Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Real-Time Estimation of the Case Fatality Ratio and Risk Factors of Death

Hiroshi Nishiura

Graduate School of Medicine, Hokkaido University, Sapporo, Japan

Corresponding author: e-mail: nishiurah@med.hokudai.ac.jp

ABSTRACT
During the course of an epidemic, estimating the risk of death and identifying risk factors of death are of utmost importance for public health assessment of the severity of infection. The real-time estimation involves a number of important statistical problems to consider, and this chapter comprehensively describes commonly used estimation methods and their pitfalls. When estimating the case fatality risk (CFR) during the course of an epidemic, the data are right-censored because of the time delay from illness onset to death. A conventional survival analysis technique is employed for addressing right censoring. Identification of risk factors of death also requires the care for censored nature of the data, and we have devised a method that combines the survival analysis and logistic regression. Ascertainment bias is always a practical issue in interpreting the absolute value of the CFR or comparing CFR between different groups, and recent studies have shown that observational effort including seroepidemiological survey has to be made to overcome this bias.

Keywords: Case fatality risk, Statistical estimation, Death, Censored data, Ascertainment

1 INTRODUCTION
As an emerging infectious disease appears in human society, one of our early epidemiological tasks is to estimate the risk of death caused by infection, which may be measured by the fraction of deaths among infected individuals. The virulence, i.e., the severity of infection, is assessed at a population level by estimating the case fatality risk (CFR), which is also referred to as the case fatality ratio or the case fatality rate abbreviated as the same, but we use the case fatality risk throughout this chapter due to its epidemiological nature to
represent the risk (Kelly and Cowling, 2013). CFR is defined as the risk of deaths from an infectious disease among the total number of cases (Ma and van den Driessche, 2008). The virulence could be partly assessed by experimentally exploring specific molecular markers of the pathogen that should be known to be associated with severe complications, but available scientific evidence seldom rests on human infection, and moreover, the absence of such information during the early course of an epidemic always imposes substantial uncertainty.

Estimating the CFR in real time is thus of utmost importance for public health so as to modelers. Nevertheless, the estimation involves a number of important statistical problems to consider, and experts of mathematical and statistical modeling are expected to understand surrounding issues well. To enhance our technical understanding of the estimation of the CFR as a measurement of the virulence in real time, it is useful to review its practical and technical aspects. This chapter aims to describe the practical utility and the definition and concept of the epidemiological assessment of the CFR during the course of an epidemic. Common pitfalls are thoroughly covered so that readers can avoid any associated errors of estimation in the future.

2 RIGHT CENSORING: CORE ISSUE OF REAL-TIME ESTIMATION

Since the real-time estimation handles the datasets that are available by the most up-to-date time of observation, one would have to adjust possible underestimation due to the time delay from illness onset to death (Nishiura, 2010c). In the following, we use three different statistical measurements of the CFR, namely, (i) $b_t$, which is a crude, biased estimate of the CFR calculated at time $t$; (ii) $\pi$, which is an unbiased cCFR (confirmed case fatality risk) to be estimated; and (iii) $p_t$, a random variable, which yields an estimator of $\pi$ and is regarded as the realized value in one particular outbreak. First, $b_t$, a biased estimate of the CFR, calculated at time $t$, is given by the division of the cumulative number of deaths $D_t$ by the cumulative number of confirmed cases $C_t$:

$$b_t = \frac{D_t}{C_t}. \tag{1}$$

During the outbreak of severe acute respiratory syndrome (SARS) in 2002–03, it was shown that this estimator, $b_t$, yielded considerably underestimated value of the CFR (Ghani et al., 2005), and moreover, the time-dependent increase in the calculated result from (1) has erroneously led public health experts to suspect possible time-dependent increase in the virulence. This can be demonstrated by disentangling the data-generating mechanisms of $C_t$ and $D_t$ using the incidence $c_t$ (i.e., the transient number of new illness onset on day $t$), and the conditional probability density function $f_s$ of the time $s$
from onset to death, given death. Counting the time $t$ from the beginning of an epidemic, $C_t$ is the cumulative number of cases up to time $t$:

$$C_t = \sum_{i=0}^{t} c_i.$$  \hfill (2)

$D_t$ represents the cumulative number of deaths up to time $t$:

$$D_t = p_t \sum_{i=0}^{t} \sum_{j=0}^{\infty} c_{i-j} f_j,$$  \hfill (3)

where $p_t$ is the realized proportion of cases to die from the infection and is a random variable. $b_t$ can be rewritten as

$$b_t = p_t \frac{\sum_{i=0}^{t} \sum_{j=0}^{\infty} c_{i-j} f_j}{\sum_{i=0}^{t} c_i}.$$  \hfill (4)

As can be observed in Eq. (4), the estimator $b_t$ would frequently be smaller than $p_t$, because of the time delay from onset to death, leading the numerator being smaller than the denominator. Nevertheless, due to stochasticity, actual $b_t$ can also be greater than $p_t$ especially during the very early stage of the epidemic. Supposing that we observe the entire course of the epidemic (i.e., $t \to \infty$), $b_t$ tends to $p_t$ and becomes an unbiased estimator. A public health challenge is how we obtain unbiased estimator of the CFR rather than relying on frequently underestimating $b_t$.

