

Original Research Paper

Estimation of the Case Fatality Ratio of Middle East Respiratory Syndrome based on Patients' Comorbidities with Application to the MERS-CoV Epidemics in the Republic of Korea, 2015

Changhyuck Oh and Seonyeong Hwang

Department of Statistics, Yeungnam University, Gyeongsan, Gyeongbuk 38541, Republic of Korea, Korea

Article history

Received: 26-09-2016

Revised: 25-10-2016

Accepted: 05-11-2016

Corresponding Author:

Changhyuck Oh
Department of Statistics,
Yeungnam University,
Gyeongsan, Gyeongbuk 38541,
Republic of Korea, Korea
Email: aristoh99@hotmail.com

Abstract: In an early stage of the course of the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) outbreak in Korea in 2015, the Centers for Disease Control and Prevention of Korea and several studies reported estimated Case Fatality Ratios (CFRs) that were significantly different. Here, we propose an estimation method based on the commonly quoted naive estimator of CFR utilizing the number of in-hospital patients with comorbidities, as well as the numbers of cumulative confirmed and dead patients up to a certain date. We compared the proposed estimator with two naive and an integral method estimator by simulation experiments under the individual based susceptible-exposed-infected-recovered model and analysis of data from the 2015 epidemic of MERS-CoV in Korea. The proposed estimator better simulated and analyzed the Korean MERS data.

Keywords: Estimation, Case Fatality Ratio, MERS-CoV, Comorbidity

Introduction

During the early stages of a pandemic of a novel influenza, an estimate of the Case Fatality Ratio (CFR) can be utilized to determine how to distribute resources to supervise, prevent and treat the disease; therefore, this information plays an important role in response. Accordingly, when a novel virus occurs and spreads, estimation of the CFR is essential. However, because the CFR is determined after the epidemic, it is unknown during the course of the epidemic and must be estimated. In this study, we considered the CFRs among lab-confirmed cases. The first confirmed case of the Middle East Respiratory Syndrome Corona Virus (MERS-CoV) in Korea was reported on May 20, 2015. Prior to entering Korea, the CFR in other regions known to be affected, such as the Middle East, was known to be about 40%. Additionally, secondary infections occurred through some hospitals and the Centers for Disease Control and Prevention of Korea (CDCPK) announced a daily estimate of CFR using the naive method, which is employed by the World Health Organization (WHO). The CFRs reported in the early stage of the epidemic were very low compared to the known CFR of MERS-CoV in Middle East countries. About one month after

the first confirmed patient on June 19, 2015, the CDCPK reported a CFR of 14.5%.

During the epidemic in Korea, several articles were published that estimated the CFR of the epidemic. Cowling *et al.* (2015) estimated the CFR as 21% using an integral method proposed by Garske *et al.* (2009) based on epidemic data of 24 deaths and 30 recovered persons among 166 confirmed cases up to June 19, 2015. Mizumoto *et al.* (2015) estimated the CFR in real time using an integral equation method and calculated estimates of 40 to 80% in the early stages of the epidemic. On the other hand, when data collected up to July 15 were used (after which there were no more new confirmed patients), Majumder *et al.* (2015) estimated the CFR to be 22% using the improved naive estimator, while CDCPK reported it as 17.7% based on the naive estimator. These wide ranging estimates in the early stages of the outbreak led to difficulties in developing preventative policies and treatments, as well as to uncertainty regarding which estimation method is more reliable.

Many studies have investigated methods for estimation of CFR of infectious diseases. The naive estimator is defined as the ratio of cumulative deaths to cases. Ma and Driessche (2008) represented the naive estimator of CFR using parameters of a compartment

model with susceptible, infectious and recovered individuals. The naive estimator is widely used. For example, Atkins *et al.* (2014) estimated the CFR of Ebola in West Africa from 2013 to 2014 by the naive estimator. However, the naive estimator appears to underestimate the actual value since it assumes all patients in the hospital will recover and hence be discharged. Conversely, the improved naive estimator, which is given as the ratio of the number of dead patients to in-hospital patients, assumes that not all patients in the hospital have been infected, resulting in overestimation in the early stages of an epidemic. Therefore, trials have been conducted to improve two naive estimators by using auxiliary information of patients other than the number of confirmed cases and the number of dead patients.

Upon estimation of the CFR of SARS in Hong Kong and worldwide in 2003, Donnelly *et al.* (2003) and Ghani *et al.* (2005) used additional information regarding the time from onset to death or discharge from the hospital. Garske *et al.* (2009) investigated an integral method of CFR estimation of a novel flu (A/H1N1) that occurred in 2009 worldwide, resulting in 100,000 known patients and 429 deaths. Their proposed estimator utilized the ratio of deaths to a modified number of confirmed patients by using an estimated cumulative distribution of the time from onset to death, such that it improves underestimation of the naive estimator.

Among these, two naive estimators and an integral method estimator using the cumulative distribution of delay times from onset to death were employed for estimation of CFR during the 2015 Korean MERS-CoV epidemic. However, the integral method estimators depend on estimates of the times from onset to death, resulting in the estimated CFR differing among studies.

Here, we propose a novel estimator of CFR that utilizes information regarding comorbidities of patients in the hospital on the day of estimation and compare it to the two naive estimators and an integral method estimator. We demonstrate that the proposed estimator is better than the naive estimators and the integral method estimator used in the 2015 Korean MERS-CoV pandemic based on simulation experiments and analysis of data from the 2015 Korean MERS-CoV.

