_i^{(out)}(t) = S_i(t) + I_i^{(out)}(t) + R_i^{(out)}(t)$. This mechanism can be seen as a delayed SIR model and we expressed it as a set of stochastic differential equations, as follows:

$$
\begin{aligned}
dS_i &= -\alpha_i \frac{S_i(t) I_i^{(out)}(t)}{N_i^{(out)}(t)} dt - \sqrt{\alpha_i \frac{S_i(t) I_i^{(out)}(t)}{N_i^{(out)}(t)}} dW_i^{(\alpha)}(t) \\
&\quad + \Big\lfloor \textstyle\sum_{j=1}^{M} w_{ji} S_j(t) dt - \sum_{j=1}^{M} w_{ij} S_i(t) dt \\
&\quad + \textstyle\sum_{j=1}^{M} \sqrt{w_{ji} S_j(t)} dV_{ji}(t) \\
&\quad - \textstyle\sum_{j=1}^{M} \sqrt{w_{ij} S_i(t)} dV_{ij}(t) \Big\rfloor \\
dI_i^{(out)} &= \alpha_i \frac{S_i(t-\tau) I_i^{(out)}(t-\tau)}{N_i^{(out)}(t-\tau)} dt \\
&\quad + \sqrt{\alpha_i \frac{S_i(t-\tau) I_i^{(out)}(t-\tau)}{N_i^{(out)}(t-\tau)}} dW_i^{(\alpha)}(t-\tau) \\
&\quad - \beta I_i^{(out)}(t) dt - \sqrt{\beta I_i^{(out)}(t)} dW_i^{(\beta_o)}(t) \\
&\quad - \gamma_i I_i^{(out)}(t) dt - \sqrt{\gamma_i I_i^{(out)}(t)} dW_i^{(\gamma)}(t) \\
&\quad + K_q \Big\lfloor \textstyle\sum_{j=1}^{M} w_{ji} I_j^{(out)}(t) dt - \sum_{j=1}^{M} w_{ij} I_i^{(out)}(t) dt \\
&\quad + \textstyle\sum_{j=1}^{M} \sqrt{w_{ji} I_j^{(out)}(t)} dV_{ji}(t) \\
&\quad - \textstyle\sum_{j=1}^{M} \sqrt{w_{ij} I_i^{(out)}(t)} dV_{ij}(t) \Big\rfloor \\
dD_i^{(out)} &= \rho^* \beta I_i^{(out)}(t) dt + \rho^* \sqrt{\beta I_i^{(out)}(t)} dW_i^{(\beta_o)}(t) \\
dR_i^{(out)} &= (1-\rho^*) \beta I_i^{(out)}(t) dt \\
&\quad + (1-\rho^*) \sqrt{\beta I_i^{(out)}(t)} dW_i^{(\beta_o)}(t) \\
&\quad + \Big\lfloor \textstyle\sum_{j=1}^{M} w_{ji} R_j^{(out)}(t) dt - \sum_{j=1}^{M} w_{ij} R_i^{(out)}(t) dt \\
&\quad + \textstyle\sum_{j=1}^{M} \sqrt{w_{ji} R_j^{(out)}(t)} dV_{ji}(t) \\
&\quad - \textstyle\sum_{j=1}^{M} \sqrt{w_{ij} R_i^{(out)}(t)} dV_{ij}(t) \Big\rfloor \\
dI_i^{(in)} &= \gamma_i I_i^{(out)}(t) dt + \sqrt{\gamma_i I_i^{(out)}(t)} dW_i^{(\gamma)}(t) \\
&\quad - \beta I_i^{(out)}(t) dt - \sqrt{\beta I_i^{(out)}(t)} dW_i^{(\beta_i)}(t) \\
dD_i^{(in)} &= \rho_i \beta I_i^{(in)}(t) dt + \rho_i \sqrt{\beta I_i^{(in)}(t)} dW_i^{(\beta_i)}(t) \\
dR_i^{(in)} &= (1-\rho_i) \beta I_i^{(in)}(t) dt + (1-\rho_i) \sqrt{\beta I_i^{(in)}(t)} dW_i^{(\beta_i)}(t) \\
&\quad + \Big\lfloor \textstyle\sum_{j=1}^{M} w_{ji} R_j^{(in)}(t) dt - \sum_{j=1}^{M} w_{ij} R_i^{(in)}(t) dt \\
&\quad + \textstyle\sum_{j=1}^{M} \sqrt{w_{ji} R_j^{(in)}(t)} dV_{ji}(t) \\
&\quad - \textstyle\sum_{j=1}^{M} \sqrt{w_{ij} R_i^{(in)}(t)} dV_{ij}(t) \Big\rfloor
\end{aligned}
\tag{1}
$$

where $W_i^{(\alpha)}(t)$, $W_i^{(\beta_i)}(t)$, $W_i^{(\beta_o)}(t)$, $W_i^{(\gamma)}(t)$, $V_{ij}(t)$ are Wiener processes with $\langle W_i^{(s)}(t) \rangle = 0$,

$\langle W_i^{(s_1)}(t_1) W_j^{(s_2)}(t_2) \rangle = \delta_{i,j}\delta_{s_1,s_2} \min(t_1, t_2)$, $\langle V_{ij}(t) \rangle = 0$, $\langle V_{ik}(t_1) V_{jl}(t_2) \rangle = \delta_{i,j}\delta_{k,l} \min(t_1, t_2)$ ($s_1, s_2 \in \{\alpha, \beta_o, \beta_i, \gamma\}$). $\lfloor \cdot \rfloor$ means the maximum integer less than $\cdot$. $\delta_{i,j}$ is Kronecker delta function.