An adjustment of the estimator $b_t$ is achieved by rearranging Eq. (4):

$$p_t = \frac{\sum_{i=0}^{t} c_i}{\sum_{i=0}^{t} \sum_{j=0}^{\infty} c_{i-j} f_j}.$$  \hfill (5)

Here $p_t$ is used as the unbiased estimator of $\pi$, which is computed by three pieces of information: the cumulative number of deaths $D_t$, the incidence $c_t$, and the distribution of the time from onset to death $f_s$. The former two are observed during the course of an epidemic. When there are a few deaths or none at all, an assumption has to be made for $f_s$, e.g., from literature based on previous outbreak. The multiplicative factor in Eq. (4) may be referred to as the factor of underestimation, $u_t$, defined by

$$u_t = \frac{\sum_{i=0}^{t} \sum_{j=0}^{\infty} c_{i-j} f_j}{\sum_{i=0}^{t} c_i}.$$  \hfill (6)

The estimator $p_t$ can be written as $p_t = b_t/u_t$. Let $C_t$ be the cumulative number of cases among which a proportion $u_t$ has been at risk for dying by time $t$, while the remaining proportion $1 - u_t$ is still unobserved. That is, in the
observed empirical data, among the $u_iC_i$ cases that have been at risk, $D_i$ have died. This is a sample from a binomial distribution with sample size $u_iC_i$ and probability $\pi$:

$$\Pr(X = D_i) = \left(\frac{u_iC_i}{D_i}\right) \pi^{D_i}(1 - \pi)^{u_iC_i - D_i}. \quad (7)$$

This method was adopted for the estimation of the CFR during epidemics of SARS (Ghani et al., 2005), influenza A (H1N1-2009) (Garske et al., 2009; Nishiura et al., 2009), and Middle East respiratory syndrome (MERS) (Cowling et al., 2015; Mizumoto et al., 2015a).

3 RIGHT CENSORING AND IDENTIFICATION OF DEATH RISK FACTORS

Despite the availability of useful methods to estimate the CFR in real time, it is crucial to identify risk factors associated with death, as the risk of death may vary significantly with age, occupation, and underlying comorbidities. During the epidemic of SARS, confirmed or probable cases aged 60 years or older were found to be at greater risk of death than younger cases by conducting a real-time analysis (Donnelly et al., 2003). Moreover, infections with SARS-associated coronavirus with comorbidities were found to be 1.7 times more likely to die than those without comorbidities. While the methods for identifying the risk of death in real time during the SARS epidemic have relied on survival analysis techniques including nonparametric Kaplan–Meier-like methods (Donnelly et al., 2003; Jewell et al., 2007), there was a need to develop a simple yet tractable method which is inspired on the adjustment of censoring for CFR with particular application to small outbreak sizes such as the MERS outbreak in the Republic of Korea in which the cumulative number of confirmed cases was as small as 181.

We developed an estimation model composed of a mixture of a survival model and logistic regression model (Mizumoto et al., 2015b). Let $S(\tau)$ and $p$ be the survival probability at time $\tau$ since illness onset and the CFR, respectively. The relationship of the two is described as

$$p = 1 - S(\infty), \quad (8)$$

where $S(\tau)$ is given by

$$S(\tau; p) = 1 - p \int_0^\tau f(x)dx, \quad (9)$$

where $f(\tau)$ is the conditional probability density function of the time from illness onset to death, given fatal outcome. In the majority of the following analyses, $f(\tau)$ was assumed as known and based on the data from first 10 reported cases who eventually resulted in death in South Korea, using the moment-based estimates of the mean (13.2 days) and standard deviation.
(7.1 days). Since we aim to identify risk factors (or explanatory variables) of \( p \), the notation \( p \) is set to be changeable by individual \( i \), i.e., \( p_i \). Such individual variation is modeled using the logit model:

\[
\ln \left( \frac{p_i}{1-p_i} \right) = a_0 + \sum_{k=1}^{N} a_k x_{k,i},
\]

where \( a_0 \) is the intercept, \( a_k \) the coefficient of variable \( k \), \( x_{k,i} \) the \( k \)th variable of individual \( i \) of the linear predictor, and \( N \) the total number of independent variables. Let \( A \) and \( B \) represent the groups of cases who have survived and died by the most recent calendar time \( t_m \), the likelihood function to parameterize the linear predictor in (10) is

\[
L(a; \alpha, \beta, t_m) = \prod_{i \in A} S(t_m - \alpha_i; p_i) \prod_{i \in B} [p_i f(\beta_i - \alpha_i)],
\]

where \( \alpha_i \) and \( \beta_i \) represent the observed dates of illness onset and death of an individual \( i \), respectively, with coefficient vector \( a = (a_0, a_1, \ldots, a_N) \).