Materials and Methods

Let s be the days from when the first MERS-CoV patient was confirmed in a country and $C(s)$, $D(s)$ and $R(s)$ be the cumulative numbers of confirmed patients, dead patients and recovered patients, respectively, up to day s . When an epidemic is completed on day T , the case fatality ratio π of a nation or a region is defined as the ratio of the cumulative number of deaths to that of confirmed patients up to T , that is, $\pi = D(T)/C(T)$. Here,

we do not count patients who have not come to the hospital for various reasons and therefore not been diagnosed as confirmed.

There are two types of naive estimators, the naive and improved naive estimator, which are as follows:

$$\hat{\pi}_1(s) = D(s) / C(s) \tag{1}$$

$$\hat{\pi}_2(s) = D(s) / \{D(s) + R(s)\} \tag{2}$$

For $s = 1, 2, \dots, T$, respectively. These estimators are also quoted as simple crude estimators as reported by Ghani *et al.* (2005) and Garske *et al.* (2009). Note that at day $s = 1, \dots, T$, the estimator $\hat{\pi}_1$ assumes that every patient in the hospital has recovered, while $\hat{\pi}_2$ assumes that all patients in the hospital are free from infection. When the epidemic is completed, the two estimators become the same, resulting in the true CFR. At each observation time during the epidemic process, the confidence intervals of the two estimators are given using normal approximation as in Ghani *et al.* (2005).

The estimator suggested by Garske *et al.* (2009) reduced the number of confirmed patients by employing an estimated distribution function of the time from onset to death to give a bigger value than the naive estimator. The estimator developed by Garske *et al.* (2009) is as follows:

$$\hat{\pi}_3(s) = \frac{D(s)}{\sum_{u=0}^s c(u)F(s-u)} \tag{3}$$

where, $c(u)$ is the number of confirmed patients who begin symptoms at day u and $F(s)$ is an estimated cumulative distribution of delay times from symptom onset to death. Therefore, $F(s-u)$ is the probability that a patient who develops symptoms on day u will die by day s . However, if the day of symptom onset for a patient is not known, it must be estimated. If it is over-estimated, the Garske estimator itself becomes over estimated by reducing the denominator. The confidence interval for $\hat{\pi}_3$ on each day can be obtained by using the bootstrap method with the estimated distribution function $F(\cdot)$ using data describing the delay times. Garske *et al.* (2009) compared his proposed estimator to the naive estimator and the improved naive estimator by simulation under the SEIR model. Cowling *et al.* (2015) used a log normal distribution and Mizumoto *et al.* (2015) a gamma distribution for $F(\cdot)$. The Garske *et al.* (2009) method is widely used in the literature. Cowling *et al.* (2015) estimated Korean MERS-CoV with Garske, while Mizumoto *et al.* (2015) used an integral method, of which Garske's method is included. Many studies have been conducted using

Garski's method (Cowling *et al.*, 2015; Mizumoto *et al.*, 2015; Mishra *et al.*, 2010; Presanis *et al.*, 2009; Balcan *et al.*, 2009; Khandaker *et al.*, 2011; Nishiura, 2010; Yu *et al.*, 2013; Nguyen-Van-Tam *et al.*, 2010; Echevarría-Zuno *et al.*, 2009; Merler *et al.*, 2011; Yang *et al.*, 2009).

When information regarding comorbidities of confirmed patients are available, we propose an estimator using this information in Equation 1 of the naive estimator. On day s , let the number of patients with and without comorbidities be $H_s(s)$ and $H_u(s)$, respectively. The proposed estimator is:

$$\hat{\pi}_4(s) = \frac{D(s) + \alpha H_u(s) + \beta H_s(s)}{D(s) + R(s) + H(s)} \quad (4)$$

The estimator $\hat{\pi}_4$ assumes $100\alpha\%$ of patients with comorbidities are dead and $100\beta\%$ of patients without it are as recovered on day s , where α and β are given constants obtained from previous pandemics of MERS, if available. Note that the naive estimator assumes $\alpha = \beta = 0$ and the improved naive estimator assumes $H(s) = 0$ in addition to $\alpha = \beta = 0$, i.e., not taking into account in-hospital patients. The confidence interval of π can be obtained using normal approximation as follows (see appendix for proof):

$$\hat{\pi}_4 \pm z_{1-\alpha/2} \sqrt{\frac{\hat{V}(\hat{\pi}_4)}{n}} \quad (5)$$

where:

$$\hat{V}(\hat{\pi}_4) = \frac{1}{n^3} \left\{ \begin{array}{l} D(s)(n - D(s)) + \alpha H_s(s)(n - H_s(s)) \\ + \beta H_u(s)(n - H_u(s)) - \alpha D(s)H_s(s) \\ - \beta D(s)H_u(s) - \alpha\beta H_s(s)H_u(s) \end{array} \right\} \quad (6)$$

Results

To compare the proposed estimator with the two naive estimators and the integral method estimator, we conducted simulation experiments and analyzed the 2015 Korean MERS-CoV epidemic data.