As shown in Figure 1 (a), we assume two types of areas outside and inside isolation wards and two SIR models inside and outside isolation wards. The dynamics outside isolation wards are expressed by $S_i(t)$, $I_i^{(out)}(t)$, $R_i^{(out)}(t)$ and $D_i^{(out)}(t)$. The dynamics inside isolation wards are expressed by $I_i^{(in)}(t)$, $R_i^{(in)}(t)$ and $D_i^{(in)}(t)$. We further assume that we can catch infectious persons and treat them in isolation wards with rate γ_i, which leads to a flow of infectious persons expressed by $I_i^{(in)}$ and $I_i^{(out)}(t)$.

Susceptible, infectious and recovered persons who live outside isolation wards can travel abroad via a global air transport system. As shown in Figure 1 (b), this can be expressed by a metapopulation network model for $S_i(t)$, $I_i^{(out)}(t)$ and $R_i^{(out)}$. Each term contains fluctuation, which is expressed by a multiplicative noise term including a Wiener process.

Table I
VARIABLES OF A NUMERICAL SIMULATION MODEL FOR EPIDEMICS. SUBSCRIPT i EXPRESSES CITY i AND t REPRESENTS TIME.

variable	explanation
$S_i(t)$	# of susceptible persons
$I_i^{(out)}(t)$	# of infected patients
$R_i^{(out)}(t)$	# of recovered persons outside isolation wards
$D_i^{(out)}(t)$	# of deaths outside isolation wards
$I_i^{(in)}(t)$	# of infected patients inside isolation wards
$R_i^{(in)}(t)$	# of recovered persons inside isolation wards
$D_i^{(in)}(t)$	# of deaths inside isolation wards

Table II
A SET OF PARAMETERS

parameters	variables	before 1 September 2014	after 1 September 2014
incubation period	τ [day]	11.4	11.4
infection rate	α [1/day]	0.08718985	0.01918177
natural removal rate	β [1/day]	0.00853659	0.00853659
artificial removal rate	γ_i [1/day]	0.01	0.01
fatality rate inside (industrialized country)	ρ_A	0.25	0.25
fatality rate inside (developing country)	ρ_B	0.708	0.43
fatality rate outside	ρ^*	0.708	0.708
vulnerability (industrialized country)	Vul	0.001	0.001
transition probabilities	w_{ij}	OAG data	OAG data
load factor	K_c	0.65	0.65
quarantine factor	K_q	0.45	0.45

Tables I and II show variables and parameters. In our numerical simulation model of epidemics, we have seven variables in each city: ($S_i(t)$, $I_i^{(out)}(t)$, $R_i^{(out)}(t)$, $D_i^{(out)}(t)$, $I_i^{(in)}(t)$, $R_i^{(in)}(t)$, $D_i^{(in)}(t)$). In general, $S_i(t)$, $I_i^{(out)}(t)$, and $R_i^{(out)}(t)$ are unobservable, and are latent variables. We need to infer them from observable variables $I_i^{(in)}(t)$, $R_i^{(in)}(t)$, $D_i^{(in)}(t)$, and $D_i^{(out)}(t)$. Furthermore, we have six kinds of parameters to characterize the epidemic disease. Some parameters (β, ρ_i, ρ^*, and τ) can be determined from medical knowledge and statistics on past events. However, others (α_i and γ_i) should be estimated from observations based on the observable variables $I_i^{(in)}(t)$, $R_i^{(in)}(t)$, $D_i^{(in)}(t)$, and $D_i^{(out)}(t)$. w_{ij} represents transition probability from i to j through air transport networks. This can be inferred from passenger statistics or timetable data.

Assuming a unit of time as a day, we define the transition probability from city i to city j per day as

$$w_{ij} = \frac{F_{ij}}{N_i^{(out)}(0)}, \tag{2}$$

where F_{ij} is the daily number of passengers from city i and city j. We focus on air transportation only since its mobility is faster than other transport systems and the most dominant in global mobility.

Consider two types of categories for countries: industrialized countries A and developing countries B. If city i is included in category A, then α_i should be less than $\alpha_j = \alpha$ ($j \in B$). We assume that fatality rate (non-treatment) ρ^* is independent of the categories. However, the fatality rate (treatment) $\rho_i = \rho_A$ at city i ($i \in A$) also should be less than $\rho_j = \rho_B$ ($j \in B$), i.e., $\rho_A < \rho_B$. Therefore, we express infectious and fatality rates (treatment) at city i,

$$\alpha_i = \begin{cases} \alpha \times Vul & (i \in A) \\ \alpha & (i \in B) \end{cases}, \tag{3}$$

$$\rho_i = \begin{cases} \rho_A & (i \in A) \\ \rho_B & (i \in B) \end{cases}, \tag{4}$$

where Vul ($0 \leq Vul \leq 1$) represents vulnerability ranging from 0 to 1. $Vul = 0$ expresses perfect protection, and $Vul = 1$ no protection. The vulnerability Vul for industrialized countries was set as 0.001. We also use Vul for conducting sensitivity analysis in terms of Vul.

UN's Travel and Transport Task Force on Ebola virus disease outbreak officially recommended exit screening for Ebola [15], [16]. The exit screening of all persons departing affected countries through international airports, seaports and major land crossings is recommended by WHO and could reduce the numbers of people with symptoms from travelling from the countries with high levels of Ebola transmission. Therefore, we introduced a factor K_q ($0 \leq K_q \leq 1$) for migration of infectious persons.