Using the abovementioned model, we have successfully characterized the heterogeneous risk of death associated with MERS in real time in the context of the small-scale MERS outbreak in the Republic of Korea from May to July 2015 (Mizumoto et al., 2015b). Although not covered in this chapter, our method relying on the logistic model with linear predictor enabled us not only to identify epidemiological determinants of MERS death but also to examine the time-dependent variation in CFR due to an increased case ascertainment rate influenced by extensive contact tracing efforts. We have shown that cases among senior persons under treatment were at particularly high risk of MERS death. That is, marked similarities between MERS and SARS were identified, as the CFR for MERS among patients aged 60 years or older under treatment was estimated as high as 48.2% (95% CI: 35.2, 61.3), while the CFR for younger healthy individuals was less than 15%. Compared to earlier models, the proposed model was shown to be advantageous in handling small sample data. By imposing an assumption that the heterogeneous risk of death is captured by the linear predictor of the logit model in (10), our approach was able to successfully identify risk factors from very small samples (i.e., 185 cases and 36 deaths) for the MERS outbreak in the Republic of Korea, and that identification can be put into practice in real time.

4 EXTENSIONS AND FUTURE CHALLENGES

Two other important issues need to be discussed. First, comparison of multiple CFR estimates is always more informative than subjectively judging the absolute value of single CFR estimate (Nishiura et al., 2010). The CFR estimate for the entire population is regarded as a summary measure of virulence, so the reduced order of virulence (approximately by a factor of 10) provides
not only more accurate information but a more useful measure for assessing the impact of this pandemic. Nevertheless, knowledge based on additional information would be even more useful. Ideally, we would like to compare epidemiologically the virulence between different settings (e.g., comparison between age groups or countries) and to use this knowledge in the development of relevant public health policy. Understanding the heterogeneous risks of death by age and risk groups is critical in the effective design of intervention strategies. In other words, apart from discussing the changing assessment of the CFR as a whole, CFR$_i$ for age group $i$ is probably more informative than CFR for the entire population (Nishiura, 2010b). In this regard, a method for estimating relative case fatality was proposed elsewhere (Reich et al., 2012).

Second, we cannot sometimes rely on case data due to heavy involvement of ascertainment bias. The 2009 influenza pandemic evoked a new concept in appropriately describing the virulence using a clearly defined denominator population (Nishiura, 2010a). Comparing the confirmed CFR (cCFR) and symptomatic CFR (sCFR) estimates, it is likely that the substantial difference between them is mostly attributable to the size of the denominator population. That is, statistical estimation of the CFR involves not only the tendency to underestimate the CFR due to the time delay from illness onset to death, but also the risk of overestimation due to the difficulty in ascertaining all the cases, which tend to be biased toward severely affected patients. Confirmed cases represent only a small fraction of symptomatic cases, and thus, while diagnostic coverage is less than 100%, the cCFR will always be larger than the sCFR. It should also be noted that only a portion of infected individuals develop symptoms. In an analysis of metadata of experimental challenge with interpandemic influenza viruses in humans, it has been shown that approximately one-third of infected individuals remained subclinical throughout the course of infection. Although the evidence is still scarce regarding subclinical infection with the influenza A (H1N1-2009) virus, a population-wide serological study in the United Kingdom indicated that only one-tenth of infected individuals developed an influenza-like illness that required medical attention. If we define the denominator of the CFR as the total number of infected individuals and we refer to it as the infected CFR (iCFR), the iCFR will be $0.1 \times$ the sCFR, and the sCFR may be $0.1 \times$ the cCFR. In this regard, a method for estimating the so-called infection fatality risk (IFR) was proposed elsewhere (Wong et al., 2013).

Estimation of IFR is interestingly based on two estimates in the numerator and denominator of the risk calculation. That is, the numerator is informed by the excess mortality due to uncertain number of deaths attributable to influenza, and the denominator is based on seroepidemiological survey of antibody-positive individuals. Due to that complexity of estimation, the timeliness of the estimation has remained very limited. Namely, both the estimation of excess mortality and a snapshot of seroepidemiological survey may
be completed at the end of an epidemic, and the estimates of IFR have not necessarily been obtained during the early stage of an epidemic. While many observational efforts (e.g., the use of cohort of exposed individuals in a confined unit) have been made to avoid possible biases with absolute and relative case fatality (Lipsitch et al., 2015), it remains a big challenge that the most robust and unbiased methods tend to require time and real-time features have yet to be established for those methods.

