



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.



Original Article

Application of non-parametric models for analyzing survival data of COVID-19 patients

Sarada Ghosh^a, Guruprasad Samanta^{a,*}, Juan J. Nieto^b^a Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711103, India^b Instituto de Matemáticas, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain

ARTICLE INFO

Article history:

Received 13 May 2021

Received in revised form 19 August 2021

Accepted 21 August 2021

Keywords and phrases:

Relative risk

Case–fatality ratio

Kaplan–Meier estimator

Cure model

Nadaraya–Watson weight

ABSTRACT

Background: COVID-19 Coronavirus variants are emerging across the globe causing ongoing pandemics. It is important to estimate the case fatality ratio (CFR) during such an epidemic of a potentially fatal disease. **Methods:** Firstly, we have performed a non-parametric approach for odds ratios with corresponding confidence intervals (CIs) and illustrated relative risks and cumulative mortality rates of COVID-19 data of Spain. We have demonstrated the modified non-parametric approach based on Kaplan–Meier (KM) technique using COVID-19 data of Italy. We have also performed the significance of characteristics of patients regarding outcome by age for both genders. Furthermore, we have applied a non-parametric cure model using Nadaraya–Watson weight to estimate cure-rate using Israel data. Simulations are based on R-software.

Results: The analytical illustrations of these approaches predict the effects of patients based on covariates in different scenarios. Sex differences are increased from ages less than 60 years to 60–69 years but decreased thereafter with the smallest sex difference at ages 80 years in a case for estimating both purposes RR (relative risk) and OR (odds ratio). The non-parametric approach investigates the range of cure-rate ranges from 5.3% to 9% and from 4% to 7% approximately for male and female respectively. The modified KM estimator performs for such censored data and detects the changes in CFR more rapidly for both genders and age-wise.

Conclusion: Older-age, male-sex, number of comorbidities and access to timely health care are identified as some of the risk factors associated with COVID-19 mortality in Spain. The non-parametric approach has investigated the influence of covariates on models and it provides the effect in both genders and age. The health impact of public for inaccurate estimates, inconsistent intelligence, conflicting messages, or resulting in misinformation can increase awareness among people and also induce panic situations that accompany major outbreaks of COVID-19.

© 2021 The Author(s). Published by Elsevier Ltd on behalf of King Saud Bin Abdulaziz University for Health Sciences. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

The pandemic ‘severe acute respiratory syndrome (SARS)’, a viral respiratory disease, has shown how spontaneously new infectious diseases can rapidly spread out in worldwide. In December 2019, a pneumonia (i.e. COVID-19) outbreak was first pointed out at Wuhan in China. On 31 December 2019, the outbreak has been traced to a novel strain of Coronavirus, giving the interim name 2019-nCoV by WHO. It is also known as COVID-19. The virus primarily spreads among people via exhaled respiratory droplets such

as coughing or sneezing. Among those who died from the disease, the time from development of symptoms to death is between 6 to 41 days, with a median of 14 days [1]. As of 20th March 2020 more than 11,000 deaths have been attributed to COVID-19 [2]. Most of the people who died were elderly about 80% of deaths were in those over 60% and 75% had pre-existing health conditions including cardiovascular diseases and diabetes. Scientists in Italy have found traces of the new Coronavirus in waste water collected from Milan and Turin in December 2019, suggesting COVID-19 was already circulating in northern Italy before China reported the first cases. The Italian National Institute of Health looked at 40 sewage samples collected from waste water treatment plants in northern Italy between October 2019 and February 2020. An analysis released late on Thursday (December 19, 2019) said samples taken in Milan

* Corresponding author.

E-mail addresses: saradaghosha11@gmail.com (S. Ghosh), gpsamanta@math.iests.ac.in (G. Samanta), juanjose.nieto.roig@usc.es (J.J. Nieto).

and Turin on 18th December showed the presence of the COVID-19 virus [3].

Materials

In this work we have demonstrated the data-set of Spain which reports daily cumulative deaths due to COVID-19 by age and sex available on datalab of Institute National D'Études Démographiques [4]. Next, we have considered total 73,780 cases of COVID-19 diagnosed by the regional reference laboratories (Istituto Superiore di Sanità) as positive in Italy [5]. Lastly, the data-set of recovered patient from COVID-19 of Israel includes anonymized records of 15,383 Israeli COVID-19 patients [6]. But we have used only 14,835 patients (8202 men and 6633 women) data since gender group is not clear in other group.