We assumed $K_q = 0.45$ independently of airports. This means that the exit screening at airports in affected countries may reduce infected travelers up to 45%. The long incubation period of EVD (the longest period is believed to be 21 days [8]) can cause infected patients to 'leak out'.

IV. Characteristics of the global air transport system

In order to estimate w_{ij}, we need to have knowledge about passenger flow. Traditionally, the passenger flows have been modelled by the gravity model [17]–[19]: $F_{ij} = cx_i^{b_1}(d)x_j^{b_2}(d)a_{ij}^{b_3}$, where $x_i(d)$ and a_{ij} represent the number of people living around airport i within radius d (km) and the geodesic distance between airport i and airport j. c is a positive constant and b_k $(k = 1, 2, 3)$ are coefficients. They are estimated from the past passenger flow data by using regression analysis.

This proposes that the number of passengers from airport i to airport j may depend on populations of cities i and j and the geodesic distance a_{ij}. This implies that passenger flows to and from large cities are large, and that those to and from small cities are small.

Recently, we have been able to approach this problem from an alternative dimension. The passenger flow from airport i to airport j may be estimated from capacities (the number of seats) computed from timetable data of a global aviation network.

Then, the number of passengers from airport i to airport j may be approximated as

$$F_{ij} = K_c J_{ij}, \tag{5}$$

where K_c is average load factor ranging from 0 to 1 and J_{ij} is the daily number of seats from airport i to airport j, which is computed from OAG aviation timetable data [20]. $K_c = 0.65$ is typical but depend on flights, seasons, days, and districts.

Table III shows descriptive statistics of monthly global aviation networks in 2014 [20]. The worldwide coverage of the OAG timetable data is 94.3% as of October 2014. The coverage of each region in 2014 is as follows: North America (99%), South America (90%), Europe (97%), Middle East & Africa (79%), Asia & Pacific (88%). The regional coverage ranges from 79% to 99%, depending on districts.

As shown in Table III, we showed the number of airports, the total number of routes, the total number of flights, and the total number of seats included in monthly timetable data from January to December 2014. The network structures vary in time. The number of airports used in each month is about 3,680. We found that the number of available connections ranges from 38,000 to 44,300. The total monthly number of flights is from 2,378,000 to 2,907,000. The total monthly number of seats range from 318 million to 398 million. The mean in-degree and out-degree of airports (the average number of flights landing at and taking-off from an airport) were estimated as values ranging from 10 to 12. The mean path length is about 4.1. This means that a pair of airports can be connected with about 4 changes.

Table IV shows the average number of available seats per day departing from the airport for the top 10 countries and 3 countries where EVD spread was observed [20]. The United States was the most active country, where 2,447,869.22 seats flied per day. The second most active country was China, with 1,368,010.70 seats per day. The third most active country was Japan, with 479,083.22 seats per day.

Table V shows the number of seats departing from the airport for the top 10 airports and 3 airports where EVD spread had happened. The number of seats departing from Atlanta was 144,904.32 seats per day. The second most active airport was Beijing, China with 142,964 seats per day. The third was Tokyo International Airport in Japan, with 131,969.06 seats per day.

Table IV
THE AVERAGE NUMBER OF SEATS PER DAY DEPARTING FROM A COUNTRY FOR THE TOP 10 COUNTRIES AND 3 COUNTRIES WHERE EVD SPREAD WAS OBSERVED IN 2014. THESE WERE CALCULATED FROM DATA IN JANUARY 2014.

country	seats/day	ratio to the world total
US	2,447,869.225664	21.831117%
CN	1,368,010.709697	12.200489%
JP	479,083.225806	4.272664%
BR	418,923.258063	3.736132%
IN	343,756.000001	3.065759%
GB	325,363.677451	2.901729%
ID	307,845.451617	2.745494%
DE	302,311.871000	2.696143%
AU	280,102.774183	2.498073%
FR	237,467.387099	2.117833%
SL	1,274.774193	0.011369%
LR	1,073.419354	0.009573%
GN	1,004.193548	0.008956%

Table V
THE DAILY NUMBER OF SEATS DEPARTING FROM AN AIRPORT FOR THE TOP 10 AIRPORTS. THESE WERE CALCULATED FROM DATA IN JANUARY 2014.

airport	country	seats/day	ratio to the world total
ATL	US	144,904.322574	1.292317%
PEK	CN	142,964.548386	1.275017%
HND	JP	131,969.064515	1.176955%
DXB	AE	125,938.419354	1.123171%
LHR	GB	119,063.903221	1.061861%
HKG	HK	110,075.322578	0.981698%
LAX	US	107,828.354838	0.961658%
ORD	US	103,718.677419	0.925006%
SIN	SG	103,071.935492	0.919238%
CGK	ID	101,127.870967	0.901900%
FNA	SL	1,274.774193	0.011369%
CKY	GN	1,004.193548	0.008956%
ROB	LR	961.580645	0.008576%
MLW	LR	111.838709	0.000997%

V. Simulation

The numerical simulation was executed based upon computational resources of the HPCI system (Institute of Statistical Mathematics). We used parallel computation techniques for different ensembles due to mutual independence among tasks. We assigned different cores to simulate Eq. (1) with different random seeds. We conducted a numerical

