MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **12**, Number **5**, October **2015**

doi:10.3934/mbe.2015.12.1055

pp. **1055–1063**

# EBOLA OUTBREAK IN WEST AFRICA: REAL-TIME ESTIMATION AND MULTIPLE-WAVE PREDICTION

Xiang-Sheng Wang and Luoyi Zhong

Department of Mathematics
Southeast Missouri State University
Cape Girardeau, MO 63701, USA

(Communicated by Yang Kuang)

Abstract. Based on the reported data until 18 March 2015 and numerical fitting via a simple formula of cumulative case number, we provide real-time estimation on basic reproduction number, inflection point, peak time and final outbreak size of ongoing Ebola outbreak in West Africa. From our simulation, we conclude that the first wave has passed its inflection point and predict that a second epidemic wave may appear in the near future.

1. **Introduction.** Ebola virus disease (EVD) is a severe disease in humans which has infected nearly 25 thousand individuals and claimed more than ten thousand deaths during the recent outbreak in West Africa, according to the report of World Health Organization dated 18 March 2015 [3, 12]. The most affected countries are Guinea, Liberia and Sierra Leone. This study aims to provide some real-time estimations on the outbreak in these three countries using the reported cumulative case data. Specifically, we will estimate the following quantities:

1. basic reproduction number $R_0$, which is defined as the average new cases caused by a single infective individual during one infectious period;
2. inflection point $t_c$, which marks the time when the increment speed of cumulative case numbers starts to slow down;
3. final outbreak size $K$, which indicates the total number of infectious cases throughout the outbreak wave.
4. peak time $t_p$, which is defined as the critical time when daily infectious number reaches its maximum.

All of these indicators provide quantitative information about severity of a disease outbreak.

2. **Methods.** Following [10], we study the epidemic model:

$$S'(t) = -\frac{\beta S(t) I(t)}{S(t) + I(t)}; \text{ and } I'(t) = \frac{\beta S(t) I(t)}{S(t) + I(t)} - \gamma I(t), \tag{1}$$

2010 *Mathematics Subject Classification.* Primary: 92B05; Secondary: 62P10.

*Key words and phrases.* Ebola outbreak in West Africa, basic reproduction number, inflection point, final outbreak size, peak time.

The first author is partially supported by the faculty start-up fund of Southeast Missouri State University.

where $S(t)$ and $I(t)$ are the numbers of susceptible and infective individuals at time $t$, respectively. The constant $\beta$ denotes the transmission rate of the disease, and the constant $\gamma$ corresponds to the removal rate of infective individuals. The basic reproduction number [6, 9] is given by

$$R_0 = \frac{\beta}{\gamma}. \tag{2}$$

It is noted that a disease outbreak occurs if and only if $R_0 > 1$. The differential system (1) can be solved explicitly and its solution is given by

$$\begin{aligned} S(t) =& K[1 + (1 - 1/R_0)e^{\gamma(R_0-1)(t-t_c)}]^{-R_0/(R_0-1)}; \\ I(t) =& K[1 + (1 - 1/R_0)e^{\gamma(R_0-1)(t-t_c)}]^{-1/(R_0-1)} \\ & - K[1 + (1 - 1/R_0)e^{\gamma(R_0-1)(t-t_c)}]^{-R_0/(R_0-1)}, \end{aligned} \tag{3}$$

where $K$ and $t_c$ are two constants of integration. Now, we define the cumulative infective case number at time $t$ as

$$C(t) = \int_{-\infty}^{t} \frac{\beta S(r) I(r)}{S(r) + I(r)} dr. \tag{4}$$

From (1) and (3), we have

$$C(t) = K - K[1 + (1 - 1/R_0)e^{\gamma(R_0-1)(t-t_c)}]^{-R_0/(R_0-1)}. \tag{5}$$

Here, the constant $K = C(\infty)$ has the biological meaning of final outbreak size. It can be verified that $C''(t_c) = 0$. Hence, $t_c$ is the inflection point of $C(t)$. We remark that the inflection point $t_c$ is related to but different from another commonly used quantity: the peak time, denoted by $t_p$. The peak time is defined as the time when infective case number achieves its maximum, namely, $I'(t_p) = 0$. It follows from (3) that

$$t_p = t_c + \frac{\ln R_0}{\gamma(R_0 - 1)}. \tag{6}$$

In the case when $R_0$ is close to 1, namely, $\ln R_0 \approx R_0 - 1$, we can approximate the difference $t_p - t_c$ by $1/\gamma$. Thus, the peak time occurs about one infectious period after the inflection point [10].