Methodology

Non-parametric approach for estimating RR & OR

In this work, we have estimated relation between mortality due to COVID-19 and age specific gender using non-parametric approach with logistic regression [7,8,11]. We have calculated sex-specific cumulative mortality rates together with relative risks with 95% CIs for the covered time periods. Since the function $f(x)$ is arbitrary and unknown, it cannot be explicitly expressed through coefficients. The odds ratio is:

$$OR(x, x_r) = \exp(f(x) - f(x_r)) \tag{1}$$

where x_r is a specific value of the exposure taken as the reference and the CIs of the odds ratio in (1) be such as $\exp(\ln \hat{OR}(x, x_r) \pm z_{\alpha/2} SE(\ln \hat{OR}(x, x_r)))$ where SE be standard error and z be the corresponding statistics and α represents the level of significance.

Statistical approach for estimating CFR

Preliminaries

Let, $\zeta(t)$, $R(t)$, and $\delta(t)$ denote as cases, recoveries and the cumulative number of deaths respectively, where t denotes time point. Now we have considered such as:

$$e_1(t) = \frac{\delta(t)}{\zeta(t)} \tag{2}$$

$$e_2(t) = \frac{\delta(t)}{\delta(t) + R(t)} \tag{3}$$

In Eq. (2) the censoring is ignored which arises since some affected person persist in hospital. It is assumed in the second estimator that the CFR for those who persist in hospital will be same for known outcome. Apart from this, from Eq. (3), it is also a clear vision that at any certain time t elapsed from admitting to hospital, the hazards of death and recovery must be proportional.

KM method (extension part)

Let, $h_0(t)$ and $h_1(t)$ as hazard functions death and recovery states respectively & $t_{max}(s)$ be referred the maximum time of observation from admitting in hospital for death or recovery, occurring at s time in the pandemic, the probability of discharge $\theta_1(s)$ or death $\theta_0(s)$ at or before the time s be obtained such as:

$$\theta_i(s) = \int_0^{t_{max}(s)} \Theta(t) h_i(t) dt, \quad \text{where } i = 0, 1 \tag{4}$$

where $\Theta(t)$ denotes survival function whenever both endpoints are considered as a single composite endpoint. If the epidemic is end then, $\theta_0(s) + \theta_1(s) = 1$, & $\hat{\theta} = \hat{\theta}_0(s)$ is an estimate of the CFR. But, during pandemic, $S_i(t)$ are not complete, therefore, $\theta_0(s) + \theta_1(s) < 1$. The

CFR at time s should lie between $\hat{\theta}_0(s)$ and $1 - \hat{\theta}_1(s)$ followed by our estimation. We have estimated the CFR at time s is as follows:

$$\hat{\theta}(s) = \frac{\hat{\theta}_0(s)}{\hat{\theta}_0(s) + \hat{\theta}_1(s)} \tag{5}$$

At any certain time s , the estimation of hazard function estimate such as: $\hat{h}_{ij}(s) = \frac{\delta_{ij}(s)}{n_j(s)}$, where, $\delta_{ij}(s)$ be the number of events of i th type on j th day (whereas j being the time elapsed from admission). Lastly, $n_j(s)$ be the number who are remained at risk at j days. The corresponding matrix of $\hat{\Theta}_j(s)$ be $\Omega(s)$ defined as follows:

$$\omega_{jj} = \frac{\hat{\Theta}_j(s)\{1 - \hat{\Theta}_j(s)\}}{n^*(s)} \tag{6}$$

$$\omega_{jk} = \frac{\hat{\Theta}_j(s)\{1 - \hat{\Theta}_k(s)\}}{n^*(s)} \tag{7}$$

where $j > k$ and at time point s , $n^*(s)$ be the total sample size which is considered as halfway between the total uncensored sample size and total sample size in the pandemic.