Table III
THE DESCRIPTIVE STATISTICS OF MONTHLY GLOBAL AVIATION NETWORKS IN 2014

month	#airports	#routes	#flights	#seats	mean in-degree	mean out-degree	mean path length
January/2014	3,672	38,602	2,600,058	349,049,121	10.512527	10.512527	4.222798
February/2014	3,663	38,415	2,378,364	318,901,378	10.487305	10.487305	4.219340
March/2014	3,682	39,080	2,667,665	357,558,837	10.613797	10.613797	4.197044
April/2014	3,685	40,800	2,655,991	359,074,363	11.071913	11.071913	4.188108
May/2014	3,670	41,935	2,758,588	374,199,404	11.426431	11.426431	4.165208
June/2014	3,690	43,736	2,752,694	374,011,449	11.852575	11.852575	4.115193
July/2014	3,687	44,243	2,906,234	397,190,537	11.999729	11.999729	4.104424
August/2014	3,690	44,271	2,903,803	398,874,253	11.997561	11.997561	4.114957
September/2014	3,696	43,400	2,737,858	375,040,614	11.742424	11.742424	4.148268
October/2014	3,671	42,054	2,782,234	380,620,602	11.455734	11.455734	4.151767
November/2014	3,661	39,008	2,567,941	350,869,864	10.655012	10.655012	4.193759
December/2014	3,640	38,787	2,665,038	365,761,938	10.655769	10.655769	4.187635

simulation with the actual international air transportation networks [20] and population estimates [21].

We extracted 3,916 ($M = 3,916$) places having an airport that is connected with others in airline timetable data for each month in 2014 [20]. From Eqs. (2) and (5), we estimated transition probability w_{ij} from the number of seats for each month, which are $M \times M$ asymmetric matrices.

The initial population $N_i^{(out)}(0)$ was set as the total number of people who live within $d = 200$ km of an airport. The population of each city was computed from 2.5 arc-minute grid square data on world population estimates in 2010 provided by Socioeconomic Data and Applications Center (SEDAC) [21], including $K = 29,652,480$ records. Suppose that pop_k represents the population in grid square k and that $\boldsymbol{r}_k$ is its representative position. Then, the initial population $N_i^{(out)}(0)$ is computed as

$$N_i^{(out)}(0) = \sum_{k=1}^{K} pop_k H(d - |\boldsymbol{r}_k - \boldsymbol{r}_i|), \qquad (6)$$

where $\boldsymbol{r}_i$ denotes representative position of city (airport) i. $H(\cdot)$ represents Heaviside function, which is defined as

$$H(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases} .$$

Figure 2 shows populations of people who live around an airport for six megacities.

The initial number of susceptible people, $S_i(0)$, was set as the initial population $N_i(0)$, excepting three airports: Conakry (CKY) in Guinea, Freetown (FNA) in Sierra Leone, and Monrovia (ROB) in Liberia. The initial conditions of the numerical simulation were set as the actual number of cases and deaths on June 10, 2014, which is shown in Table VI.

The fatality rate in industrialized countries is estimated from cases number and deaths number at the time of 4 November 2014 [22]. Since we observed 4 deaths for 16 cases (we ignored cases in treatment) as shown in Table VII, we assume that $\rho_A = 4/16 = 0.25$. Fatality rate in developing countries ρ_B was set as 0.708 before October 2014 following the report from the WHO Ebola Response Team [9] and ρ_B was set as 0.43 after October 2014 following the report by Bah et al. [11]. The UNMEER response seems to have been working in the three countries of West Africa and can reduce infection rate α_i at these countries.

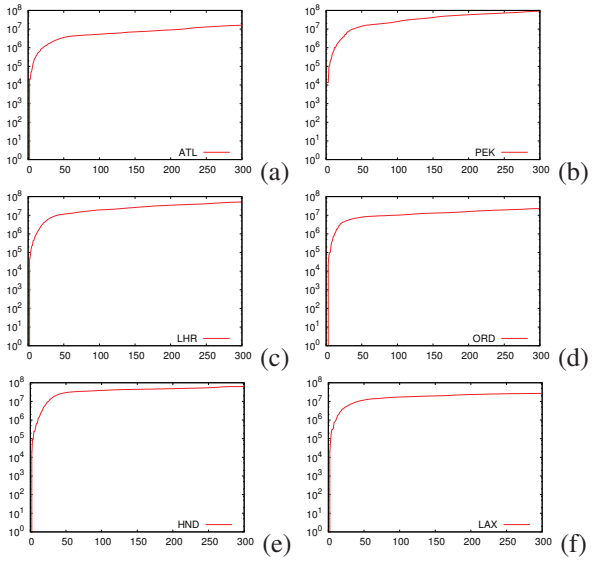


Figure 2. The relationship between population of people who live around an airport within radius and the distance for six megacities. (a) Hartsfield Jackson Atlanta International Airport (ATL), ATLANTA (GA, US), (b) Beijing Capital International Airport (PEK), BEIJING (CN), (c) London Heathrow Airport (LHR), LONDON (GB), (d) Chicago O'Hare International Airport (ORD), CHICAGO (IL,US), (e) Tokyo International Airport (HND), TOKYO (JP), and (f) Los Angeles International Airport (LAX), LOS ANGELES (CA, US).

Table VI
THE INITIAL CONDITIONS OF THE NUMERICAL SIMULATION. THE NUMBERS OF CASES AND DEATHS WERE REPORTED BY WHO ON JUNE 10, 2014.

City	IATA	Country	cases [persons]	deaths [persons]
Conakry	CKY	Guinea	372	236
Freetown	FNA	Sierra Leone	13	9
Monrovia	ROB	Liberia	89	7

We changed infection rate α_i for developing countries after October 1, 2014.