Simulation

An individual based Susceptible-Exposed-Infected-Removed (SEIR) model was used to simulate epidemic outbreaks. The EpiFast algorithm of Bisset *et al.* (2009) for the individual based SEIR model was implemented in C++. We assumed the population was homogeneously mixed. The number of susceptible individuals in the group was set to 500. The number of the initial infectious patients and the initial latent patients was set to 1 and 0, respectively. The case fatality ratio of the

entire group was set to 0.2 (0.1) 0.4 and the ratio of patients with comorbidities in the population was set as 0.10, 0.20 and 0.30. For the given CFR and the ratio of patients with comorbidity in the population, the CFR of patients with comorbidity was set as 0.9 (0.1) 0.5 and the CFR of patients without comorbidity was set such that the CFR of the whole group was satisfied. Removed individuals were classified as dead or recovered. The days from exposure to onset were set to 3, 4 and 5 with probabilities of 0.25, 0.5 and 0.25, respectively. Simulations were repeated until we had 1,000 selected observed processes with more than 30 infected patients at the end of the epidemic process for the pandemic. Because we do not know how high or low the CFR of patients with or without comorbidities is, we set $\alpha = 0$ and $\beta = 0.6$ for simplicity.

To compare performance of the estimators, we proposed the following measurement: Assume we generate an epidemic process R times independently. For each epidemic process simulated, we first calculated the CFR π^j , $j = 1, \dots, R$ after the epidemic was completed. Let $\hat{\pi}_k^j(i)$ be the estimate of π^j on day $i = 1, \dots, N_j$ and N_j the days for which the epidemic continued in the j -th simulation, corresponding to $\hat{\pi}_k$, $k = 1, \dots, 4$. The square root of the mean square error of the daily estimates for estimator $\hat{\pi}_k$ on the j -th repetition is defined by:

$$d_j^{(k)} = \sqrt{\sum_{i=i_0}^{N_j} \frac{(\hat{\pi}_k^j(i) - \pi^j)^2}{N_j - 1}}, \quad j = 1, \dots, R$$

where, i_0 is the day on which the number of cumulative deaths is greater than or equal to a given number. In our simulation, i_0 was set to 5. This was because, in each epidemic process, if the number of cumulative deaths is too small, the estimate of the final CFR becomes unstable. For $d_j^{(k)}$, mean $\bar{d}^{(k)}$ and standard deviation $m^{(k)}$ are calculated. The simulation results are listed in Table 1.

The values of $\bar{d}^{(4)}$ were the smallest among $\bar{d}^{(k)}$, $k = 1, 2, 3, 4$ for all cases, which means the performance of the proposed estimator $\hat{\pi}_4$ was better than that of the others. As expected, the values of $\bar{d}^{(2)}$ were largest among $\bar{d}^{(k)}$, $k = 1, 2, 3, 4$ for all cases. The values of $\bar{d}^{(3)}$ were larger or the same as $\bar{d}^{(1)}$, which was somewhat unexpected because the Garske type estimator is known to be an improvement of the simple estimator. One explanation for this is that the Kaplan Meier type estimate for the cumulative distribution function utilized in the present study affected the result. Other type estimates for the cumulative distribution function might change the result, as reported by Cowling *et al.* (2015) and

Mizumoto *et al.* (2015). The standard deviations $m^{(1)}$, $m^{(3)}$ and $m^{(4)}$ are comparable with each other, while $m^{(2)}$ was larger than the others.

Korea 2015 MERS-CoV Epidemic

We used publicly available MERS-CoV data of confirmed MERS cases in the Republic of Korea from May 20 to December 23, 2015 (KICH, 2015), as well as some newspapers published in Korea. In the data set,

patient comorbidity, date of confirmed infection and recovery, death or discharged are given. The first confirmed case of MERS was recorded on May 20, 2015 and 185 more were confirmed through August 10, 2015. Among these, 36 died (19.4%), 141 were discharged (75.3%) and nine were still in the hospital (5.4%). Among the nine in-hospital patients, seven were without comorbidities and two were with comorbidities.

Table 1. Means and standard deviations $\bar{d}^{(k)}$ and $m^{(k)}$, $k = 1,2,3,4$ for $R = 1000$ simulations. The number of susceptibles in the group was 500. The number of the initial infectious patients was 1. The CFRs were given as 0.2 (0.1) 0.4 and the comorbidity ratios in the population were set as 0.1 (0.1), 0.3. For each comorbidity ratio, the death ratio of patients with comorbidity were set as 0.9 (0.1) 0.5.