Non-parametric approach for cure-rate estimator

One of the most important cure models is the proportional hazards cure (PHC) model defined as:

$$S(t | z, x) = \exp(-\theta(z)F_0(t | x)) \tag{8}$$

where $S(t|z, x)$ is the improper population survival function, $F_0(t|x)$ is a proper baseline cdf (cumulative distribution function) and $\theta(z)$ is a positive function of z and is formulated as $\exp(\beta^T z)$. Let, Y_i be the failure time of the i th subject. Let C_i be the i th censoring time and $t_i = \min(Y_i, C_i)$ be the observed failure time. Let δ_i be the censoring indicator with $\delta_i = 0$ when $t_i = C_i$ and $\delta_i = 1$ when $t_i = Y_i$; z_i and x_i are the corresponding covariates of t_i . The observed data is

$$D(\{(t_i, \delta_i, z_i, x_i), i = 1, 2, \dots, n\}) \tag{9}$$

In this work we use the dataset as follows:

$$D(\{(Y_i, \delta_i, z_i), i = 1, 2, \dots, n\}) \tag{10}$$

Compared with the definition in Eq. (9), x is suppressed as it only affects the survival function of patients who are not cured. A non-parametric methods for cure-rate based on KM method at the largest uncensored failure time such as [12–14]:

$$\hat{\pi} = \hat{S}(T_{(n)}) \tag{11}$$

where $\hat{S}(\cdot)$ is the KM estimate of the population survival function and $T_{(n)}$ is the largest uncensored failure time. The generalized maximum likelihood estimation of the population survival function and the cure-rate are as follows [15]:

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \left(1 - \frac{\delta_i}{n + m - i + \delta_i} \right) \tag{12}$$

and

$$\hat{\pi} = \prod_{i=1}^n \left(1 - \frac{\delta_i}{n + m - i + \delta_i} \right) \tag{13}$$

where $t_{(1)} < t_{(2)} < \dots < t_{(n)}$ are the ordered failure times. Whenever $m = 0$, the survival function estimation reduces to the Kaplan–Meier estimation and the cure-rate estimation reduces to the estimate of Maller and Zhou [14] as mentioned in (11). After getting the estimations of θ , where covariate z is suppressed, the cure-rate was estimated as follows:

$$\hat{\pi} = \exp(-\hat{\theta}) \tag{14}$$

For survival data $(Y_i, \delta_i, z_i), i = 1, \dots, n$ without a cure fraction, the generalized Kaplan-Meier estimator is as follows [16]:

$$\hat{S}(t | z) = \begin{cases} \prod_{j=1}^n \left[\frac{1 - \sum_{r=1}^n I(Y_r \leq Y_j) B_r(z)}{1 - \sum_{r=1}^n I(Y_r < Y_j) B_r(z)} \right]^{I(Y_j \leq t; \delta_j = 1)} & \text{if } t \leq Y_{(n)} \\ 0 & \text{if } t > Y_{(n)} \end{cases} \quad (15)$$

where $I(A)$ is the indicator function.

The proper weight function $B_j(z)$ satisfy the following conditions:

$$(i) B_j(z) \geq 0 \quad \text{and} \quad (ii) \sum_{i=1}^n B_j(z) = 1 \quad (16)$$

Here $Y_{(n)}$ refers as the largest failure time. When the covariate z is compressed, (14) reduces to the KM method. For survival data mentioned in (9) with a cure fraction, we have proposed for estimating $\pi(z)$ as:

$$\hat{\pi}(z) = \hat{S}(Y_{(n)}^1 | z) \quad (17)$$

where $\hat{S}(y | z)$ is the generalized Kaplan–Meier estimate with data as mentioned in (9), if there is no cure fraction. Let us consider $Y_{(n)}^1$ as the largest uncensored failure time. The bandwidth in the weight functions is as follows $h = \alpha n^{-\frac{1}{3}}$ [17]. We have tried α from 0.1 to 2 and observed that the bandwidth $h_m = 0.2$ for male and $h_f = 0.1$ for female work very well [17]. The weight functions for $B(z)$ are considered in this work. The Nadaraya–Watson weight function is as follows:

$$B_j(z) = \frac{K(\frac{z_j - z}{h})}{\sum_{r=1}^n K(\frac{z_r - z}{h})}, \quad j = 1, 2, \dots, n \quad (18)$$

$K(\cdot)$ is a proper kernel function. If the covariate is absent, $B_r(z) = \frac{1}{n}, r = 1, 2, \dots, n$, and the proposed estimator becomes the KM estimator at $Y_{(n)}^1$. Therefore, the proposed estimator as mentioned in (15) reduces to the cure-rate estimator of no-covariate case [14].