We obtained values of β and γ_i by fitting the actual number of cases and deaths reported by WHO [6] between 25 March 2014 to 10 June 2014. The average incubation period τ was set as 11.4 days, based on the estimation by the WHO Ebola Response Team [9]. We summarized a set of parameters used in our numerical simulation in Table II.

Table VII
CASES OF EBOLA IN INDUSTRIALIZED COUNTRIES. *DATE OF EBOLA DIAGNOSIS.

Job title	Arrival date	State	Country
Aid worker	Aug. 2, 2014	Recovered	USA
Missionary	Aug. 2, 2014	Recovered	USA
Doctor	Sept. 5, 2014	Recovered	USA
Doctor	Sept. 9, 2014	Recovered	USA
Visitor	Sept. 30, 2014*	Died	USA
NBC Cameraman	Oct. 6, 2014	Recovered	USA
Nurse	Oct. 11, 2014*	Recovered	USA
Nurse	Oct. 15, 2014*	Recovered	USA
Doctor	Oct. 23, 2014*	In treatment	USA
Priest	Aug. 7, 2014	Died	Spain
Missionary	Sept. 22, 2014	Died	Spain
Nurse	Oct. 6, 2014*	Recovered	Spain
Nurse	Sept. 19, 2014	Recovered	France
Medical worker	About Nov. 2, 2014	In treatment	France
Nurse	Aug. 24, 2014	Recovered	Britain
Doctor	Aug. 27, 2014	Recovered	Germany
Physician	Oct. 3, 2014	In treatment	Germany
U.N. medical official	Oct 9, 2014	Died	Germany
Aid worker	Oct. 6, 2014	Recovered	Norway

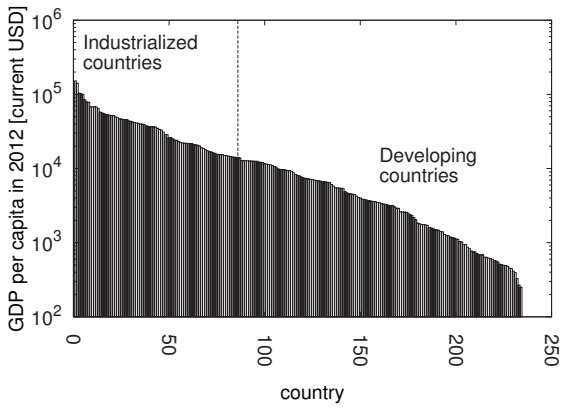


Figure 3. Rank plots of GDP per capita for 234 countries and territories. We used the latest GDP per capita estimated from economic performance from 2000 to 2013. A dashed line shows a boundary between industrialized and developing countries.

Next, we will explain the definitions of industrialized and developing countries. Economic performance has a positive correlation with medical capacities and technologies. We assume that a developing country is a country where economic performance per person is not large. In this sense, GDP per capita, which is computed as a value of GDP for a country divided by the country's population, is a representative quantity for determining the country's health care levels.

Figure 3 shows a rank plot for GDP per capita of 234 countries and territories. The data were collected from World Databank by the World Bank [23]. We assumed a country where the GDP per capita is more than 4,000 USD to be an industrialized country (A). A developing country is assumed to have a GDP per capita of less than 4,000 USD (B). By these definitions, we identified 85 developing countries and 149 industrialized countries from GDP per capita, as measure in current (2012) USD.

Table VIII
MONTHLY TRAVEL CAPACITIES FROM THE MOST AFFECTED COUNTRIES (CONAKRY(CKY), FREETOWN (FNA), AND MONROVIA (ROB)) TO OTHER PLACES, AND THE NUMBER OF PASSENGERS ESTIMATED FROM OAG TIMETABLE DATA IN DECEMBER 2014.

operator	origin	destination	distance (km)	seats/month
AF	CKY	NKC	974.519218	3536
AF	CKY	CDG	4618.774386	2771
HF	CKY	ABJ	1170.330857	1656
AT	CKY	CMN	2702.088012	4884
SN	CKY	FNA	115.512553	2556
DN	CKY	DKR	709.334385	700
AT	FNA	ROB	408.663575	159
AT	FNA	CMN	2796.644626	1272
E2	FNA	DKR	822.396809	588
HF	FNA	ROB	408.663575	536
HF	FNA	ABJ	1089.090896	603
SN	FNA	BRU	4956.972413	2556
YO	FNA	DKR	822.396809	600
YO	FNA	ACC	1476.950283	750
YO	FNA	ABJ	1089.090896	600
YO	FNA	ROB	408.663575	750
AT	ROB	CMN	3013.357268	1431
HF	ROB	FNA	408.663575	603
HF	ROB	ABJ	720.900381	536
SN	ROB	BRU	5133.758056	2374
YO	ROB	FNA	408.663575	750
YO	ROB	ABJ	720.900381	750

We further regarded flight suspension of several connections. Table VIII[1] shows flights and their capacity which are available from airports contained in endemic countries during December, 2014. We found that the total number of seats departing from three countries was 30,961 seats/month (998.74 seats/day),

Several airline companies suspended their connecting flights with three airports (Conakry (CKY) in Guinea, Freetown (FNA) in Sierra Leone, and Monrovia (ROB) in Liberia) after July, 2014. Figure 4 (a) shows the number of seats departing from the most affected countries (CKY, FNA, and ROB). The number of seats departing from three airports decreased steeply after July, 2014.