Richards' empirical model [8] was suggested to provide real-time estimation of a disease outbreak; see [7] for example. However, some of the parameters in Richards' model do not have clear biological meanings [10]. The advantage of formula (5) is that all of the parameters in this formula have significant biological interpretations. We will use the explicit formula of $C(t)$ in (5) to fit the reported cumulative case numbers of 2014 Ebola outbreak in West Africa.

As pointed out in [10], one should fix the value of $\gamma$, the removal rate of infective individuals, to resolve possible over-fitting problems. Note that $1/\gamma$ can be regarded as the infectious period which characterizes the average duration of an individual being infective. In most cases, an individual is removed from the infective group either by recovery or death. For the fatal cases of Ebola virus disease, death usually occurs between 6 and 16 days (with mean 7.5 days) after onset of symptom; and for the non-fatal cases, patients may improve their symptoms at around day 6 but need more time to recover [4]. Convalescent patients may still be infective because the Ebola virus RNA may remain in the body fluid for a couple of weeks even though the risk of transmission from them is low [2]. It is thus reasonable to assume the infectious period to be 7.5 days with some possible perturbations in the interval between 6 and 16 days. In our simulation, we first fix $1/\gamma = 7.5$ days to

estimate the basic reproduction number $R_0$, inflection point $t_c$, peak time $t_p$ and final outbreak size $K$ using reported cumulative case data of Ebola virus in Guinea, Liberia and Sierra Leone, respectively [3, 12]. Also, we provide the 95% confidence intervals of each estimated parameter value using bootstrap method. Next, we vary the value of the parameter $1/\gamma$ from 6 to 16 days and investigate the sensitivity of fitted parameter values.

In the simulation, we use least squares method to estimate the parameter values and bootstrap method to calculate the confidence interval for each fitted parameter. To be specific, given a sequence of data points $(t_k, C_k)$ for $k = 1, \cdots, n$, where $C_k$ is the cumulative case number at time $t_k$, we fix the value of $\gamma$ and use the formula (5) to find optimal values for $K$, $t_c$ and $R_0$ that minimize the least squares $\sum_{k=1}^{n}[C(t_k) - C_k]^2$. Furthermore, we sample a set of $n$ indices $k_1, k_2, \cdots, k_n$ with each index randomly choosing from $1, \cdots, n$. These indices are not necessarily to be distinct. We then fit the parameter values using the sampled data $(t_{k_i}, C_{k_i})$ with $i = 1, \cdots, n$, and repeat the procedure $N = 1000$ times so as to obtain $N$ fitted values of $K$, $t_c$ and $R_0$. The corresponding mean, variance, as well as confidence interval for each parameter, are calculated using the following formulas

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i;\ \sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2};\ X_{\pm} = \bar{X} \pm \frac{\sigma_X}{\sqrt{N}}\Phi^{-1}(1 - \alpha/2),$$

where $X_1, \cdots, X_N$ are estimated parameter values from $N$ samples, $\alpha$ is the significance level of confidence interval which is commonly chosen as 5%, $\Phi^{-1}$ is the inverse of cumulative distribution function of the standard normal distribution, and $X_{\pm}$ are the upper and lower bounds of the confidence interval.