Results

Example 1

This is based on COVID-19 dataset of Spain available on the Institut National D’Études Démographiques (INED) website. Sex-related cumulative mortality rates (CMRs) from COVID-19 per 100,000 men and women are calculated for each day for the covered time periods of Spain. Furthermore RRs and ORs with 95% CIs are also calculated for four age-groups (<60, 60–69, 70–79 and ≤80 years) using COVID-19 data-set of Spain.

In Spain, the CMRs are increased with advancing age (from Table 1). The largest sex differences are found at ages from 60 to 69 (RR = 2.56, 95% C.I. 2.31–2.83) years in Spain and smallest sex difference is found at ages ≥80 years (RR = 1.65, 95% C.I. 1.60–1.70) whenever illustrating both RRs and ORs, i.e. 4.13 with 95% C.I. 3.88–4.41(for 60–69 years) and 2.35 with 95% C.I. 2.30–2.42 (for ≥80 years) for men vs. women. The result also provides the largest sex difference for people aged above 80 years whenever considering the cumulative mortality rate.

Example 2

The non-parametric KM based approach provides sensible estimates of the CFR whenever the degree of censoring is moderate. In Italy, there are various reasons for higher CFR such as: (i) the age structure of the Italian population (i.e. 2nd oldest in the world), (ii) the highest rate of antibiotic resistance deaths in Europe.

Table 3

Age effects (for both genders) on cure-rates for COVID-19 patients.

Age group (in years)	Cure rate	
	Male	Female
≤19	0.0789	0.0679
19–29	0.0855	0.0609
29–39	0.0557	0.0412
39–49	0.0537	0.0414
49–59	0.0528	0.0407
≥60	0.0601	0.0521

In this work, the data are collected from the Istituto Superiore di Sanità (ISS), Rome at 4 pm on March 26, 2020. It was reported that total 73,780 cases of COVID-19 diagnosed by the regional reference laboratories as positive for COVID-19 (15,781 more cases than in the previous bulletin referred to March 23rd, 2020). The diagnosis of COVID-19 infection was confirmed in 99% of the samples sent by regional reference laboratories and processed by the national reference laboratory (ISS). There is a clear vision in Fig. 1, that the number of censored cases are gradually increasing mode according as different time points. Information on sex is known for 73,044 out of 73,780 cases. The difference in the number of cases reported by sex increases progressively in favor of sex subjects male up to the age-group ≥70–79, with the exception of the 20–29 and 30–39 years in which the number of female subjects is slightly higher. For ≥90 years, the number of cases of female sex exceeds that of male cases probably for the demographic structure of the population.

The observed CFR are increasing and reflect a vulnerable effect for the age-group in both cases. In such pandemic, the simple estimator from Eq. (2), based on deaths and cases ratio, denoted as e_1 , underestimates the CFR. Therefore, the numerator in this work can able to underestimate the total deaths related to COVID-19, occurring eventually in sample. The another estimate obtained from Eq. (3) based on the ratio of deaths whose outcome is already perceived denoted as e_2 , which is sensible in such pandemic. In different scenarios, specific group-wise estimates of CFR are very desirable for both age and gender. Table 2 provides the estimates of the CFR obtained for several age-groups. It is also observed that CFR differences for various age-group are changing for larger Italy cohorts. This analysis provides that patients aged 70–79 years and >79 years reveal remarkable effects on age due to COVID-19 in CFR purpose for both gender cases. It also shows a significantly slower mortality rate in case of female patients (specially in case of aged person over 80 years or more). Relatively, <70 years age-groups signify average rate of CFR which indicates the prognosis comparatively low rate for young aged people.