The number of seats departing from CKY was recovering from November, 2014. Figure 4 (b) shows the number of cities with non-stop flights from the most affected countries

[1]Operator abbreviations use AF: Air France, HF: Air Cote d'Ivoire, AT: Royal Air Maroc, SN: Brussels Airlines N.V., DN: Senegal Airlines, E2: Eagle Atlantic Airlines Limited, YO: Heli Air Monaco.

(CKY, FNA and ROB). The number of cities with non-stop flights from FNA and CKY decreased after July 2014. The number of cities from ROB increased from July to August, but it became 4 in December 2014.

Therefore, we assume that the number of seats departing from and arriving at the airports in the most affected countries remained at the same level as in December, 2014, after January, 2015.

We conducted sensitivity analysis on the effects of quarantine in airports. We computed 10 ensembles which have different random seeds by using parallel computation techniques. Figure 5 (a) shows the average number of cities with more than 10 patients from November 1, 2014, to June 30, 2015, regarding effects on exit screening to epidemic spread. As shown in Figure 5 (b), the average numbers of cities with more than 10 patients on December 31, 2015 and December 31, 2016 strongly depended on vulnerability Vul. If Vul is less than 0.25, we will not expect to observe the pandemic scenario by the end of 2016.

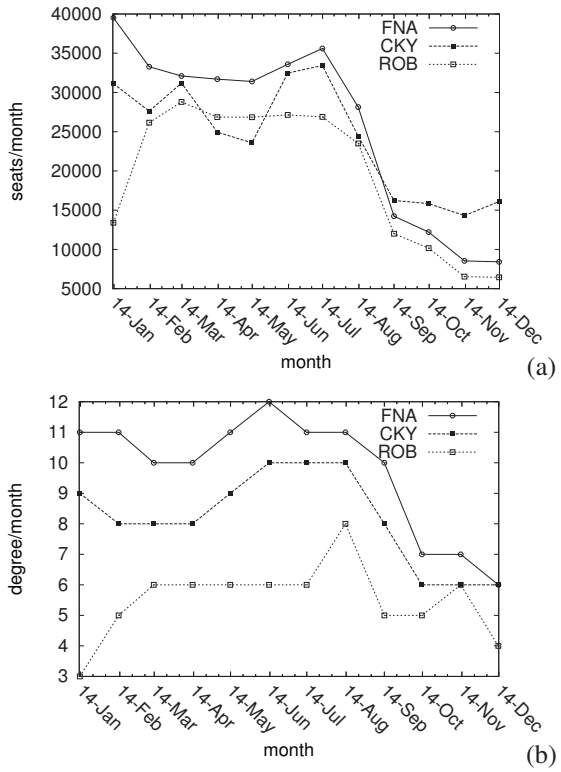


Figure 4. (a) The monthly number seats departing from the most affected countries (Conakry (CKY) in Guinea, Freetown (FNA) in Sierra Leone, and Monrovia (ROB) in Liberia). (b) The number of cities reachable from the most affected countries in every month.

VI. Discussion

If the infection rates in industrialized countries is completely zero, several cities in developing countries would become epidemic only. This is because the basic reproduction

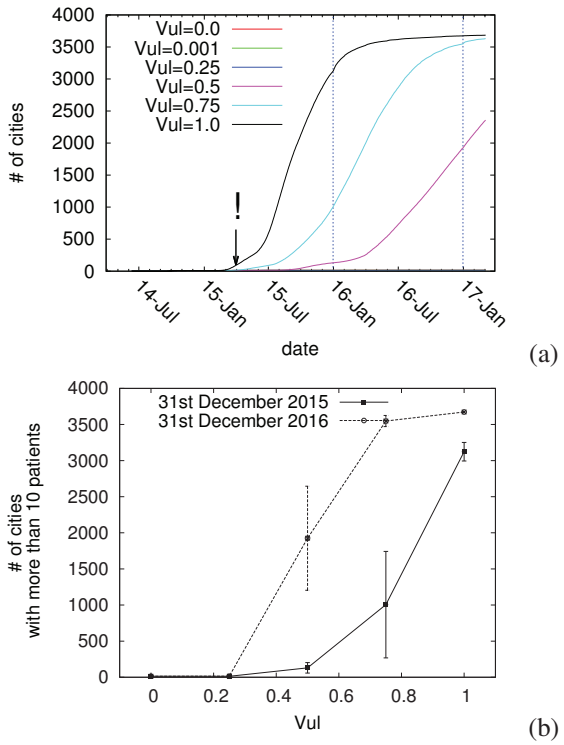


Figure 5. (a) The number of cities with more than 10 patients averaged over nine samples at $Vul = 0.0$, 0.001, 0.25, 0.5, 0.75, and 1.0 for the period from March, 2014, to March, 2017. (b) The number of cities with more than 10 patients averaged over nine samples on December 31, 2015, (solid curve) and December 31, 2016, (dashed curve) in terms of Vul. Error bars show 1-σ range.

number in developing countries may be more than 1, however, international travel of people in developing countries is partial. The infected people from endemic countries travel and migrate between cities in developing countries. These are reasons why the pandemic does not start if the infection rate in industrialized countries is zero. Furthermore, Figure 5 (a) shows that the pandemic does not happen till December, 31, 2016 if the infection rate in industrialized countries is 25% less than in developing countries.

Thus, we need to make efforts to reduce the infection rate in both industrialized and developing countries in order to prevent the pandemic. To do so, industrialized countries should assist in improving medical capacities in developing countries for EVD prevention before an epidemic in developing countries happens.

