Survival analysis using the Covid-Death Mean-Imputation (CoDMI) algorithm: a first clinical application in radiation oncology

Francesca De Felice (fradefelice@hotmail.it)
Policlinico Umberto I “Sapienza” University of Rome
Luca Mazzoni
University of Florence
Daniela Musio
Policlinico Umberto I “Sapienza” University of Rome
Vincenzo Tombolini
Policlinico Umberto I “Sapienza” University of Rome
Franco Moriconi
University of Perugia

Study protocol

Keywords: Survival, CoDMI algorithm, Covid-19, death, Kaplan-Meier, event, cancer

Posted Date: April 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1539967/v1

Abstract

Purpose. We illustrate a clinical application of Covid-Death Mean-Imputation (CoDMI) algorithm in survival analysis.

Material and methods. We analyzed 94 patients treated for primary locally advanced rectal cancer (LARC).

Overall survival was calculated in months from diagnosis to first event (last follow-up/death).

Because Covid-19 death events potentially bias survival estimation, to eliminate skewed data due to Covid-19 death events the observed lifetime of Covid-19 cases was replaced by its corresponding expected lifetime in absence of the Covid-19 event using CoDMI algorithm.

In a traditional Kaplan-Meier approach, patient died of Covid-19 (DoC) can be: i) excluded to the cohort, or ii) counted as censored (Cen), or iii) considered as died of disease (DoD). CoDMI algorithm offers an additional, more satisfactory option: iv) DoC events are mean-imputed by the Kaplan-Meier estimator. With this approach, observed lifetime of each DoC patient is considered as an “incomplete data” and is extended by an additional expected lifetime computed using the classical Kaplan-Meier model.

Results. 16 patients were DoD, 1 patient was DoC and 77 cases were Cen. The DoC patient died of Covid-19 52 months after diagnosis. CoDMI algorithm computed the expected future lifetime provided by the Kaplan-Meier estimator applied to the no-DoC observations as well as to the DoC data itself. Given the DoC event at 52 months, CoDMI algorithm estimated that this patient would be died after 79.5 months of follow-up.

Conclusion. CoDMI algorithm leads to “unbiased” probability of overall survival in LARC patients with Covid-19 infection, compared with that provided by a naïve application of Kaplan-Meier approach. This allows for a proper interpretation/use of Covid-19 events in survival analysis. A user-friendly version of CoDMI is freely available at https://github.com/alef-innovation/codmi.

Introduction

Since the World Health Organization (WHO) declaration of 2019 coronavirus disease (Covid-19) on 2020 March 11, it has been proven that cancer patients have a higher risk of Covid-19 infection and death than individuals without cancer diagnosis [1]. This clinical evidence may introduce problems in the survival estimates in oncologic clinical trials. Once pandemic data will be fully collected, the trials population is expected to include a relevant proportion of death observations caused by Covid-19. Therefore, there is a need to define a more suitable way to correctly categorize this information for the estimation of the survival rates [2]. A possible, and perhaps more appropriate, approach to survival analysis when Covid deaths are present in an oncology clinical trial in addition to cancer deaths, could be based on the theory of *competing risks*, see e.g. [3], Chap. 8. This would lead to dealing with sub-distributions (the marginal probabilities of each competing risk) and would require appropriate statistical tests to be used. However, most of the clinical researchers perform survival analysis by routinary applications of standard statistical tools, like Kaplan-Meier estimator, which are actually not well suited for competing events. As noticed in [4], “Kaplan-Meier curves have become familiar friends to medical researchers, a *lingua franca* for reporting clinical trial results." When these standard tools cannot be directly applied, there is a danger that Covid deaths observations will simply be discarded from the study sample, with serious loss of information. With this in mind, we limit ourselves to a more pragmatic approach, consisting in operating directly on the data through mean-imputation, transforming each observed competing event into a virtual event-of-interest or a censoring. This is done by a purpose-built algorithm recently proposed. The advantage of this approach is that any statistical tool suitable for

the analysis of non-competing events can be used, without requiring additional skills. The cost of this choice is that it is necessary to accept some approximations and a certain level of bias.