Example 3

The summary of COVID-19 dataset of Israel used in this work are posted (June 28, 2020) by the Israel Ministry of Health (IMH). The mortality and morbidity due to COVID-19 are higher for old aged patients [18]. This important information can improve our knowledge of COVID-19 and helps for assisting public healthcare for future Coronavirus pandemics or epidemics. The subgroups categorized by age are shown on underlying dataset according to decades for the range of age from 20 to 59 years, while for older or younger COVID-19 patients age-groups are assigned here as >60 years and 0–19 years respectively. In dataset, recovery is determined by taking two consecutive negative results of COVID-19 after testing ‘test of throat swab’.

For proposing cure-rate estimation, we have considered the bandwidth $h_m = 0.2$ for male and $h_f = 0.1$ for female and (0, 1) for standardized age variable. For non-parametric model, it is shown in Table 3 that the rates of cure are substantial for age-group under

Table 1
CMRs of COVID-19 per 100,000 men and women and RRs and ORs for men vs. women of Spain.

Age-group	CMR with 95% C.I. (for men)	CMR with 95% C.I. (for women)	RR with 95% C.I.	OR with 95% C.I.
<60	3.76 (3.48–4.06)	1.74 (1.55–1.95)	2.17 (1.89–2.48)	2.86 (2.50–3.17)
60–69	50.41 (47.70–53.26)	19.72 (18.11–21.47)	2.56 (2.31–2.83)	4.13 (3.88–4.41)
70–79	187.42 (181.13–193.93)	73.52 (69.95–77.28)	2.55 (2.40–2.70)	4.11 (3.96–4.29)
≥ 80	597.80 (583.26–612.70)	362.22 (353.52–371.13)	1.65 (1.60–1.70)	2.35 (2.30–2.42)

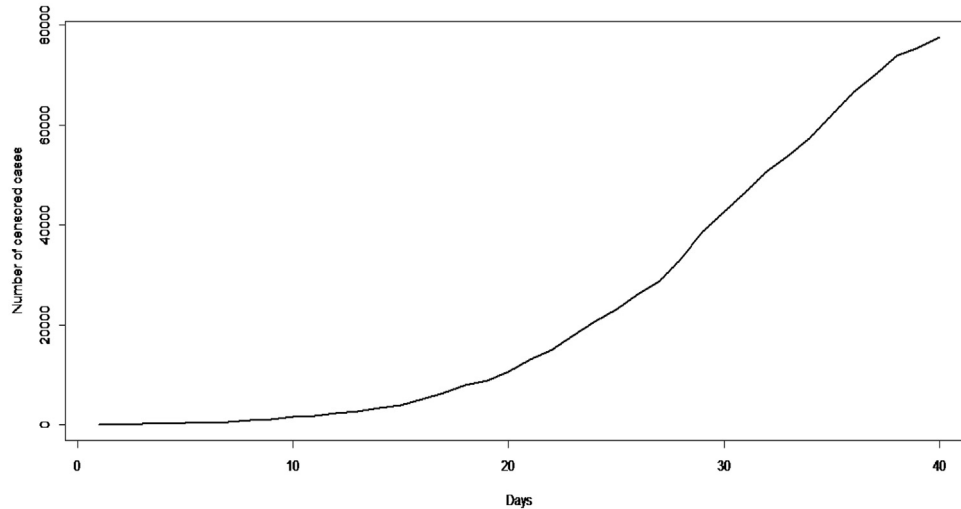


Fig. 1. Estimate of the CFR by age (for both genders) for epidemic of Covid-19.

Table 2
Estimate of the CFR by age (for both genders) for COVID-19.

Age group (in years)	Male			Female		
	Case fatality ratio	Estimate	95% confidence interval	Case fatality ratio	Estimate	95% confidence interval
30–39	0.6	0.43	0.27, 0.69	0.2	0.26	0.16, 0.42
40–49	1.1	0.91	0.72, 1.16	0.4	0.55	0.4, 0.7
50–59	2.4	2.05	1.81, 2.33	1.1	1.23	1.08, 1.40
60–69	6.9	6.67	6.22, 7.15	4.28	4.02	3.71, 4.34
70–79	19.78	19.71	18.98, 20.47	11.6	11.86	11.26, 12.50
80–89	29.2	29.81	28.78, 30.87	18.9	17.94	17.11, 18.80
≥90	30.8	32.26	30.01, 34.68	20.4	19.1	18.05, 20.88

consideration and the cure-rate ranges from 5.3% to 9% approximately for male cases and from 4% to 7% approximately for female in the non-parametric model. Whenever age is very large or very small, the cure-rates tend to increase as age increases and demonstrate S-shaped curve. For a full vision of how age affects the cure-rate for both gender purpose, we have estimated the cure-rate corresponds to the actual age from ≤ 19 to ≥ 60 and plotted the relation between the cure-rate and age in Fig. 2.