| CMB | | Death ratio | | | | | |
|------|--------------------|---------------------|------------------|---------------|---------------|---------------|---------------|
| CFR | Ratio ¹ | no CMB ² | CMB ³ | $\hat{\pi}_1$ | $\hat{\pi}_2$ | $\hat{\pi}_3$ | $\hat{\pi}_4$ |
| 0.20 | 0.10 | 0.12 | 0.90 | 0.053 (0.031) | 0.066 (0.050) | 0.057 (0.036) | 0.052 (0.033) |
| 0.20 | 0.10 | 0.13 | 0.80 | 0.054 (0.032) | 0.065 (0.050) | 0.057 (0.036) | 0.053 (0.033) |
| 0.20 | 0.10 | 0.14 | 0.70 | 0.053 (0.030) | 0.065 (0.050) | 0.057 (0.035) | 0.052 (0.032) |
| 0.20 | 0.10 | 0.16 | 0.60 | 0.055 (0.033) | 0.070 (0.055) | 0.059 (0.038) | 0.054 (0.035) |
| 0.20 | 0.10 | 0.17 | 0.50 | 0.054 (0.032) | 0.066 (0.052) | 0.058 (0.037) | 0.053 (0.034) |
| 0.20 | 0.20 | 0.02 | 0.90 | 0.046 (0.022) | 0.051 (0.034) | 0.048 (0.025) | 0.043 (0.025) |
| 0.20 | 0.20 | 0.05 | 0.80 | 0.048 (0.024) | 0.055 (0.038) | 0.050 (0.028) | 0.046 (0.027) |
| 0.20 | 0.20 | 0.07 | 0.70 | 0.049 (0.025) | 0.057 (0.041) | 0.051 (0.029) | 0.046 (0.028) |
| 0.20 | 0.20 | 0.10 | 0.60 | 0.051 (0.027) | 0.058 (0.044) | 0.053 (0.031) | 0.048 (0.030) |
| 0.20 | 0.20 | 0.12 | 0.50 | 0.052 (0.029) | 0.063 (0.047) | 0.055 (0.034) | 0.050 (0.032) |
| 0.20 | 0.30 | 0.03 | 0.60 | 0.044 (0.020) | 0.049 (0.030) | 0.046 (0.022) | 0.041 (0.022) |
| 0.20 | 0.30 | 0.07 | 0.50 | 0.049 (0.026) | 0.058 (0.042) | 0.052 (0.029) | 0.047 (0.030) |
| 0.30 | 0.10 | 0.23 | 0.90 | 0.069 (0.031) | 0.080 (0.051) | 0.072 (0.035) | 0.067 (0.032) |
| 0.30 | 0.10 | 0.24 | 0.80 | 0.069 (0.028) | 0.078 (0.048) | 0.071 (0.032) | 0.066 (0.029) |
| 0.30 | 0.10 | 0.26 | 0.70 | 0.068 (0.029) | 0.080 (0.051) | 0.071 (0.034) | 0.066 (0.030) |
| 0.30 | 0.10 | 0.27 | 0.60 | 0.069 (0.031) | 0.082 (0.054) | 0.073 (0.036) | 0.067 (0.033) |
| 0.30 | 0.10 | 0.28 | 0.50 | 0.071 (0.031) | 0.086 (0.053) | 0.075 (0.036) | 0.070 (0.032) |
| 0.30 | 0.20 | 0.15 | 0.90 | 0.066 (0.027) | 0.071 (0.043) | 0.067 (0.030) | 0.061 (0.029) |
| 0.30 | 0.20 | 0.18 | 0.80 | 0.067 (0.028) | 0.075 (0.046) | 0.069 (0.032) | 0.063 (0.030) |
| 0.30 | 0.20 | 0.20 | 0.70 | 0.066 (0.030) | 0.076 (0.048) | 0.069 (0.033) | 0.062 (0.031) |
| 0.30 | 0.20 | 0.23 | 0.60 | 0.069 (0.030) | 0.078 (0.049) | 0.071 (0.034) | 0.065 (0.032) |
| 0.30 | 0.20 | 0.25 | 0.50 | 0.069 (0.031) | 0.081 (0.052) | 0.072 (0.035) | 0.066 (0.033) |
| 0.30 | 0.30 | 0.04 | 0.90 | 0.062 (0.023) | 0.061 (0.031) | 0.062 (0.024) | 0.054 (0.024) |
| 0.30 | 0.30 | 0.09 | 0.80 | 0.063 (0.025) | 0.066 (0.037) | 0.064 (0.027) | 0.056 (0.027) |
| 0.30 | 0.30 | 0.13 | 0.70 | 0.064 (0.025) | 0.069 (0.042) | 0.065 (0.029) | 0.057 (0.028) |
| 0.30 | 0.30 | 0.17 | 0.60 | 0.066 (0.027) | 0.072 (0.045) | 0.068 (0.031) | 0.060 (0.030) |
| 0.30 | 0.30 | 0.21 | 0.50 | 0.069 (0.031) | 0.081 (0.053) | 0.073 (0.036) | 0.065 (0.035) |
| 0.40 | 0.10 | 0.34 | 0.90 | 0.083 (0.029) | 0.092 (0.050) | 0.085 (0.032) | 0.080 (0.030) |
| 0.40 | 0.10 | 0.36 | 0.80 | 0.085 (0.029) | 0.093 (0.048) | 0.086 (0.031) | 0.082 (0.029) |
| 0.40 | 0.10 | 0.37 | 0.70 | 0.085 (0.030) | 0.096 (0.050) | 0.087 (0.033) | 0.082 (0.031) |
| 0.40 | 0.10 | 0.38 | 0.60 | 0.086 (0.030) | 0.096 (0.049) | 0.088 (0.032) | 0.083 (0.030) |
| 0.40 | 0.10 | 0.39 | 0.50 | 0.082 (0.029) | 0.091 (0.047) | 0.084 (0.032) | 0.080 (0.029) |
| 0.40 | 0.20 | 0.27 | 0.90 | 0.081 (0.028) | 0.084 (0.044) | 0.081 (0.030) | 0.074 (0.029) |
| 0.40 | 0.20 | 0.30 | 0.80 | 0.082 (0.029) | 0.087 (0.046) | 0.083 (0.032) | 0.076 (0.030) |
| 0.40 | 0.20 | 0.33 | 0.70 | 0.084 (0.029) | 0.090 (0.047) | 0.085 (0.031) | 0.079 (0.030) |
| 0.40 | 0.20 | 0.35 | 0.60 | 0.084 (0.028) | 0.091 (0.047) | 0.085 (0.030) | 0.078 (0.029) |
| 0.40 | 0.20 | 0.37 | 0.50 | 0.085 (0.029) | 0.093 (0.051) | 0.086 (0.033) | 0.079 (0.031) |
| 0.40 | 0.30 | 0.19 | 0.90 | 0.080 (0.028) | 0.078 (0.042) | 0.079 (0.030) | 0.070 (0.030) |
| 0.40 | 0.30 | 0.23 | 0.80 | 0.080 (0.028) | 0.081 (0.042) | 0.080 (0.030) | 0.071 (0.029) |
| 0.40 | 0.30 | 0.27 | 0.70 | 0.083 (0.030) | 0.087 (0.046) | 0.084 (0.032) | 0.075 (0.032) |
| 0.40 | 0.30 | 0.31 | 0.60 | 0.081 (0.028) | 0.086 (0.045) | 0.082 (0.031) | 0.073 (0.030) |
| 0.40 | 0.30 | 0.36 | 0.50 | 0.083 (0.029) | 0.091 (0.048) | 0.084 (0.032) | 0.076 (0.031) |

