

Transcatheter aortic valve durability and the dark side of Kaplan-Meier analysis



Davide Capodanno*, MD, PhD, *Deputy Editor*

Ferrarotto Hospital, University of Catania, Catania, Italy

As interventional cardiologists, we deal with images every day. In fact, many of our daily decisions are driven by the first bias induced by a single picture, be it an angiographic projection or an intravascular ultrasound cross-section. To keep pace with the rapid evolution of our discipline, the busy doctor may be tempted to adopt the same approach to clinical research. In the era of “fast-food” data consumption, when a large number of results are disseminated through large-scale international meetings before the studies have even been peer reviewed never mind published, images try to capture the essence and sell the soul of a study.

Many readers will agree that one of the most hotly debated images presented at the last EuroPCR meeting in Paris was a Kaplan-Meier curve presented by Dvir and colleagues¹. During the “Late breaking trials, registries and innovation” session on 17 May 2016 in the Main Arena, the presenter showed a graph summarising transcatheter aortic valve implantation (TAVI) durability over an eight-year time horizon, from a multicentre experience. Up to five years, the profile of the curve for survival free from prosthesis degeneration looked similar to that of many other

Kaplan-Meier curves in the literature, with a gentle slope indicating acceptably few cases of valve deterioration. At one point in time, however, the curve had a steep collapse, resulting in more than half of the patients showing some form of valve deterioration at the end of follow-up. Some local bystanders (or worldwide spectators, as the image travelled fast through the web) reacted emotionally, as if this drop in the curve represented an allegory of the fall over the precipice for TAVI. Interestingly, depicting the concern over long-term durability issues of TAVI in the shape of a Kaplan-Meier curve was enough to make this message legitimate in the eyes of many.

Kaplan-Meier analysis is a popular method of summarising “time-to-event” outcome data within a study. This information is typically displayed in the form of a stepwise survival plot, where each step represents an event (or a cluster of events) at a certain point during the follow-up period. By looking at a glance, one may estimate from a Kaplan-Meier curve the fraction of subjects living free from an unfavourable outcome at different time points. Is this approach free from weaknesses and threats? There are

*Corresponding author: Cardio-Thoracic-Vascular Department, Ferrarotto Hospital, University of Catania, Via Citelli 6, 95100 Catania, Italy. E-mail: dcapodanno@gmail.com

a number of considerations that we should take into account when interpreting this kind of analysis, and tips to understand how much data conveyed by a Kaplan-Meier graph are robust and informative to our practice.

For example, look at the “number of patients at risk”. This is usually reported below the x-axis of the graph (and, if not, that is unfortunate). Ideally, in a clinical study, patients should be followed starting at the same time and have the same amount of follow-up. If so, the number of patients at risk at the end of the curve typically reflects the initial sample size minus patients who had the event of interest and therefore do not contribute to the curve anymore. More frequently, particularly in observational studies, patients are followed for survival starting at different times and therefore have varying lengths of follow-up. As a result, there is a progressive decrease over the course of time in the number of patients at risk who contribute to the curve. This phenomenon, called “censoring”, refers to mathematically removing a patient from the curve. When a patient is censored, the curve does not take a step down as happens when a patient has the event of interest.

Also, ideally, the curve should report marks along its trajectory to indicate where patients were censored, but for some reason these marks are rarely depicted. Otherwise, one may understand that the number of patients at risk is shrinking over time by looking at the 95% confidence interval of the curve, which is also a missing item in most of the Kaplan-Meier analyses presented or even published. That confidence interval indicates the degree of uncertainty over the survival estimate, and its widening reflects the notion that few patients actually have the final follow-up available. Therefore, most of the time there is nothing on the curve or around the curve to tell the viewer where (and when) a patient was censored, but in fact this can be indirectly derived by looking at the number of patients at risk (when available!).

In the Kaplan-Meier survival curve for TAVI durability mentioned above as an example, the number of patients at risk at eight years was only seven. At that point, survival free from valve degeneration was estimated at about 40%. Even so, that very low number of patients at risk indicates that some censoring was at play. In other words, adding to the proportion of patients who actually had valve degeneration, a high number of subjects in the study simply did not reach eight years of follow-up, due to late recruitment or to death exerting a competing effect in a typically old and frail population (i.e., the TAVI prosthesis has no time to deteriorate, because the patient dies earlier). Because censoring a patient reduces the number of those at risk who are actively contributing to the curve, each event after that point represents a higher proportion of the remaining population, and therefore every drop in the curve afterwards will be a little bit larger than it would have been with less censoring. One way to avoid disseminating overestimation

of clinically relevant events is to cut the Kaplan-Meier curve at a point in time where follow-up is consistent in most patients or, simply ... to wait and collect more follow-up data.

While these methodological reflections would be enough to detract from overinterpreting the TAVI durability curve mentioned in our example, there are other issues to consider. Firstly, valve degeneration cannot be considered a true clinical event, but rather a time-dependent process. As a matter of fact, many patients may not report symptoms and are therefore not referred to echocardiography. Conversely, for such a type of endpoint to be credible, echocardiographic assessment should be performed in all living patients at pre-specified time points rather than occasionally or because of symptom development, or the single events will have an exaggerated impact on the shape of the curve. Of course, this kind of ascertainment bias may act in both directions, because it is also possible that echocardiography could not be performed because valve degeneration was present and actually caused worsening of heart failure and death. Secondly, the definition of valve degeneration should be pragmatic and consistent, with an established link to hard clinical outcomes. Finally, methods other than the Kaplan-Meier analysis (e.g., Fine-Gray model) should be considered to address the problem of competing risk due to frequent deaths.

In general, and outside the context of the above example, it is the responsibility of investigators to choose the most appropriate strategy for the analysis of a set of data, preventing unsupported conclusions. In parallel, it is the responsibility of editors and reviewers to avoid publication of analyses that may be methodologically unsound. Expert consensus documents can contribute to better reporting by standardising definitions and strategies for analysing specific outcomes of interest. Finally, we, as viewers, should remember that the complexity of biological phenomena cannot be reduced to the perception induced by a single image from an abstract presentation. We should recognise that the results of any study, be it randomised or observational, are only as good as the methods.

Conflict of interest statement

The author has no conflicts of interest to declare.

Reference

1. Dvir D, Eltchaninoff H, Ye J, Kan A, Durand E, Bizios A, Cheung A, Aziz M, Simonato M, Tron C, Arbel Y, Moss R, Leipsic J, Ofek H, Perlman G, Barbanti M, Seidman M.A, Blanke P, Yao R, Boone R, Lauck S, Lichtenstein S, Wood D, Cribier A, Webb J. Accessed on July 28, 2016 at <http://www.pcronline.com/Lectures/2016/First-look-at-long-term-durability-of-transcatheter-heart-valves-assessment-of-valve-function-up-to-10-years-after-implantation>