The S-shape relationship between cure-rate and age from underlying the data-set is depicted here. The original data-set is divided into 6 subgroups based on age: ≤ 19 , (19, 29], (29, 39], (39, 49], (49, 59) and ≥ 60 for both male (red line in Fig. 2) and female (blue line in Fig. 2). The approach proposed by Maller and Zhou (1992) is used in each of the 6 subgroups to estimate the cure-rates, which are shown in Fig. 2 as per age subgroup [14]. The Table 3 shows that the cure-rates for male purpose are decreasing whenever age increases from subgroup 2 to subgroup 5. But the cure-rates of subgroup 1 and subgroup 6 do not obey such pattern and the former is greater (except group 2) than others and the latter is second greater (except group 2) than the cure-rates of their respective adjacent groups. Table 3 also shows that the cure-rates for female purpose are decreasing whenever age increases from subgroup 1 to subgroup 5. However, the cure-rates of subgroup 6 do not follow this pattern.

Results of the non-parametric approach investigate the influence of covariates on these models and also provide that both genders and age have a significant effect (shown in Tables 2 and 3) and the depicted curves are displayed in Fig. 2 for individual group purpose.

Conclusion

In epidemiology, case fatality rate typically is used as a measure of disease severity and is often used for prognosis. It also can be used for evaluating the effect due to new treatments, with measures that are decreasing as treatments improve. On daily cumulative deaths by age and sex due to COVID-19 in Spain, it is concluded that the risk of death increases with age, and that men have higher mortality from COVID-19 than women in almost all age-groups using population estimates. Sex differences are increased from ages <60 years to 60–69 years but decreased thereafter with the smallest sex difference at ages ≥ 80 years in case for estimating both purposes RR and OR. A recent work provides the survival in seven populations under extreme conditions from famine, epidemics and slavery: even when mortality is very high, women can survive on average better than men [19,20]. The work provides that the survival advantage of women has fundamental biological underpinnings, but the female advantage is modulated by a complex