In a traditional Kaplan-Meier approach, patients died from Covid-19 could be: i) excluded from the cohort, or ii) counted as lost-to-follow-up, or iii) considered as died of disease. But all these three options are unsatisfactory. Option (i) represents a loss of data (we know that Covid-19 patients were alive since the time of observed death); both option (ii) and (iii) imply substantial bias in survival estimates. A new statistical method recently presented [5] allows for a further, more satisfying option: iv) Covid-19 data is adjusted using the Kaplan-Meier model itself, so that the traditional model can be consistently applied to the entire sample of observations. This is realized by a freely available algorithm named *Covid-Death Mean-Imputation* (CoDMI) [5]. Essentially, in CoDMI algorithm the observed lifetime of each Covid-19 case is replaced by the corresponding expected lifetime in absence of the Covid-19 event, where the expectation is derived consistently with the Kaplan-Meier estimator (this replacement is usually referred to as *mean-imputation*).

Here we present a first clinical application of CoDMI algorithm in a radiation oncology scenario. This observational study was planned to describe a population (P) and observe outcomes (O). The hope is to give a picture of what happens in a group of subjects when researcher actively uses CoDMI algorithm in survival analysis. In the sample considered in this study, we have only one Covid-19 death out of 94 oncological patients. Therefore, the effect of the Covid-19 event on the survival estimates is expected to be marginal with any option considered. However, this simplified situation is also interesting, because it makes it easier to interpret the effects of different options and illustrate how CoDMI algorithm can be used.

Methods And Materials

Study design. An observational study was carried out with data prospectively collected from locally advanced rectal cancer patients, treated with curative intent, from July 2014 to September 2020, at the XXX. Medical records of adults patients (≥ 18 years of age), who presented with a histologically confirmed diagnosis of adenocarcinoma of the rectum were collected. Approval for the study was obtained from Ethics Committee (ref. 6452) and written informed consent was obtained from all patients before treatment. Data were collected in a digital form and included patient-related, tumor-related and follow-up-related variables. The possible diagnosis of the Covid-19 infection was recorded. All data were anonymized prior to any analysis.

Treatment. All patients referred to a multidisciplinary gastrointestinal tumor board prior to treatment initiation. Clinical examinations – complete medical history and careful physical examination – were combined with imaging – trans-rectal ultrasound, total-body contrast-enhanced computed tomography (CT) and magnetic resonance imaging (MRI) of the pelvis – to properly define the local (T), regional nodal (N) and distant (M) extent of the disease. External beam definitive radiotherapy with an intensity modulated technique (IMRT) was delivered to a total dose of 50.4–54 Gy (1.8 Gy per fraction) to the tumor volume and 45 Gy (1.8 Gy per fraction) to the whole pelvis. Concurrent chemotherapy consisted of weekly oxaliplatin (50 mg/m^2 on the first day of each week of RT) and 5-FU (200 mg/m^2/5 daily continuous infusions). Surgery was planned 7–9 weeks after the end of concomitant treatment and its type was left to the surgeon's discretion. Induction chemotherapy was recommended in case of bulky primary or nodal disease from October 2015 [6].

Follow-up. According to local protocol [7], patients were followed at 3-month intervals for two years and every 6 months for subsequent years with a physical examination, complete blood count, blood chemistry and trans-rectal

ultrasound. Patients were monitored closely to detect local recurrent disease and distant metastasis by clinical exam and imaging, including total body CT and pelvic MRI recommended annually for up to 5 years after treatment.

Statistical analysis. Statistical analysis was performed using R 4.0.

In standard descriptive analysis, continuous variables were reported as means ± 1 standard deviation (SD) and categorical variables as frequencies or percentages.