3. **Results.** The basic reproduction number $R_0$ is estimated as 1.116 (95% CI: 1.115-1.116) for Guinea, 1.226 (95% CI: 1.225-1.228) for Liberia, and 1.181 (95% CI: 1.181-1.182) for Sierra Leone. The inflection point $t_c$ is estimated as 21 November 2014 for Guinea, 24 October 2014 for Liberia, and 28 November 2014 for Sierra Leone. As shown in Table 1, the lengths of 95% confidence intervals for the estimated inflection points are no more than one day. The tight confidence intervals for $R_0$ and $t_c$ are due to the fact that the structure of the S-shaped curve does not vary much even though some data are perturbed or missing.

| | Guinea | Liberia | Sierra Leone |
|---|---|---|---|
| $K$ | 3268 [3257, 3274] | 8630 [8605, 8660] | 11227 [11198, 11253] |
| $R_0$ | 1.116 [1.115, 1.116] | 1.226 [1.225, 1.228] | 1.181 [1.181, 1.182] |
| $t_c$ | 266 [265, 266] | 238 [237, 238] | 273 [272, 273] |
| $t_p$ | 273 [272, 273] | 244 [244, 245] | 279 [279, 280] |

TABLE 1. Estimated parameter values with 95% confidence intervals. Here, day 1 corresponds to 1 March 2014. So, days 266, 238 and 273 correspond to 21 November 2014, 24 October 2014, and 28 November 2014, respectively.

The fitted curves together with reported cumulative case data are illustrated in Figure 1 (Guinea), Figure 2 (Liberia) and Figure 3 (Sierra Leone), respectively. It is noted that in each of these three figures, there is a jump on the reported cumulative case numbers in late October 2014. This is due to a more comprehensive assessment

of patient databases on the World Health Organization report dated 29 October 2014 [11]. Among these three countries, Liberia has the most significant gap, which may account for the result that the inflection point for Liberia is about one month earlier than the other two countries.

We follow the ideas in [5] to calibrate our model using reduced number of date points. For Guinea, the final outbreak sizes are estimated as 14577, 27149, 25789 and 3327 when the end dates of reduced date points are chosen as 08 September 2014, 10 October 2014, 14 November 2014, and 07 January 2015, respectively. For Liberia, the final outbreak sizes are estimated as 4539, 4553, 7368 and 7883 when the end dates of reduced date points are chosen as 08 September 2014, 10 October 2014, 14 November 2014, and 07 January 2015, respectively. For Sierra Leone, the final outbreak sizes are estimated as 2598, 84247, 8341 and 10260 when the end dates of reduced date points are chosen as 08 September 2014, 10 October 2014, 14 November 2014, and 07 January 2015, respectively. From the fitted graphs, we observe that our model provides stable and reasonable estimations once the outbreak has passed its inflection point; namely, the end date used for simulation is later than the estimated inflection point.

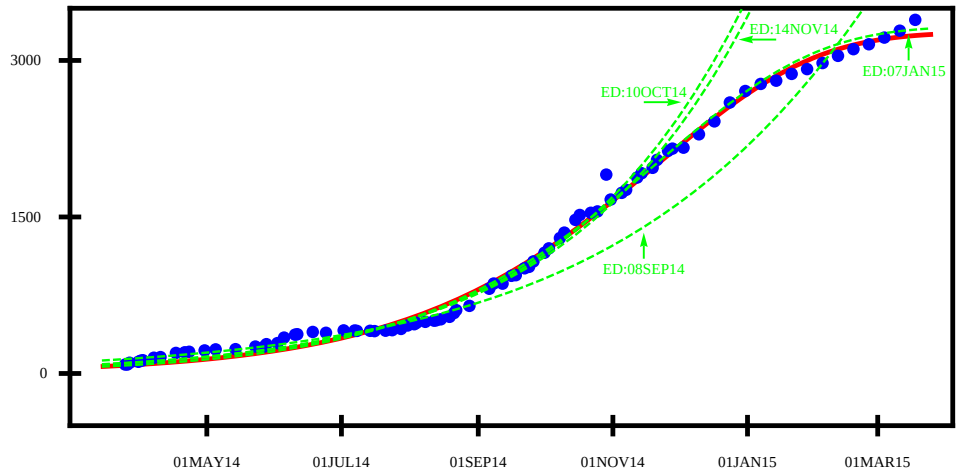


FIGURE 1. Fitted graph for the reported cumulative cases in Guinea. The dots are real data points, the solid curve is the trajectory of fitted cumulative case number using all data points until 18 March 2015, and the dashed curves are trajectories of fitted cumulative case number using reduced data points with different end dates (ED)