¹Comorbidity patient ratio in the population; ²Death ratio among patients without comorbidity; ³Death ratio among patients with comorbidity

Table 2. Number of patients, deaths and discharges by comorbidities. Ratios are in parentheses

| Comorbidity | Result | | No. of patients |
|-------------|---------------|----------------|-----------------|
| | Death | Discharge | |
| Without | 11 (0.071) | 143 (0.929) | 154 (1.000) |
| With | 27 (0.844) | 5 (0.156) | 32 (1.000) |
| Total | 38 | 148 | 186 |

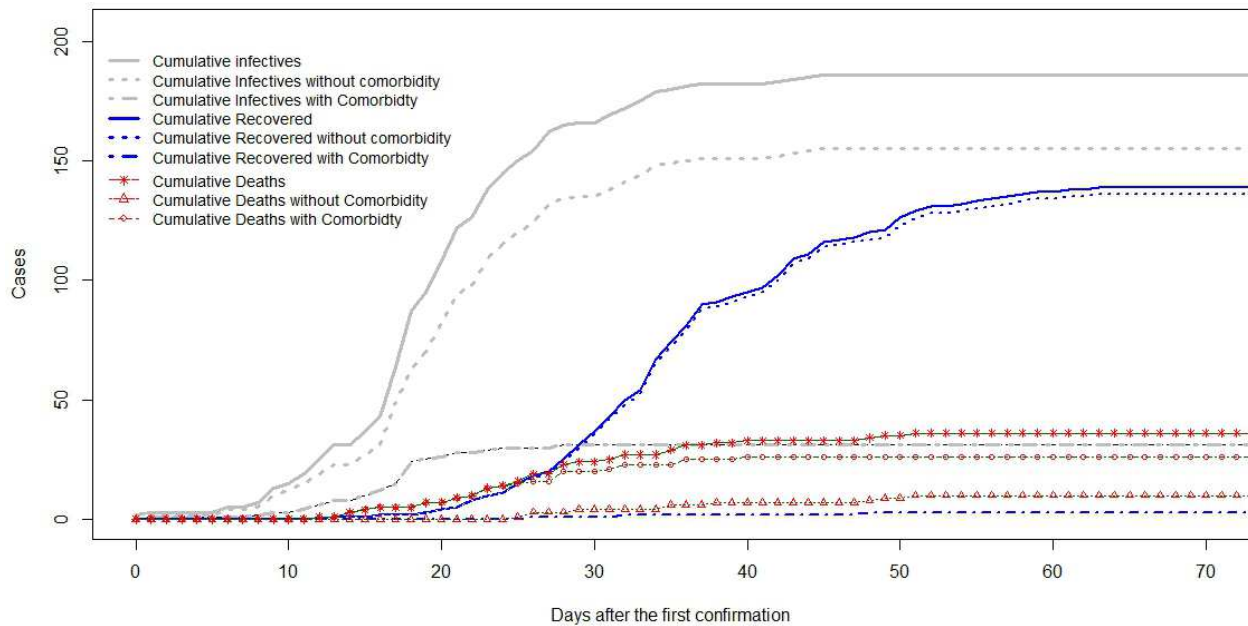


Fig. 1. Plot of daily cumulative events

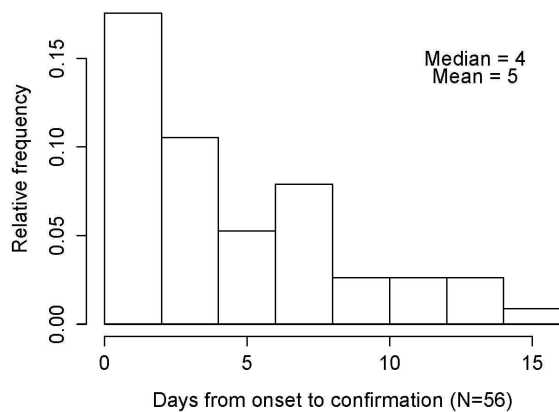


Fig. 2. Histogram of days from onset to confirmation

After July 5, 2015, the number of confirmed patients (186) did not change and the number of deaths (36) did not change between July 11 and October 25, 2015, at which time the number of deaths increased to 37. The first death was recorded on June 1, 2015; therefore, CFR estimates were obtained after that day. In the data reported, patients

with number 25, 36 and 64 were discharged before or on the confirmed day. For the three cases, we use the confirmed day minus one day as a modification. Figure 1 is a plot showing the cumulative number of patients or infectious patients, as well as the cumulative number of patients discharged from hospitals or recovered and cumulative death by comorbidity condition.