As the study was designed to analyze the usefulness of CoDMI algorithm, the primary outcome focused solely on overall survival (OS), which was defined as the time between the initial diagnosis to the first event (death from any cause). In the calculation of the OS rates, patients observed to die of Covid-19 were taken into account following alternative options. They were: i) excluded from the data analysis; ii) counted as lost-to-follow-up; iii) considered as died of disease; iv) “adjusted” using CoDMI algorithm [5]. Survival rates were estimated using the Kaplan-Meier model directly in the first three cases and after mean imputation by CoDMI in the fourth case. Moreover, the mean imputation, that is the replacement of the observed lifetime with an estimated virtual life expectancy, was made under two additional options: iv.a) assuming a virtual death of disease (CoDMI applied in standard form); iv.b) assuming a virtual censoring (CoDMI with *adjustment for censoring*). A detailed analysis of CoDMI algorithm, as well as the underlying model, is beyond the aim of this study; a comprehensive overview and a tutorial of the fundamental aspects of the algorithm is provided in [5]. Briefly, we only recall here that the CoDMI adjustment is obtained by running the Kaplan-Meier estimator itself starting from the no-DoC observations, then providing a consistent estimate of the life expectancy beyond the DoC date. A user-friendly version of CoDMI programmed in R is freely available at https://github.com/alef-innovation/codmi.

Results

Overall, 94 patient records were collected. All patients had been treated with a curative intent. Patient and tumor characteristics are listed in Table 1. Mean age was 67.4 years and 47 patients (50.0%) were female. There were no patients with coexisting serious medical conditions.

Table 1
Study population characteristics

Characteristic	Total (N = 94)
Gender	
Female	47 (50.0%)
Male	47 (50.0%)
Age	
Mean (SD)	67.4 (11.5)
Range	22.0–87.0
Smoker	
No	51 (54.3%)
Yes	43 (45.7%)
Performance status	
0	87 (92.6%)
1	7 (7.4%)
Body mass index	
≤ 35 kg/mq	89 (94.7%)
> 35 kg/mq	5 (5.3%)
Co-morbidity	
None	21 (22.3%)
Cardiovascular	61 (64.9%)
Other	12 (12.8%)
Distance from anal verge (cm)	
> 8	32 (34.0%)
6–8	26 (27.7%)
< 6	36 (38.3%)
Clinical tumor stage (cT)	
2	13 (13.8%)
3	56 (59.6%)
4	25 (26.6%)
Clinical nodal stage (cN)	
0	9 (9.6%)

SD: standard deviation; kg/mq: kilograms per square meter; cm: centimeters

Characteristic	Total (N = 94)
1	25 (26.6%)
2	60 (63.8%)
cT dimension (cm)	
> 5	33 (35.1%)
≤ 5	61 (64.9%)
SD: standard deviation; kg/mq: kilograms per square meter; cm: centimeters	

The vast majority of patients (*n* = 85, 90.4%) had regional lymph node involvement at diagnosis.

Most tumors were located in the low rectum (*n* = 60, 63.8%).

In the sample, the minimum and the maximum observed survival time was *t* = 2 months and *t* = 82 months, respectively. In the entire cohort, 17 patients (18.1%) died and the last death event occurred 53 months after diagnosis. Of these 17 death events, one case (5.9%) was attributed to Covid-19 and was recorded after 52 months. Therefore, on the entire sample of 94 patients, there were 16 patients *dead of disease* (DoD), 1 patient *dead of Covid-19* (DoC) and 77 cases of *censoring* (Cen). This is summarized in Table 2. The basic problem in this kind of samples is how to treat the DoC cases in order to obtain a usual sample composed of only DoD and Cen cases and to which the classical Kaplan-Meier estimator, *as well as any statistical tool for survival analysis*, can be applied.

Table 2
Details of exit causes in the entire sample

exit cause	number	min *t* (months)	max *t* (months)
DoD	16	2	53
Cen	77	11	82
DoC	1	52	52
Total	94		
DoD: dead of disease; Cen: censored; DoC: dead of Covid-19; min: minimum;			
t: survival time; max: maximum			

Let us consider the Kaplan-Meier estimator. Using the standard model, it is possible to estimate the survival probabilities in our sample according to three different options: i) excluding the DoC event from the data analysis (accounting for a total of 93 observations); ii) considering the DoC event as a Cen event at the same time point; iii) considering the DoC event as a DoD event at the same time point. The corresponding Kaplan-Meier survival plots are provided in Fig. 1a. In the figure, the black curve (*without DoC*) corresponds to option (i), the green curve (*DoC as Cen*) refers to option (ii) and the red curve (*DoC as DoD*) corresponds to option (iii). Using CoDMI algorithm an additional option is available. This essentially consists in considering the DoC event as an “incomplete data” which is adjusted by *mean imputation*, that is the observed lifetime of the DoC patient is extended by an additional expected lifetime. The crucial point is that this lifetime extension is computed by the algorithm using the Kaplan-Meier model itself and therefore does not introduce inconsistencies in the survival estimates.

