



## COVID-19 and Cancer: Implications for Survival Analysis

Francesca De Felice<sup>1</sup>, and Franco Moriconi<sup>2</sup>

<sup>1</sup>Department of Radiotherapy, Policlinico Umberto I “Sapienza” University of Rome, Rome, Italy; <sup>2</sup>Department of Economics, University of Perugia, Perugia, Italy

Since the outbreak of the coronavirus disease 2019 (COVID-19) pandemic.<sup>1</sup> An important issue in survival data analysis has emerged—how to consider COVID-19 positive patients when the outcome event of interest occurs?

Independent from the anatomical extent of disease and its clinical stage classification at diagnosis, the main objective of an anticancer treatment strategy is to improve survival rates. Survival analysis is based on time to the binary event of interest, mainly death (no or yes). Traditionally, overall survival (OS), disease-free survival (DFS), and progression-free survival (PFS) are calculated in months from the date of diagnosis (or other time reference such as date of randomization and start of treatment) to the first event, including date of the last follow-up or death (OS) and/or relapse (DFS) and/or progression (PFS).

First described in 1958, Kaplan–Meier analysis is a univariate approach able to calculate the probability of each event at the time it occurs.<sup>2</sup> When patients are lost for any reason, they are considered ‘censored’. Censored cases are included in a probability estimation until the timepoint preceding their censoring, and eliminated thereafter. However, are censored observations independent from survival? This is crucial to legitimately support clinically relevant evidence-based data.

For instance, in a clinical trial in which cancer death is the primary endpoint, those patients who died due to COVID-19 respiratory disease should be (1) censored at the time of death because they are lost to follow-up

(*censored* option); or (2) included in the number of events (death) due to his/her advanced disease status (*event* option)? It is difficult to isolate the cause of death due to COVID-19 (dying from disease) or cancer progression (dying of disease). These misclassification/uncertainties in the cause of death allocations are essential to provide accurate and non-biased results when performing data analysis. Similarly, this issue is critical to define whether a new treatment strategy will potentially improve (censored option) or disprove (event option) survival time. As a result, the ability to judge the quality and authenticity of survival information could be biased due to COVID-19 infection.

A further, more subtle issue should be considered. We observed that the event of an oncological patient being COVID-19-positive materially increases his/her probability of death, which will be interpreted as the probability of being censored or dying of cancer, depending on whether the censored option or the event option is chosen. How could this information be properly taken into account in the survival analysis?

In addition, in the near future, the early development of national vaccination strategies may introduce additional bias. The COVID-19 pandemic should be considered a secular change (see, for example, Szklo and Nieto).<sup>3</sup> In fact, when the vaccine becomes available, it would not be appropriate to include all patients in the same cohort because this would mean combining heterogeneous cases, those accrued both before and after the introduction of the vaccine, and those who had the COVID-19 vaccination and those who had not. Again, as a result, the true survival rates will be biased.

We have briefly described the main COVID-19-related barriers in the process of conducting a reliable survival analysis. The hope is to capture the attention of the scientific community in the interpretation of these issues and to define an agreed modality on how COVID-19 deaths

should be classified in a survival analysis, in order to obtain comparable results. As a further, more ambitious goal, it should be important to agree on a standard approach on how the positivity information should be included in the survival model.

**DISCLOSURES** Francesca De Felice and Franco Moriconi have no disclosures to declare.

## REFERENCES

1. Our World in Data. Statis and research: coronavirus pandemic (COVID-19). <https://ourworldindata.org/coronavirus>
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53(282):457–81.
3. Szklo M, Nieto FJ. *Epidemiology—beyond the basics.* 4th edn. Burlington, MA: Jones & Bartlett Learning; 2019.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.