We also fit the final outbreak size as 3268 for Guinea, 8630 for Liberia, and 11227 for Sierra Leone, using all data points until 18 March 2015. All of these estimated values are smaller than cumulative case numbers reported on 18 March 2015. This indicates that another potential outbreak wave may be approaching [7]. A multiple epidemic outbreak can be interpreted as a trajectory of cumulative case number with multiple S-shaped segments. In other words, the cumulative case number has more than one inflection point and it goes through multiple acceleration-deceleration stages; see [7]. Clearly, the formula (5) is not suitable for multiple-wave estimation

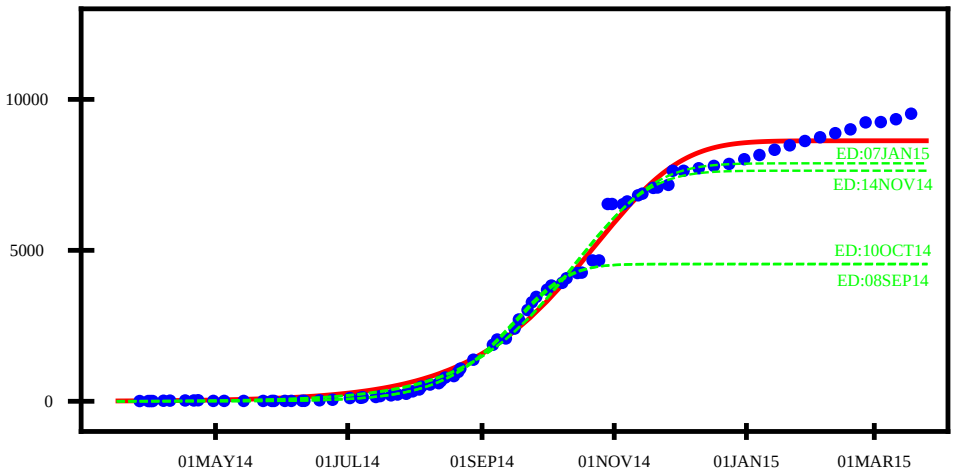


FIGURE 2. Fitted graph for the reported cumulative cases in Liberia. The dots are real data points, the solid curve is the trajectory of fitted cumulative case number using all data points until 18 March 2015, and the dashed curves are trajectories of fitted cumulative case number using reduced data points with different end dates (ED). Here, the two dashed curves corresponding to end dates 10 October 2014 and 08 September 2014 are very close to each other.

because its graph has only one S-shaped segment. However, one could use it to predict a possible second epidemic wave when the final outbreak size is continually underestimated. Following the ideas in [7], we fit our model using reported data of selected time periods with different end dates. By comparing the estimated final outbreak size with the case number at the end date of the selected time period, we propose a simple criterion for multiple-wave prediction: if, in three simulations using successive end dates, the final outbreak sizes are all underestimated; namely, they are less than the reported case numbers at the corresponding end dates, we conclude that a second epidemic outbreak is approaching.

From Table 2, we observe that the final outbreak size is underestimated if the end date is later than 04 March, 2015. This indicates that a second epidemic wave in Guinea is starting in early March of 2015. From Table 3, we observe that the final outbreak size is underestimated if the end date is chosen to be later than 04 February, 2015. This indicates that a second epidemic wave in Sierra Leone is starting in early February of 2015. We also conduct simulations using selected time periods for Liberia and find that the final outbreak size is underestimated if the end date is chosen to be later than 10 December, 2014. This indicates that the second epidemic wave in Liberia starts in December 2014. The differences in the estimated initial dates of second epidemic wave in three countries may result from a significant gap of the reported case number on 29 October, 2014.