In the worst-case scenario ($Vul = 1.0$), the number of cities with more than 10 patients on July 1, 2015 was estimated to be 583.22 (std. = 207.52). In the best-case scenario ($Vul = 0.0$), the number would reach 13.22 (std. = 0.62). These numbers may be used for estimating the number of vaccine doses delivered to medical staffs, who need immunization. By mid-2015, WHO expected that around 200,000 Ebola vaccine doses would be produced [24]. This

would be a safe amount even in the worst-case scenario.

It could be a sufficient number of vaccine doses to stop secondary infection from patients to health care workers in each city at the initial stage. All we have to do is to vaccinate medical staffs and persons who have the potential to come in contact with EVD patients, so as to prevent secondary transmissions. This could contribute to keeping the infection rate small. It was also confirmed that favipiravir (Toyama Chemical, Japan) has increased benefit in patients with low Ebola viraemia compared with patients with high viraemia [25].

The flight bans worked well in preventing infected people from reaching other cities. If we conduct strategic bans of international air aviation and manage the international flow of people, then the pandemic would be delayed slightly. The economic losses by strategic bans seem to be enormous. However, EVD pandemics would cause a more serious crisis than short economic losses from the strategic bans.

Our method has several limitations. The first is the complexity of parameter calibration. We needed to generate a flow matrix and initial conditions based on several assumptions. The uniformity of population in each city also resulted in a limitation of data resolution. The regional dependence of parameters for the SIR model essentially reflects the basic reproduction number R_0, and seasonality of the number of passengers strongly influences the tendency for epidemic spread. Our future work is to develop a method to estimate the basic reproduction number R_0 based on the proposed model expressed in Eq. (1). The second limitation originates from uncertainty of future traffic and flows of people. We assumed that the present traffic flow should be similar to the past flow. Under this assumption, we infer the current flow from the gravity model with an adequate set of parameters estimated from the past number of passengers.

Currently, we used flight volume data and assumed a homogeneous load factor in order to infer future passenger flows. However, this provides the upper limit of passenger flows only, and might underestimate or overestimate future passenger flows. The load factor should depend on flights and seasons. We should consider heterogeneity of load factors. However, such extensions make it more difficult to calibrate parameters. Moreover, the load factor data can be obtained from the past number of passengers and the past number of seats. The future load factor should be inferred based on a model. A correlation between the number of seats and the number of passengers may be used. This is a future work.

The ratio of passengers who travel with Ebola virus significantly affects the pandemic speed. In fact, there is a difficulty in determining potential Ebola patients without symptoms. Therefore, we may observe Ebola patients in countries other than endemic countries repeatedly.

The third limitation is related to uniformity of human behavior. We always have free will and make decision from both current and past experiences and information. Therefore, if we use results of numerical simulation for warning or precaution, and announce the projection, then the future flow of people may be influenced, and thus change the future outcome. This is also deeply related to circular causality, resulting from an interaction between decision-making and projection.

VII. Conclusion

We developed a numerical simulation model for Ebola virus disease (EVD) based on the susceptible, infectious, removed (SIR) model, with flows of people in international air transportation networks. We calibrated parameters of the SIR model from aviation timetable data purchased from OAG [20], grid statistics on world population estimates provided by SEDAC [21], the number of cases and deaths by EVD reported in 2014 by WHO [6] and economic statistics provided by the World Bank [23]. We simulated the number of cases and the number of deaths for 3,916 areas having an airport with regard to heterogeneity of infectious rates in industrialized and developing countries.

We conducted sensitivity analysis in terms of medical efforts in industrialized countries. We found that the infection rate in industrialized countries is sensitive to epidemic spread. The results of numerical simulation implied that the pandemic does not happen till December, 31, 2016 if the infection rate in industrialized countries is 25% less than in developing countries. We confirmed that the number of cities where some infected patients are confirmed eventually increases after a given date, which slightly depends on the passenger flows. We estimated the number of cities where we should deliver several doses of vaccine from the numerical simulation.

This study contributes to providing insights on a combination of an efficient vaccination procedure and strategic bans from an international point of view. As future work, we need to use high-resolution data of aviation transportation. If we can use data on both the real-time and precise number of passengers, our numerical simulation model can be used as a real-time monitoring system for prevention controls as well as scenario-based simulation.

Acknowledgement

This work is financially supported by a Grants-in-Aid for Scientific Research (KAKENHI) (C) (#25390152). This research used computational resources of the HPCI system provided by (Institute of Statistical Mathematics) through the HPCI System Research Project (Project ID:hp140076) and (Project ID:hp150106).

References

[1] WHO. “Update 49 – SARS case fatality ratio, incububation period,” May 7, 2003. [Online] Available: http://www.who.int/csr/sarsarchive/2003_05_07a/en/. Accessed on 30 June, 2015.

[2] D. Brockmann and D. Helbing, "The Hidden Geometry of Complex, Network-Driven Contagion Phenomena", *Science*, vol. 342, no. 6164, pp. 1337–1342, Dec. 2013. DOI: 10.1126/science.1245200.

[3] WHO. "WHO Statement on the Meeting of the International Health Regulations Emergency Committee Regarding the 2014 Ebola Outbreak in West Africa," WHO, Aug. 8, 2014. [Online] Available: http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en/, Accessed on August 10, 2014.

[4] J. Kim, "WHO calls emergency meeting on 'large, complex' South Korea MERS outbreak," Reuters, Jun. 13, 2015. [Online] Available: http://www.reuters.com/article/2015/06/13/us-health-mers-southkorea-idUSKBN0OS2HI20150613. Accessed on June 23, 2015.