In Fig. 1a a blue curve (*DoC Imputed*) is also provided corresponding to option (iv.a), where the DoC event is considered as an "incomplete data" which is adjusted by mean imputation using CoDMI algorithm in its standard form, that is assuming that it would have been a virtual DoD event at a later time point. In this case, CoDMI algorithm estimates that the DoC patient that died due to Covid-19 at time $t = 52$ months, without Covid-19 infection would have survived for additional 27.5 months and would have died of tumor at time $t = 79.5$ months (52 + 27.5 months). Therefore, the observed DoC event at time 52, indicated by a red triangle on the blue line in the figure, is changed as a *virtual DoD* event at time 79.5, which is indicated by a circle.

If the virtual DoC event is considered unlikely, it is possible to modify the standard-form result by applying CoDMI algorithm with an additional option, called *adjustment for censoring*, where the survival estimate is obtained counting the DoC event as a *virtual Cen* rather than a virtual DoD. If this option is used, a reverse Kaplan-Meier estimate is computed by CoDMI, providing an expected survival of 13.98 months beyond the observed DoC event. In this case, the DoC event at time 52 is then changed as a *virtual* Cen event at time $t = 65.98$ months (52 + 13.98 months). This is illustrated by the blue line in Fig. 1b, where the circle on the blue curve (*DoC Imputed*) corresponds now to the time point of the virtual Cen event.

The choice between the standard and the no-standard mode of CoDMI application can be a matter of the clinician's discretion, but it can also be made directly by the algorithm, which computes the probability of a virtual DoC event *vs* that of a virtual Cen event.

The survival rates illustrated in Fig. 1 and Fig. 2 are reported in Table 3 for two selected time points, 2 and 5 years, together with the corresponding 95% confidence intervals (CI). For options (i), (ii) and (iii) the confidence intervals are computed using the classical Greenwood's formula. For options (iv.a) and (iv.b), where the observed DoC time point is replaced by the expected DoD and Cen time point, respectively, the confidence intervals must be computed taking into account that changing an observed value with an expected value (i.e. applying the mean imputation) provides an increase of the estimation uncertainty. This correction is provided by CoDMI algorithm, which includes a built-in extension of Greenwood's formula. The table shows that the 2-year OS rate was 89.8% (CI 83.7–96.4) and the 5-year OS rate was 74.8% (CI 63.7–0.87.9), when the DoC event was not included in survival analysis (option i). Considering that Covid-19 patient died after 52 months of follow-up, the 2-year OS rate was equal (81.1%, CI 72.3–91.0) in all other cases (option ii, iii, iv.a and iv.b). The 5-year OS where different in case of DoC as Cen, DoC as DoD, DoC as virtual DoD and DoC as virtual Cen, and survival rates were 75.1%, 71.9%, 75.3% and 75.3%, respectively. As concerning the confidence intervals, the differences between those provided by the classical Greenwood's formula and those computed with the extended formula, result to be immaterial in this case with a single DoC event.

Table 3
Survival rate estimates for different treatment of Covid-19 death event

Option	2y-OS	5y-OS
i) Without DoC	89.8% (CI 83.7–96.4%)	74.8% CI 63.7–87.9%)
ii) DoC as Cen	89.9% (CI 83.9–96.4%)	75.1% (CI 64.1–88.1%)
iii) DoC as DoD	89.9% (CI 83.9–96.4%)	71.9% (CI 59.9–86.2%)
iv.a) DoC as virtual DoD	89.9% (CI 83.9–96.4%)	75.3% (CI 64.2–88.3%)
iv.b) DoC as virtual Cen	89.9% (CI 83.9–96.4%)	75.3% (CI 64.2–88.3%)
DoC: dead of Covid-19; Cen: censored; DoD: dead of disease;		
2y: 2 years; 5y: 5 years; OS: overall survival; CI: 95% confidence intervals		