Following the procedure as given in [7], one may also simulate the second epidemic wave using the data after the starting time of the second wave. However, due to the insufficient number of reported data, we are not able to provide accurate and satisfactory estimations of the final outbreak size or other parameters for the second

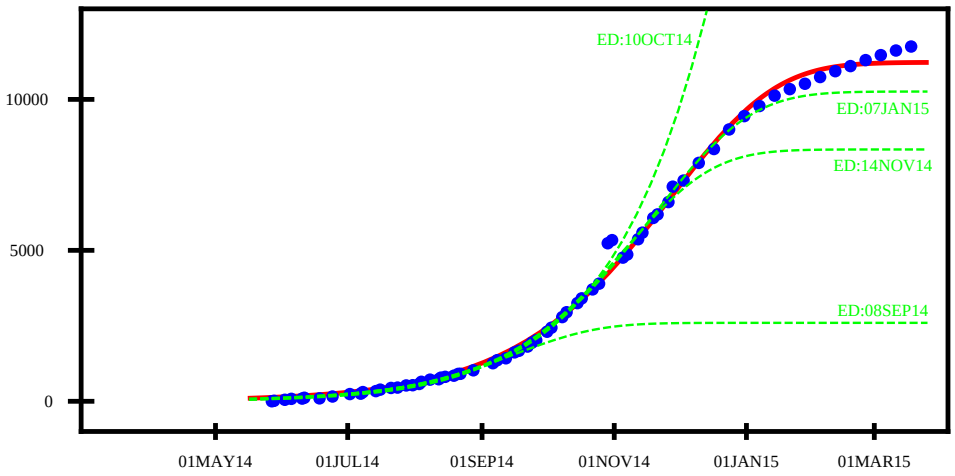


FIGURE 3. Fitted graph for the reported cumulative cases in Sierra Leone. The dots are real data points, the solid curve is the trajectory of fitted cumulative case number using all data points until 18 March 2015, and the dashed curves are trajectories of fitted cumulative case number using reduced data points with different end dates (ED).

epidemic wave. Moreover, since the outbreak is still ongoing, it is more important to prevent, rather than predict, the second epidemic wave. Based on the data until 18 March 2015, we conclude that the Ebola outbreak is starting a second epidemic wave. Some more effective disease control policy should be implemented to inhibit this possible second epidemic wave.

| End date | Cases | $K$ | $R_0$ | $t_c$ | $t_p$ | Least squares |
|---|---|---|---|---|---|---|
| 28JAN15 | 2917 | 3165 | 1.119 | 15NOV14 | 24NOV14 | 416757 |
| 04FEB15 | 2975 | 3150 | 1.120 | 14NOV14 | 23NOV14 | 417512 |
| 11FEB15 | 3044 | 3152 | 1.120 | 14NOV14 | 23NOV14 | 417541 |
| 18FEB15 | 3108 | 3165 | 1.119 | 15NOV14 | 24NOV14 | 418592 |
| 25FEB15 | 3155 | 3180 | 1.119 | 15NOV14 | 24NOV14 | 420539 |
| 04MAR15 | 3219 | 3201 | 1.118 | 16NOV14 | 25NOV14 | 425999 |
| 11MAR15 | 3285 | 3229 | 1.117 | 17NOV14 | 26NOV14 | 436834 |
| 18MAR15 | 3389 | 3268 | 1.116 | 21NOV14 | 28NOV14 | 464698 |

TABLE 2. Estimated parameter values using reported Guinea data during selected time periods with different end dates. Here, the numbers in the second column are the case numbers reported on the corresponding end dates. The least squares are the minimized values of the objective function defined as the summation of squares of estimation errors.

Now, we regularly increase the value of infectious period $1/\gamma$ from 6 to 16 days, and conduct numerical simulations. It is noted that the fitted values of basic reproduction number $R_0$ will also increase from 1.092 to 1.252 for Guinea, from 1.179

| End date | Cases | $K$ | $R_0$ | $t_c$ | $t_p$ | Least squares |
|---|---|---|---|---|---|---|
| 14JAN15 | 10124 | 10408 | 1.195 | 21NOV14 | 27NOV14 | 2082655 |
| 21JAN15 | 10340 | 10506 | 1.194 | 21NOV14 | 28NOV14 | 2110722 |
| 28JAN15 | 10518 | 10580 | 1.193 | 22NOV14 | 29NOV14 | 2136258 |
| 04FEB15 | 10740 | 10667 | 1.192 | 23NOV14 | 29NOV14 | 2189903 |
| 11FEB15 | 10934 | 10758 | 1.190 | 23NOV14 | 30NOV14 | 2275122 |
| 18FEB15 | 11103 | 10848 | 1.188 | 24NOV14 | 01DEC14 | 2392058 |
| 25FEB15 | 11301 | 10945 | 1.187 | 25NOV14 | 02DEC14 | 2575726 |
| 04MAR15 | 11466 | 11042 | 1.185 | 26NOV14 | 03DEC14 | 2811987 |
| 11MAR15 | 11619 | 11137 | 1.183 | 27NOV14 | 04DEC14 | 3100556 |
| 18MAR15 | 11751 | 11227 | 1.181 | 28NOV14 | 04DEC14 | 3428974 |