Table 2 shows the number of deaths and discharges classified by comorbidities. After the epidemic was completed, among 186 confirmed patients, 38 died, giving a CFR of 20.4%. Additionally, 154 patients had no comorbidities, giving a death ratio of 7.14%. The number of patients with comorbidities was 32, giving a death ratio of 84.4%.

Only 56 of the 186 records had information regarding the time from onset to confirmation. From these 56 records, we calculated the median as 4 days and the mean as 5 days. Therefore, we used the median 4 days for death records that did not specify the time from onset to confirmation (Fig. 2).

Figure 3 shows the estimates generated using $\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3$ and $\hat{\pi}_4$.

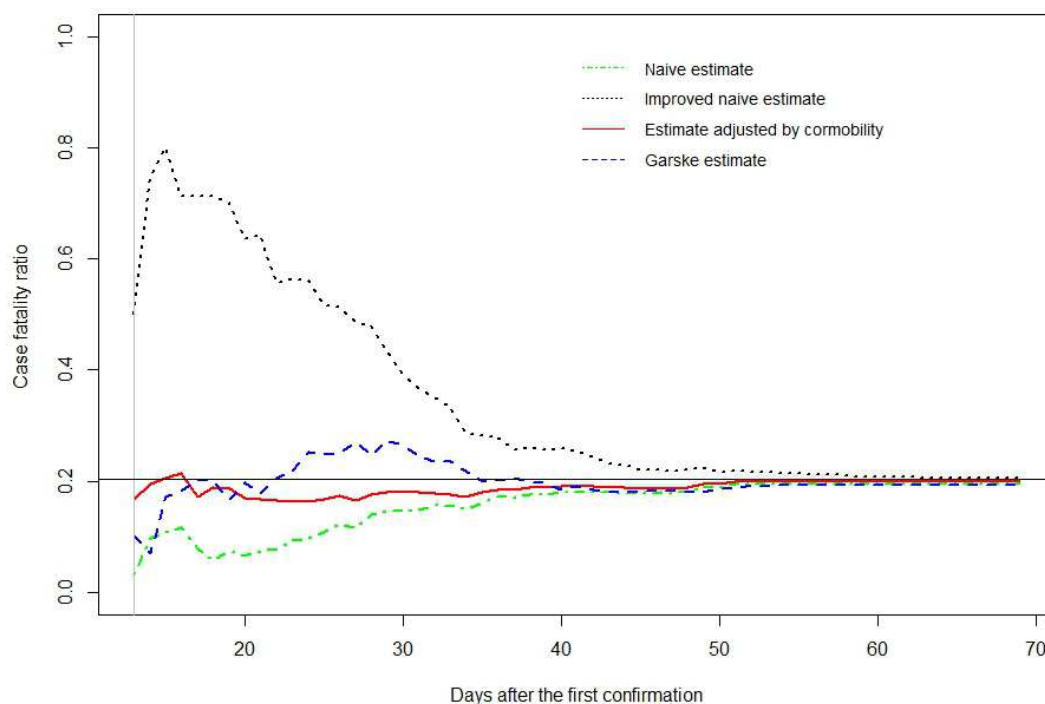


Fig. 3. Daily estimates of MERS-CoV in Korea, 2015

Daily estimates for the CFR using each estimator became stable after July 11, 2015, after which there were no changes in the number of confirmed patients and dead patients. The estimates generated using $\hat{\pi}_1$ were underestimated in the early stage of epidemics and became stable after about 40 days after the first confirmed patient. The estimates generated by $\hat{\pi}_2$ were greatly overestimated in the early stage of the epidemics. Estimator $\hat{\pi}_3$, proposed by Garske showed large fluctuations compared to the proposed estimator in the early stage. Estimator $\hat{\pi}_4$ showed a very stable pattern in the early stages of the epidemic relative to other estimators and overall good performance.

Discussion

In this study, we proposed a new estimator for the CFR of MERS-CoV that considers the comorbidities of MERS-CoV patients. The estimator uses the same denominator of the naive estimator, but a different numerator in conjunction with the number of dead patients and patients with comorbidities. The developed estimator improves the underestimation problem of the naive estimator and the large fluctuation of the integral method in the early stage of the 2015 Korea-MERS-CoV epidemic. Since doctors check comorbidities of patients upon diagnosis, it is assumed that this information will be available under normal conditions. The mean square errors of daily estimates of estimators

considered were calculated as a measure of comparison of estimators used in simulation experiments. To the best of our knowledge, this is the first study to employ such a method of comparison.

The data describing Korea MERS-CoV only included 56 cases for which the onset times were recorded; therefore, onset times for other cases must be estimated for Garske's estimator. This estimated time influences Garske's CFR estimate. For Garske's estimation, we used the Kaplan-Meier type estimator for the cumulative distribution function for censored data. However, Cowling *et al.* (2015) and Mizumoto *et al.* (2015) used a parametric estimation method for the cumulative distribution function. In a later study, comparison of estimators depending on the type used for the cumulative distribution function might be of interest.

For patients with comorbidities, it is not clear whether some symptoms were from MERS-CoV or other diseases and therefore it is not possible to know the day of onset from MERS-CoV.