It should be noted that, independently on these survival rate estimates, the DoC imputations provided by CoDMI algorithm under option (iv.a) or (iv.b) provide, correspondingly, an adjusted sample including a total of 94 observations (the total number of original observations), where, however, only DoD and Cen events are now present. This is summarized in Table 4 for the two options. As shown in [5], the mean imputations provided by CoDMI are roughly unbiased. Therefore, all the usual statistical tools can be applied to these “standardized” samples and the information conveyed by DoC events is then consistently used. As an example, a standard proportional hazard model was applied to our data. Table 5 summarizes results from the Cox regression obtained according to the five possible DoC options. Several variables deemed to be relevant to overall survival (including clinical T stage, clinical N stage, tumor diameter and age) were included in the multivariate analysis. As expected, given the presence of only one DoC case in the sample, even in this application the differences between the estimates with the different options are almost immaterial. Obviously, in the estimation of the regression coefficients, options (ii) and (iv.b) provide exactly the same results.

Table 4
Details of exit causes after CoDMI imputation in standard form (option iv.a) and with adjustment for censoring (option iv.b)

CoDMI option	exit cause	number		min t (months)	max t (months)
option iv.a: DoC imputed as DoD	DoD	17	2		79.5
	Cen	77	11		82.0
	total	94			
option iv.b: DoC imputed as Cen	DoD	16	2		53.0
	Cen	78	11		82.0*
	total	94			
(*) Including 1 event at $t = 65.98$					

Table 5
Multivariate analysis of prognostic factors for overall survival according to DoC options

	Without DoC (option i)		DoC as Cen (option ii)		Doc as DoD (option iii)		DoC as virtual DoD (option iv.a)		DoC as virtual Cen (option iv.b)	
Prognostic factor	HR (95%CI)	p value	HR (95%CI)	p value	HR (95%CI)	p value	HR (95%CI)	p value	HR (95%CI)	p value
cT4 (no versus yes)	0.53 (0.16–1.77)	0.303	0.56 (0.17–1.84)	0.336	0.46 (0.14–1.51)	0.200	0.46 (0.14–1.50)	0.198	0.56 (0.17–1.84)	0.336
cN2 (no versus yes)	1.83 (0.55–6.02)	0.323	1.80 (0.54–5.95)	0.337	2.00 (0.61–6.59)	0.256	1.90 (0.58–6.28)	0.292	1.80 (0.54–5.95)	0.337
Lesion diameter > 5 cm (no versus yes)	2.90 (0.96–8.80)	0.060	2.80 (0.92–8.55)	0.069	2.98 (1.01–8.82)	0.048	3.00 (1.00-8.98)	0.049	2.80 (0.92–8.55)	0.069
Age > 70 (no versus yes)	3.58 (1.14–11.21)	0.028	3.76 (1.19–11.87)	0.024	3.42 (1.14–10.27)	0.028	3.67 (1.18–11.41)	0.025	3.76 (1.19–11.87)	0.024

DoC: dead of Covid-19; Cen: censored; DoD: dead of disease; HR: hazard ratio; CI: confidence interval

Discussion

The main clinical conclusion of this observational study is that the CoDMI algorithm can be easily applied for taking into account the information conveyed by DoC events in a sample of locally advanced rectal cancer patients with Covid-19 infection. The data adjustments provided by the algorithm through mean imputation allow for consistent Kaplan-Meier estimates of the survival curve and, more generally, produce a complete sample of survival observations which can be safely used for any further statistical analysis, as illustrated in Table 5. A possible alternative approach based on the theory of competing risks, would require a multidimensional point of view, characterized by more specific, and complex, statistical techniques.

The presence of just one DoC event in our study sample allows a better illustration of the effects of the CoDMI adjustments and a more direct comparison of these effects with those provided by simpler (and more naïve) options. Obviously, the usefulness of the algorithm is greater the greater the impact of Covid-19 deaths in the study sample and, correspondingly, the greater the potential loss of information caused by a poor treatment of this Covid-19 data.