TABLE 3. Estimated parameter values using reported Sierra Leone data during selected time periods with different end dates. Here, the numbers in the second column are the case numbers reported on the corresponding end dates. The least squares are the minimized values of the objective function defined as the summation of squares of estimation errors.

to 1.503 for Liberia, and from 1.144 to 1.400 for Sierra Leone. The estimated final outbreak size stays in a range of [3261, 3307] for Guinea, [8620, 8677] for Liberia, and [11209, 11319] for Sierra Leone. On the other hand, the inflection point $t_c$ and peak time $t_p$ do not vary too much. For Guinea, $t_c$ decreases from 266 (21 November 2014) to 264 (19 November 2014). For Liberia, $t_c$ decreases from 238 (24 October 2014) to 235 (21 October 2014). For Sierra Leone, $t_c$ decreases from 273 (28 November 2014) to 270 (25 November 2014). The peak time $t_p$ increases from 272 (27 November 2014) to 279 (4 December 2014) for Guinea, from 244 (30 October 2014) to 248 (3 November 2014) for Liberia, and from 279 (4 December 2014) to 284 (9 December 2014) for Sierra Leone. We observe that the fitted inflection point $t_c$ and peak time $t_p$ are stable under perturbations on the infectious period $1/\gamma$.

4. **Discussion.** This study provides real-time estimation of basic reproduction number, final outbreak size, inflection point and peak time for the ongoing Ebola outbreak in West Africa using reported cumulative case data.

The fitted basic reproduction numbers are smaller than those estimated in [1] where only the data until 20 August 2014 was used. This indicates that the disease control policy became more effective during the late stage of the outbreak.

The estimated inflection points for Guinea and Sierra Leone are close to each other, but the one for Liberia is about one month earlier. This is due to a significant increase on reported cumulative case number dated 29 October 2014; see Figure 2 and [11].

From Table 1, we note that the estimated peak time has about one week's delay after the estimated inflection point, while the infectious period is fixed as 7.5 days. This supports the conclusion in [10] that the peak time occurs about one infectious period after the inflection point.

We use bootstrap method to obtain 95% confidence intervals for estimated parameter values. Since the infectious period is fixed and we use only three parameters to conduct numerical optimization, the confidence intervals are much tighter than those obtained by minimizing the objective function with four parameters; see [7]

for example. This demonstrates that reducing the number of parameters not only resolves over-fitting problem, but also provides tighter confidence interval for each estimated parameter.

If we vary the infectious period from 6 to 16 days, the estimated inflection point and peak time are stable in the sense that they only vary within a small interval. This demonstrates that our method has a significant accuracy in capturing the inflection point and peak time.

The values of final outbreak sizes in three countries are all underestimated, which can be considered as a warning signal of a second outbreak wave.

**Appendix: Explicit solution of (1).** Here, we provide the detail in solving the system (1). First, we add the two equations in (1) to obtain $S'(t)+I'(t)=-\gamma I(t)$. Coupling this with the first equation of (1) yields

$$\frac{d(S+I)}{dS}=\frac{\gamma(S+I)}{\beta S}.$$

This equation is separable and its solution can be written as

$$C_1(S+I)=S^{\gamma/\beta},$$

where $C_1>0$ is a constant of integration. Next, we use the above relation to eliminate $I$ in the first equation of (1). It follows that

$$S'(t)=-\beta S(t)+\beta C_1[S(t)]^{2-\gamma/\beta}.$$

This is a Bernoulli equation. We set $u(t)=[S(t)]^{\gamma/\beta-1}$ and obtain

$$u'(t)=(\beta-\gamma)[u(t)-C_1].$$

The above equation becomes separable and its solution is given by

$$u(t)=C_1[1+C_2e^{(\beta-\gamma)t}]$$

with $C_2>0$ being another constant of integration. Now, we have

$$S(t)=C_1^{\beta/(\gamma-\beta)}[1+C_2e^{(\beta-\gamma)t}]^{\beta/(\gamma-\beta)}.$$