The proposed estimator generated using information describing comorbidities of patients showed that it improved underestimation of the naive estimator and that it was more stable than Garske's estimator in the early stage of infection. Although we proposed this method of estimation of CFR for Korean MERS-CoV, it can be evaluated for application to other infectious diseases using the proper CFR of patients with comorbidities. Moreover, even though we used auxiliary variables describing comorbidities, other information such as

severity of condition might be applied to improve existing estimators.

In the simulation, we used $\alpha = 0$ and $\beta = 0.6$ for the proposed estimator, which assumes a CFR for patients without comorbidities of 0 and that of patients with comorbidities of 0.6. Even though these values are quite rough estimates for α and β , we obtained satisfactory results. However, we may use more accurate estimates for α and β if such information is available from previous pandemics.

In this study, we investigated various methods of estimating the CFR of the MERS-CoV among confirmed or reported cases. We did not count patients who did not come to the hospital for confirmation. In a future study, we can evaluate methods to adjust the proposed estimator for situations in which the CFR for all MERS-CoV patients, reported or not, is of interest.

Conclusion

The proposed estimator for the CFR of MERS in Korea 2015 using information of patients' comorbidity conditions performed better than existing widely used estimators, two naïve and one integral method estimators.

Appendix. Confidence Interval of $\hat{\pi}_4$

Binomial confidence intervals for the underlying probability of death can be calculated from either exact methods or a normal approximation, as appropriate.

We first assume that given $C(s)$, $(R(s), D(s), H_s(s), H_u(s))$ has a multinomial distribution, that is:

$$(R(s), D(s), H_s(s), H_u(s)) \sim MN(C(s), p_1, p_2, p_3, p_4))$$

where:

$$1 = p_1 + p_2 + p_3 + p_4$$

and

$$p_1 \geq 0, p_2 \geq 0, p_3 \geq 0, p_4 \geq 0 \text{ and } C(s) = R(s) + D(s) + H_s(s) + H_u(s)$$

By the central limit theorem of Lehman (1999), we have, for given:

$$\begin{pmatrix} \sqrt{n} \left(\frac{R(s)}{n} - p_1 \right), \sqrt{n} \left(\frac{D(s)}{n} - p_2 \right), \\ \sqrt{n} \left(\frac{H_s(s)}{n} - p_3 \right), \sqrt{n} \left(\frac{H_u(s)}{n} - p_4 \right) \end{pmatrix} \xrightarrow{L} N(0, \Sigma)$$

where:

$$\Sigma = (\sigma_{ij})_{4 \times 4}$$

with:

$$\sigma_{ii} = C(s) p_i (1 - p_i), \quad i = 1, \dots, 4$$

and:

$$\sigma_{ij} = -C(s) p_i p_j, \quad i \neq j$$

Let:

$$\pi = p_2 + \alpha p_3 + \beta p_4$$

Then, the variance of the estimator $\hat{\pi}_4$ in (4) given $C(s) = n$ is:

$$V \left(\frac{1}{C(s)} (D(s) + \alpha H_s(s) + \beta H_u(s)) \middle| C(s) = n \right) = \frac{1}{n} \left\{ p_2(1 - p_2) + \alpha^2 p_3(1 - p_3) + \beta^2 p_4(1 - p_4) - 2\alpha p_2 p_3 - 2\beta p_2 p_4 - 2\alpha\beta p_3 p_4 \right\}$$

Therefore, we can estimate the variance as:

$$\hat{V} \left\{ \frac{D(s) + \alpha H_s(s) + \beta H_u(s)}{C(s)} \middle| C(s) = n \right\} = \frac{1}{n^3} \left\{ \begin{aligned} &D(s)(n - D(s)) + \alpha H_s(s)(n - H_s(s)) \\ &+ \beta H_u(s)(n - H_u(s)) - \alpha D(s) H_s(s) \\ &- \beta D(s) H_u(s) - \alpha\beta H_s(s) H_u(s) \end{aligned} \right\}$$

Let:

$$g(D, H_s, H_u) = D(s) + \alpha H_s(s) + \beta H_u(s) \stackrel{\text{let}}{=} x_1 + \alpha x_2 + \beta x_3$$

Then we have:

$$D = \left[\frac{\partial g(x)}{\partial x_j} \right]_{x=\mu} = (1, \alpha, \beta)$$

Thus, the variance of $g(D, H_s, H_u) = D(s) + \alpha H_s(s) + \beta H_u(s)$ is given by:

$$D \Sigma D' = \sigma_{11} + \alpha^2 \sigma_{22} + \beta^2 \sigma_{33} + 2\alpha \sigma_{12} + 2\beta \sigma_{13} + 2\alpha\beta \sigma_{23}$$

From Serfling (1980) and Lehman (1999):

$$\sqrt{n} [g(D(s), H_s(s), H_u(s)) - g(p_2, p_3, p_4)] \xrightarrow{L} N(0, D \Sigma D')$$

A confidence interval of π with $\hat{\pi}_4$ can be given by:

$$\hat{\pi}_4 \pm z_{1-\alpha/2} \sqrt{\frac{\hat{V}(\hat{\pi}_4)}{n}}$$

Author's Contribution

The authors have contributed equally to the research and writing the manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and that there are no ethical issues associated with this study.