This finding is new because there is no published studies on this topic. To further illustrate the clinical relevance of such a conclusion, direct applications of CoDMI algorithm to biggest Covid-19 death observations are welcome. The aim of this study was to simply describe a population descriptive (PO questions). Data were obtained from patients with rectal cancer who have been treated with concomitant chemoradiotherapy due to their locally advanced stage disease at diagnosis. Toxicity and treatment compliance details were deliberately not included in the present

analysis. Our first idea was to include OS data to strengthen CoDMI algorithm and thus overcome uncertainties in DoC classification. We aimed to demonstrate that our CoDMI algorithm represents an attractive and valid statistical tool to adequately perform survival analysis in the Covid-19 era. Therefore, we do not want to speculate on the relative efficacy of the treatment management, despite the homogeneity of the sample would allow to it. The main advantage of CoDMI algorithm – that combines Covid-19 and no-Covid-19 events over standard survival approaches – is to perform unbiased high-quality survival analysis in order to allow every possible data comparison. For sure, well-designed randomized clinical trials remain the reference to obtain level-one evidence. CoDMI algorithm should be used mostly to analyze data of these new trials presumably enrolling both Covid-19 positive and Covid-19 negative cases.

Conclusion

This is the first clinical application of the CoDMI algorithm in the radiation oncology scenario. We demonstrate that the CoDMI algorithm is a valid and easy to use statistical tool and it offers a satisfactory interpretation of Covid-19 events in survival analysis. Based on an homogenous cohort of patients, it provides appropriate data adjustments resulting in consistent Kaplan-Meier estimates of survival curves overcoming the direction of bias (over- or under-estimation). A user-friendly version of the CoDMI algorithm is freely available (https://github.com/alef-innovation/codmi) to facilitate and ensure accurate data analysis in the near future.

Declarations

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

Funding

The authors received no financial support for the research, authorship, and/or publication of this article

Aknowledgements

None

Author contribution

Franco Moriconi: Conceptualization; Data curation; Formal analysis; Methodology; Software; Writing – review & editing; final approval.

Francesca De Felice: Conceptualization; Data curation; Formal analysis; Writing - original draft; final approval.

Luca Mazzoni: Software; final approval.

Daniela Musio: Data acquisition; final approval.

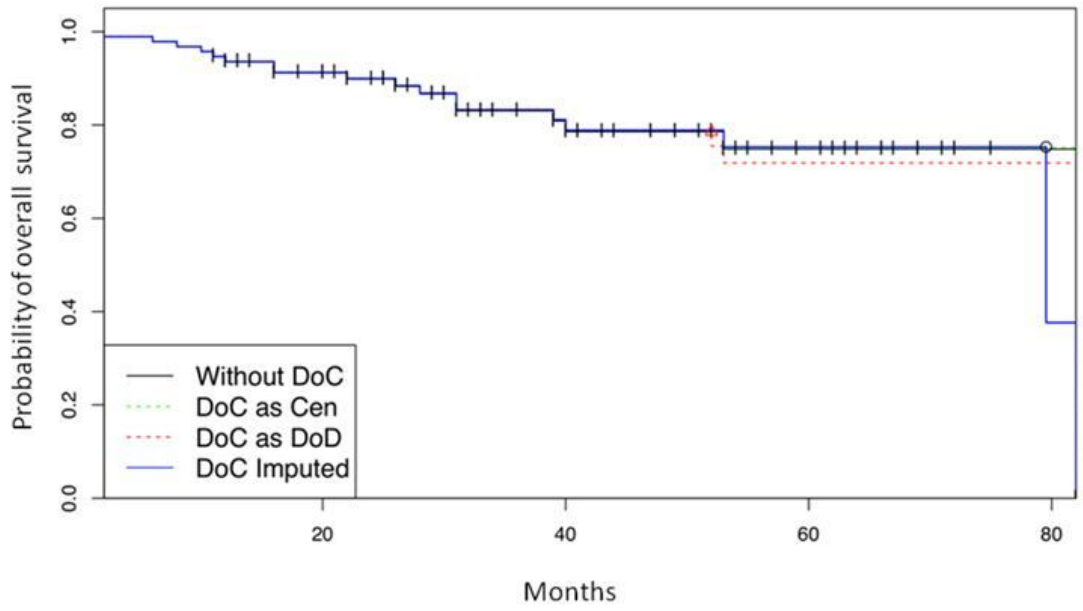
Vinncenzo Tombolini: Data acquisition; final approval.