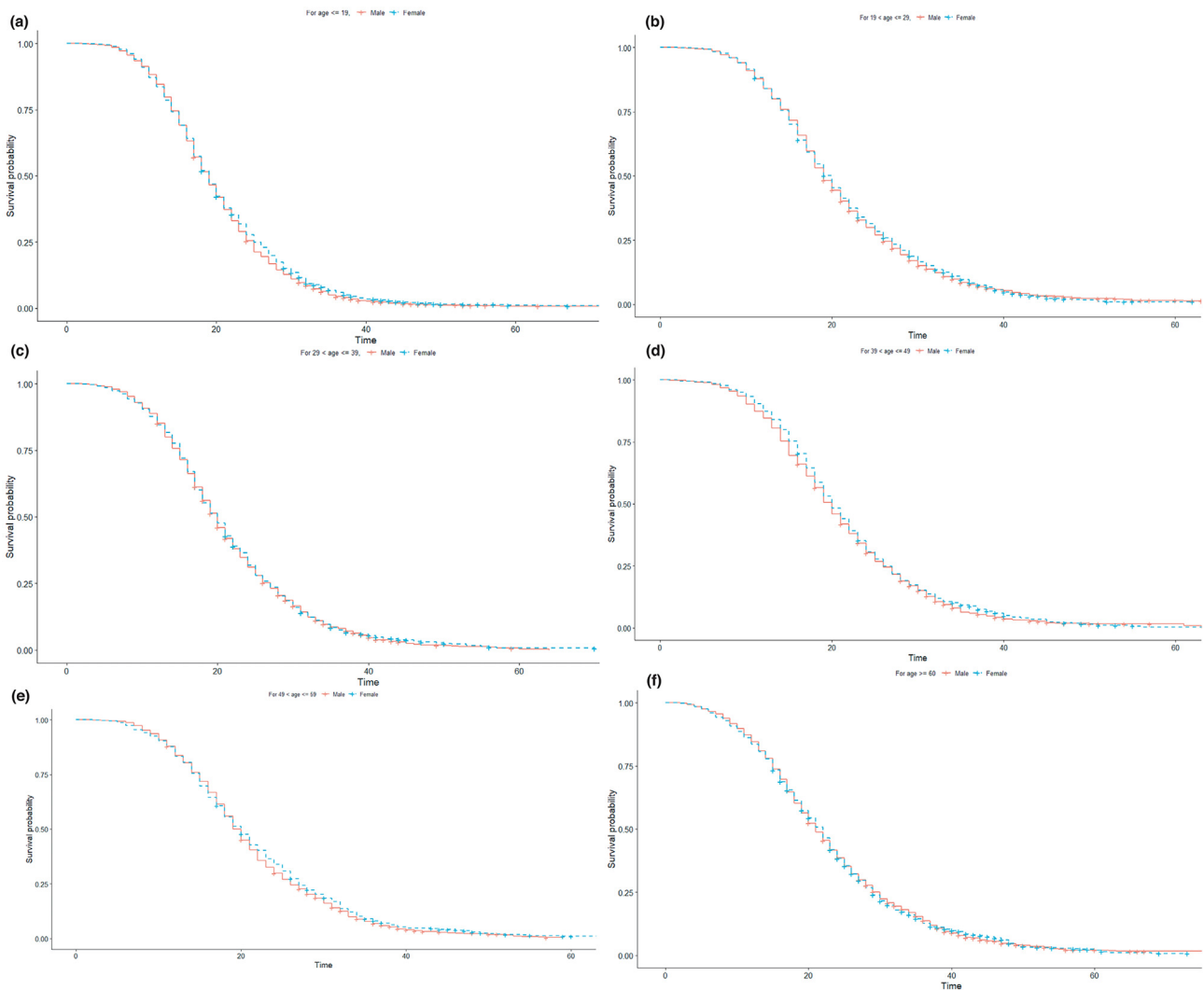


Fig. 2. KM estimates of the COVID-19 patients in 6 groups. (For interpretation of the references to color in this figure citation, the reader is referred to the web version of this article.)

interaction of biological, environmental and social factors [20]. Although we have demonstrated a higher mortality from COVID-19 for men than for women in almost all age-groups, we find, as hypothesized, a reduction in the relative risk of mortality for men at later ages, consistently with findings elsewhere [21]. This work provides the importance of addressing the impact of sex on mortality from disease epidemics, but studies using individual-level data are needed to confirm an interaction between age and sex in COVID-19 mortality in order to guide clinical care personal and to address questions of whether men require additional surveillance, prevention, and earlier intervention than women. The work also provides two approaches for estimating the CFR computed from the known outcomes and then proceed with modified KM approach that adequately estimates the CFR during the COVID-19 pandemic. The underlying first technique is appealing due to its simplicity and after that the ease with which it can be computed. The modified KM estimator performs for such censored data and detects the changes in the CFR more rapidly for both genders and age-wise. The health impact of public for inaccurate estimates, inconsistent intelligence, conflicting messages or resulting in misinformation can increase awareness between people and also induce panic situation that accompany major outbreaks of COVID-19. Among several important aims, another uttermost important challenge is to encounter during such disease epidemic. COVID-19 is perceiving the rea-

sons underlying the contrast in all CFRs which are reported for Italy.

In future epidemics (in the context of COVID-19) the careful estimation and proper analysis of the CFR can be used for evaluating the effectiveness of any new treatments. In pandemic condition, we have demonstrated the CFR initially which is defined by modified KM technique as displayed in this work. Apart from this, we have also illustrated the non-parametric approach to explore the effect of covariates on cure-rate with a cure fraction for underlying survival data. Our simulations display that the proposed non-parametric approach is flexible and can also accommodate a complex consequence of a covariate on cure-rate. The proposed non-parametric method is used for distinct age-group in both gender cases. It also can be extended to a multiple-covariate case specially whenever the number of covariates is greater than two. In the present work, both genders and age have a significant role in this prevalence predicted from the underlying approaches with current Coronavirus data. The pandemic COVID-19 shows an increased number of cases and a greater risk of severe disease with increasing age and for being male. Overall, weighted approach of cure-rate has downward tendency as age is increasing for both gender purposes.