[5] S. Su, G. Wong, Y. Liu, G.F. Gao, S. Li, and Y. Bi, "MERS in South Korea and China: a potential outbreak threat?", *Lancet*, vol. 385, no. 9985, pp. 2349–2350, Jun. 2015. DOI: 10.1016/S0140-6736(15)60859-5.

[6] WHO. Ebola Situation Reports. [Online] Available: http://apps.who.int/ebola/ebola-situation-reports. Accessed on June 30, 2015.

[7] M.I. Meltzer, C.Y. Atkins, S. Santibanez, B. Knust, B.W. Petersen, E.D. Ervin, S.T. Nichol, I.K. Damon, and M.L. Washington, "Estimating the Future Number of Cases in the Ebola Epidemic – Liberia and Sierra Leone, 20142015," *CDC Morbidity and Mortality Weekly Report, Supplement*, vol. 63, no. 3, pp. 1–14, 2014. [Online] Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/su6303a1.htm.

[8] H. Feldmann and T.W. Geisbert, "Ebola hemorrhagic fever", *Lancet*, vol. 377, no. 9768, pp. 849–862, Mar. 2011. DOI: 10.1016/S0140-6736(10)60667-8.

[9] WHO Ebola Response Team, "Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections," NEJM, vol. 371, pp. 1481–1495, Oct. 2014. DOI: 10.1056/NEJMoa1411100.

[10] J.S. Schieffelin et al.,"Clinical Illness and Outcomes in Patients with Ebola in Sierra Leone," *NEJM*, vol. 371, pp. 2092–2100, Nov. 2014. DOI: 10.1056/NEJMoa1411680.

[11] E.I. Bah et al., "Clinical Presentation of Patients with Ebola Virus Disease in Conakry, Guinea," *NEJM*, vol. 372, pp. 40–47, Jan. 2015. DOI: 10.1056/NEJMoa1411249.

[12] Y. Maeno, "Discovering network behind infectious disease outbreak", *Physica A*, vol. 389, no. 21, pp. 4755-4768, Nov. 2010. DOI: 10.1016/j.physa.2010.07.014.

[13] G. Chowell, N.W. Hengartner, C. Castillo-Chavez, P.W. Fenimore, and J.M. Hyman, "The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda," *J. Theoret. Biol.*, vol. 229, no. 1, pp. 119–126, Aug. 2004. DOI: 10.1016/j.jtbi.2004.03.006.

[14] W.O. Kermack and A.G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. Roy. Soc. Lond. A*, vol. 115, no. 77, pp. 1337–1342, Aug. 1927. DOI: 10.1098/rspa.1927.0118.

[15] II Bogoch, MI Creatore, MS Cetron, JS Brownstein, N Pesik, J Miniota, T Tam, W Hu, A Nicolucci, S Ahmed, JW Yoon, I Berry, SI Hay, A Anema, AJ Tatem, D MacFadden, M German, K Khan, "Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 west African outbreak", The Lancet 21 October 2014(Article in Press DOI: 10.1016/S0140-6736(14)61828-6)

[16] Statement from the Travel and Transport Task Force on Ebola virus disease outbreak in West Africa, UNWTO, PR No.:PR14078 Released on 10 November 2014. [Online] Available: http://media.unwto.org/press-release/2014-11-10/statement-travel-and-transport-task-force-ebola-virus-disease-outbreak-west, Accessed on 25 November 2014

[17] G.K. Zipf, "The P_1P_2/D hypothesis: On the intercity movement of persons," *Am. Social. Rev.*, vol. 11, no. 6, pp. 677–686, Dec. 1946.

[18] W-S. Jung, F. Wang, and H.E. Stanley, "Gravity model in the Korean highway," *Europhys. Lett.*, vol. 81, no. 4, p. 48005, Feb. 2008.

[19] O. Sivrikaya and E. Tunc, "Demand Forecasting for Domestic Air Transportation Turkey", *TOTJ*, vol. 7, pp. 20–26, May. 2013. DOI: 10.2174/1874447820130508001.

[20] OAG, http://www.oag.com

[21] Center for International Earth Science Information Network - CIESIN - Columbia University, United Nations Food and Agriculture Programme - FAO, and Centro Internacional de Agricultura Tropical - CIAT. 2005. Gridded Population of the World, Version 3 (GPWv3): Population Count Grid, Future Estimates. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). [Online] Available: http://dx.doi.org/10.7927/H42B8VZZ. Accessed on July 24, 2014.

[22] The New York Times Ebola Factsheet, "What is being done to improve medical treatment in Africa?" *the New York Times*, Updated Nov. 2, 2014. [Online] Available: http://www.nytimes.com/interactive/2014/07/31/world/africa/ebola-virus-outbreak-qa.html. Accessed on November 9, 2014.

[23] World Databank by the World Bank. [Online] Available: http://databank.worldbank.org/data/home.aspx. Accessed on 24 September 2014.

[24] K. Kelland, S. Nebehay, and T. Miles, "WHO expects around 200,000 Ebola vaccine doses by mid-2015," Reuters, Oct. 24, 2014. [Online] Available: http://www.reuters.com/article/2014/10/24/health-ebola-vaccines-who-idUSL6N0SJ25S20141024. Accessed on October 28, 2014.

[25] M.V. Herp, H. Declerck, T. Decroo, "Favipiravir – a prophylactic treatment for Ebola contacts?," *Lancet*, vol. 385, no. 9985, p. 2350, Jun. 2015. DOI: 10.1016/S0140-6736(15)61095-9.