References

1. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, Li C, Ai Q, Lu W, Liang H, Li S, He J. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. Lancet Oncol. 2020;21(3):335–337.
2. De Felice F, Moriconi F. COVID-19 and Cancer: Implications for Survival Analysis. Ann Surg Oncol. 2021;28(9):5446–5447.
3. DeSouza CM, Legedza AT, Sankoh AJ. An overview of practical approaches for handling missing data in clinical trials. J Biopharm Stat. 2009;19(6):1055-73.. (2016).
4. Efron, B, Hastie T. *Computer age Statistical Inference. Algorithms, Evidence, and Data. Science*. Cambridge University Press 2016.
5. De Felice F, Mazzoni L, Moriconi F. An expectation-maximization algorithm for including oncological covid-19 deaths in survival analysis. SSRN. 2022. Available at SSRN: http://ssrn.com/abstract=3998500 or http://dx.doi.org/10.2139/ssrn.3998500
6. De Felice F, D'Ambrosio G, Iafrate F, Gelibter A, Magliocca FM, Musio D, Caponetto S, Casella G, Clementi I, Picchetto A, Sirgiovani G, Parisi M, Orciuoli C, Torrese G, De Toma G, Tombolini V, Cortesi E. Intensified Total Neoadjuvant Therapy in Patients With Locally Advanced Rectal Cancer: A Phase II Trial. Clin Oncol (R Coll Radiol). 2021:S0936-6555(21)00226-0.
7. De Felice F, Musio D, Magnante AL, Bulzonetti N, Benevento I, Caiazzo R, et al. Disease control, survival, and toxicity outcome after intensified neoadjuvant chemoradiotherapy for locally advanced rectal cancer: a single-institution experience. Clin Colorectal Cancer 2016;15(2):e17ee22.

Figures

a. DoC event imputed as DoD.



b. DoC event imputed as Cen.

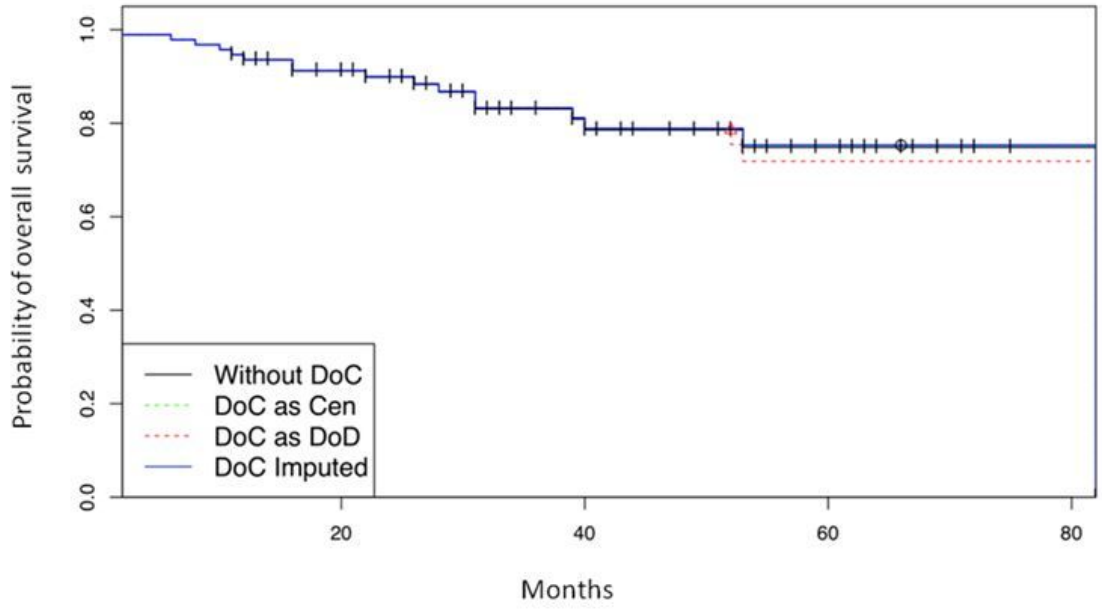


Figure 1

Overall survival Kaplan-Meier curves for alternative options on treatment of Covid-19 deaths.