Data availability statement

The data used to support the findings of the study are available within the article.

Funding

Agencia Estatal de Investigacion (AEI) of Spain, cofinanced by the European Fund for Regional Development (FEDER) corresponding to the 2014–2020 multiyear financial framework, project PID2020-113275GB-I00; and by Xunta de Galicia under grant ED431C 2019/02, and Instituto de Salud Carlos III, grant COV20/ 00617.

Conflict of interest

The authors declare that they have no conflict of interest regarding this work.

Ethical approval

Not required.

Acknowledgements

The authors are grateful to the learned reviewers and Dr. Rehab Hosny El-Sokkary (Editor) for their careful reading, valuable comments and helpful suggestions, which have helped them to improve the presentation of this work significantly. The research of J.J. Nieto has been partially supported by the Agencia Estatal de Investigacion (AEI) of Spain, cofinanced by the European Fund for Regional Development (FEDER) corresponding to the 2014–2020 multiyear financial framework, project PID2020-113275GB-I00; and by Xunta de Galicia under grant ED431C 2019/02, and Instituto de Salud Carlos III, grant COV20/00617.

References

- [1] Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmunity* 2020;109:102433. <http://dx.doi.org/10.1016/j.jaut.2020.102433>.
- [2] WHO. Coronavirus disease 2019 (COVID-19) situation report-61; 2020 <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200321-sitrep-61-covid-19.pdf>.
- [3] Kelland K. Italy sewage study suggests COVID-19 was there in December 2019. *HEALTHCARE & PHARMA*; 2020.
- [4] Institut National D'Etudes Demographiques (INED) COVID-19. Demographics of COVID-19 deaths 2020. <https://dc-covid.site.ined.fr/en/data/> [accessed 30 June 2020].
- [5] Istituto Superiore di Sanità (ISS 2020). <https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID>.
- [6] Israel Ministry of Health (IMH 2020). <https://data.gov.il/dataset/covid-19>.
- [7] Hastie TJ, Tibshirani RJ. Generalized additive models. London, United Kingdom: Chapman and Hall Ltd; 1990.
- [8] Green PJ, Silverman BW. Non-parametric regression and generalized additive models: a roughness penalty approach. New York, NY: Chapman and Hall; 1994.
- [11] Ghosh S, Samanta GP. Statistical modeling for cancer mortality. *Lett Biomath* 2019;6(2). <http://dx.doi.org/10.1080/23737867.2019.1581104>. SPECIAL ISSUE.
- [12] Tsodikov A. A proportional hazards model taking account of long-term survivors. *Biometrics* 1998;54:1508–16.
- [13] Tsodikov A. Estimation of survival based on proportional hazards when cure is a possibility. *Math Comput Modell* 2001;33:1227–36.
- [14] Maller RA, Zhou S. Estimating the proportion of immunes in a censored sample. *Biometrika* 1992;79:731–9.
- [15] Laska EM, Meisner MJ. Nonparametric estimation and testing in a cure model. *Biometrics* 1992;48:1223–34.
- [16] Beran R. Nonparametric regression with randomly censored survival data. Berkeley: University of California; 1981. Technical report.
- [17] Dabrowska DM. Non-Parametric regression with censored survival time data. *Scand J Stat* 1987;14:181–97.
- [18] Ghosh S, Samanta GP, Mubayi A. Comparison of regression approaches for analyzing survival data in the presence of competing risks: an application to COVID-19. *Lett Biomath* 2021;8(1):29–47 <https://lettersinbiomath.journals.publicknowledgeproject.org/index.php/lib/article/view/305>.
- [19] Barford A, Dorling D, Davey SG, Shaw M. Life expectancy: women now on top everywhere. *BMJ* 2006;332(7545):808.
- [20] Zarulli V, Barthold Jones JA, Oksuzyan A, Lindahl-Jacobsen R, Christensen K, Vaupel JW. Women live longer than men even during severe famines and epidemics. *Proc Natl Acad Sci U S A* 2018;115(4):832–40.
- [21] Promislow DEL. A geroscience perspective on COVID-19 mortality. *J Gerontol A Biol Sci Med Sci* 2020.