ACKNOWLEDGMENTS

H.N. received funding support from the Japan Agency for Medical Research and Development, the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant numbers 16KT0130, 16K15356, and 26700028), the Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers, the Japan Science and Technology Agency (JST) CREST program, and RISTEX program for Science of Science, Technology and Innovation Policy.

REFERENCES

Cowling, B.J., Park, M., Fang, V.J., Wu, P., Leung, G.M., Wu, J.T., 2015. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015. Euro Surveill. 20 (25), 7–13.

Donnelly, C.A., Ghani, A.C., Leung, G.M., Hedley, A.J., Fraser, C., Riley, S., Abu-Raddad, L.J., Ho, L.M., Thach, T.Q., Chau, P., Chan, K.P., Lam, T.H., Tse, L.Y., Tsang, T., Liu, S.H., Kong, J.H., Lau, E.M., Ferguson, N.M., Anderson, R.M., 2003. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet 361 (9371), 1761–1766.

Garske, T., Legrand, J., Donnelly, C.A., Ward, H., Cauchemez, S., Fraser, C., Ferguson, N.M., Ghani, A.C., 2009. Assessing the severity of the novel influenza A/H1N1 pandemic. BMJ 339, b2840.

Ghani, A.C., Donnelly, C.A., Cox, D.R., Griffin, J.T., Fraser, C., Lam, T.H., Ho, L.M., Chan, W.S., Anderson, R.M., Hedley, A.J., Leung, G.M., 2005. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. Am. J. Epidemiol. 162 (5), 479–486.

Jewell, N.P., Lei, X., Ghani, A.C., Donnelly, C.A., Leung, G.M., Ho, L.M., Cowling, B.J., Hedley, A.J., 2007. Non-parametric estimation of the case fatality ratio with competing risks data: an application to severe acute respiratory syndrome (SARS). Stat. Med. 26 (9), 1982–1998.

Kelly, H., Cowling, B.J., 2013. Case fatality: rate, ratio, or risk? Epidemiology 24 (4), 622–623.

Lipsitch, M., Donnelly, C.A., Fraser, C., Blake, I.M., Cori, A., Dorigatti, I., Ferguson, N.M., Garske, T., Mills, H.L., Riley, S., Van Kerkhove, M.D., Hernán, M.A., 2015. Potential biases in estimating absolute and relative case-fatality risks during outbreaks. PLoS Negl. Trop. Dis. 9(7), e0003846.

Ma, J., van den Driessche, P., 2008. Case fatality proportion. Bull. Math. Biol. 70, 118–133.

Mizumoto, K., Saitoh, M., Chowell, G., Miyamatsu, Y., Nishiura, H., 2015a. 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.
Mizumoto, K., Endo, A., Chowell, G., Miyamatsu, Y., Saitoh, M., Nishiura, H., 2015b. Real-time characterization of risks of death associated with the Middle East respiratory syndrome (MERS) in the Republic of Korea, 2015. BMC Med. 13, 228.

Nishiura, H., 2010a. The virulence of pandemic influenza A (H1N1) 2009: an epidemiological perspective on the case-fatality ratio. Expert Rev. Respir. Med. 4 (3), 329–338.

Nishiura, H., 2010b. Case fatality ratio of pandemic influenza. Lancet Infect. Dis. 10, 443–444.

Nishiura, H., 2010c. The relationship between the cumulative numbers of cases and deaths reveals the confirmed case fatality ratio of a novel influenza A (H1N1) virus. Jpn. J. Infect. Dis. 63, 154–156.

Nishiura, H., Klinkenberg, D., Roberts, M., Heesterbeek, J.A., 2009. Early epidemiological assessment of the virulence of emerging infectious diseases: a case study of an influenza pandemic. PLoS One 4(8), e6852.

Nishiura, H., Chowell, G., Safan, M., Castillo-Chavez, C., 2010. Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. Theor. Biol. Med. Model. 7, 1.

Reich, N.G., Lessler, J., Cummings, D.A., Brookmeyer, R., 2012. Estimating absolute and relative case fatality ratios from infectious disease surveillance data. Biometrics 68 (2), 598–606.

Wong, J.Y., Wu, P., Nishiura, H., Goldstein, E., Lau, E.H., Yang, L., Chuang, S.K., Tsang, T., Peiris, J.S., Wu, J.T., Cowling, B.J., 2013. Infection fatality risk of the pandemic A(H1N1) 2009 virus in Hong Kong. Am. J. Epidemiol. 177 (8), 834–840.