Substituting this into the relation between $S$ and $I$ gives

$$I(t)=C_1^{\beta/(\gamma-\beta)}[1+C_2e^{(\beta-\gamma)t}]^{\gamma/(\gamma-\beta)}-C_1^{\beta/(\gamma-\beta)}[1+C_2e^{(\beta-\gamma)t}]^{\beta/(\gamma-\beta)}.$$

Finally, we set $K=C_1^{\beta/(\gamma-\beta)}$ and $t_c=[\ln(1-\gamma/\beta)-\ln C_2]/(\beta-\gamma)$ to rewrite the formulas of $S(t)$ and $I(t)$ as

$$\begin{aligned}S(t)=&K[1+(1-\gamma/\beta)e^{(\beta-\gamma)(t-t_c)}]^{\beta/(\gamma-\beta)};\\ I(t)=&K[1+(1-\gamma/\beta)e^{(\beta-\gamma)(t-t_c)}]^{\gamma/(\gamma-\beta)}-K[1+(1-\gamma/\beta)e^{(\beta-\gamma)(t-t_c)}]^{\beta/(\gamma-\beta)}.\end{aligned}$$

This is equivalent with (3) in view of (2).

**Acknowledgments.** We would like to thank the anonymous referees for their valuable comments and suggestions which lead to an important improvement of this paper.

## REFERENCES


[1] C. L. Althaus, *Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa*, PLOS Currents Outbreaks, 2014.

[2] D. G. Bausch, J. S. Towner, S. F. Dowell, et al., Assessment of the risk of Ebola virus transmission from bodily fluids and fomites, *The Journal of Infectious Diseases*, **196** (2007), S142–147.

[3] *Centers for Disease Control and Prevention*, 2014 Ebola outbreak in West Africa - reported cases graphs, Available from: http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html. (Last accessed on 18 March 2015.)

[4] *Centers for Disease Control and Prevention*, Ebola virus disease (EVD) information for clinicians in U.S. healthcare settings, Available from: http://www.cdc.gov/vhf/ebola/healthcare-us/preparing/clinicians.html. (Last accessed on 18 March 2015.)

[5] G. Chowell, L. Simonsen, C. Viboud and Y. Kuang, Is West Africa approaching a catastrophic phase or is the 2014 Ebola epidemic slowing down? Different models yield different answers for Liberia, *PLOS Currents Outbreaks*, (2014), 1–12.

[6] O. Diekmann, J. A. P. Heesterbeek and J. A. J. Metz, On the definition and the computation of the basic reproduction ratio $R_0$ in models for infectious diseases in heterogeneous populations, *Journal of Mathematical Biology*, **28** (1990), 365–382.

[7] Y. H. Hsieh and Y. S. Cheng, Real-time forecast of multiphase outbreak, *Emerging Infectious Diseases*, **12** (2006), 122–127.

[8] F. J. Richards, A flexible growth function for empirical use, *Journal of Experimental Botany*, **10** (1959), 290–301.

[9] P. van den Driessche and J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, *Mathematical Biosciences*, **180** (2002), 29–48.

[10] X.-S. Wang, J. Wu and Y. Yang, Richards model revisited: Validation by and application to infection dynamics, *Journal of Theoretical Biology*, **313** (2012), 12–19.

[11] *World Health Organization*, Ebola response roadmap situation report - 29 October 2014, Available from: http://apps.who.int/iris/bitstream/10665/137376/1/roadmapsitrep_29Oct2014_eng.pdf?ua=1. (Last accessed on 18 March 2015.)

[12] *World Health Organization*, Situation reports with epidemiological data: Archive, Available from: http://apps.who.int/ebola/en/current-situation. (Last accessed on 18 March 2015.)


Received April 03, 2015; Accepted May 13, 2015.

*E-mail address*: xswang@semo.edu
*E-mail address*: lzhong1s@semo.edu