References

- Atkins, K.E., N.S. Wenzel, M. Ndeffo-Mbah, F.L. Altice and J.P. Townsend *et al.*, 2014. Under-reporting and case fatality estimates for emerging epidemics. *BMJ*, 350: h1115-h1115. DOI: 10.1136/bmj.h1115
- Balcan, D., H. Hu, B. Goncalves, P. Bajardi and C. Poletto *et al.*, 2009. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): A Monte Carlo likelihood analysis based on human mobility. *BMC Med.*, 7: 1-12. DOI: 10.1186/1741-7015-7-45
- Bisset, K., J. Chen, X. Feng, V.S.A. Kumar and M. Marathe, 2009. Epifast: A fast algorithm for largescale realistic epidemic simulations on distributed memory systems. *Proceedings of the 23rd International Conference on Supercomputing*, pp: 430-439.
- Cowling, B.J., M. Park, V.J. Fang, P. Wu and G.M. Leung *et al.*, 2015. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, *Euro Surveill*, 20: 21163.
- Donnelly, C.A., A.C. Ghani, G.M. Leung, A.J. Hedley and C. Fraser *et al.*, 2003. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet*, 361: 1761-1766. DOI: 10.1016/S0140-6736(03)13410-1
- Echevarría-Zuno, S., J.M. Mejía-Arangur, A.J. Mar-Obeso, C. Grajales-Muñiz and E. Robles-Pérez *et al.*, 2009. Infection and death from influenza A H1N1 virus in Mexico: A retrospective analysis. *Lancet*, 374: 2072-2079. DOI: 10.1016/S0140-6736(09)61638-X
- Garske, T., J. Legrand, C.A. Donnelly, H. Ward and S. Cauchemez *et al.*, 2009. Assessing the severity of the novel influenza A/H1N1 pandemic. *BMJ*, 339: b2840-b2840. DOI: 10.1136/bmj.b2840
- Ghani, A.C., C.A. Donnelly, D.R. Cox, J.T. Griffin and C. Fraser *et al.*, 2005. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. *Am. J. Epidemiol.*, 162: 479-486. DOI: 10.1093/aje/kwi230
- Khandaker, G., A. Dierig, H. Rashid, C. King and L. Heron *et al.*, 2011. Systematic review of clinical and epidemiological features of the pandemic influenza A (H1N1) 2009. *Influenza Other Respiratory Viruses*, 5: 148-156. DOI: 10.1111/j.1750-2659.2011.00199.x
- KICH, 2015. Mers portal. Korea infection control headquarter. <http://www.mers.go.kr.lastchecked:11.14.2015>
- Lehman, E.L., 1999. *Elements of Large-Sample Theory*. Springer-Verlag, New York, ISBN-10: 0-387-98595-6, pp: 313.
- Ma, J. and P.V.D. Driessche, 2008. Case fatality proportion. *Bull. Math. Biol.*, 70: 118-133. DOI: 10.1007/s11538-007-9243-8
- Majumder, M.S., S.A. Klueberg, S.R. Mekaru and J.S. Brownstein, 2015. Mortality risk factors for Middle East Respiratory syndrome outbreak, South Korea, 2015. *Emerg. Infect. Dis.*, 21: 2088-2090. DOI: 10.3201/eid2111.151231
- Merler, S., M. Ajelli, A. Pugliese and N.M. Ferguson, 2011. Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: Implications for realtime modelling. *PLoS Comput. Biol.*, 7: e1002205-e1002205. DOI: 10.1371/journal.pcbi.1002205
- Mishra, A.C., M.S. Chadha, M.L. Choudhary and V.A. Potdar, 2010. Pandemic influenza (H1N1)2009 is associated with severe disease in India. *PLoS ONE*, 5: e10540-e10540. DOI: 10.1371/journal.pone.0010540
- Mizumoto, K., M. Saitoh, G. Chowell, Y. Miyamatsu and H. Nishiura, 2015. Estimating the risk of Middle East Respiratory Syndrome (MERS) death during the course of the outbreak in the Republic of Korea, 2015. *Int. J. Infect. Dis.*, 39: 7-9. DOI: 10.1016/j.ijid.2015.08.005
- Nguyen-Van-Tam, J.S., P.J.M. Openshaw, A. Hashim, E.M. Gadd and W.S. Lim *et al.*, 2010. Risk factors for hospitalisation and poor outcome with pandemic A/H1N1 influenza: United Kingdom first wave. *Thorax*, 65: 645-651. DOI: 10.1136/thx.2010.135210
- Nishiura, H., 2010. Case fatality ratio of pandemic influenza. *Lancet Infect. Dis.*, 10: 443-444. DOI: 10.1016/S1473-3099(10)70120-1
- Presanis, A.M., D.D. Angelis, A. Hagy, C. Reed and S. Riley *et al.*, 2009. The severity of pandemic H1N1 influenza in the United States, from April to July 2009: A Bayesian analysis. *PLoS Med.*, 6: e1000207-e1000207. DOI: 10.1371/journal.pmed.1000207

- Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. John Wiley and Sons, Inc., ISBN-10: 0-471-02403-1, pp: 122.
- Yang, Y., J.D. Sugimoto, M.E. Halloran, N.E. Basta and D.L. Chao *et al.*, 2009. The transmissibility and control of pandemic influenza A (H1N1) virus. *Science*, 326: 729-733.
DOI: 10.1126/science.1177373
- Yu, H., B. Cowling, L. Feng, E.H.Y. Lau and Q. Liao *et al.*, 2013. Human infection with avian influenza A H7N9 virus: An assessment of clinical severity. *Lancet*, 382: 138-45.
DOI: 10.1016/S0140-6736(13)61207-6