



ICB digital

noviembre 2022

Boletín para los Comités de Ética de Investigación

contenido nº135

artículo breve

Como sobrevivir a la interpretación de los diseños de supervivencia más comunes en los ensayos clínicos: pautas y soluciones

Ferran Torres, Víctor Sapena, José Ríos

adendum

Adendum icb septiembre 2022

Amelia García-Luque, Ruth María Aparicio-Hernández

próximo número

nº 136 - enero 2023

© SOCIEDAD ESPAÑOLA DE FARMACOLOGÍA CLÍNICA

administracion@se-fc.org

www.se-fc.org/icbdigital

Como sobrevivir a la interpretación de los diseños de supervivencia más comunes en los ensayos clínicos: pautas y soluciones

Ferran Torres, Víctor Sapena, José Ríos

Unidad Bioestadística, Facultad de Medicina, Universitat Autònoma de Barcelona

Introducción

En muchos ensayos clínicos (EE.CC.) se evalúa el tiempo hasta un evento o un resultado de interés (puede ser un evento positivo también) en alguna variable principal o secundaria. A pesar que la metodología subyacente de los métodos más comúnmente empleados está ampliamente difundida, su uso e interpretación ha sido problemática en varias revisiones¹⁻³.

El objetivo de este artículo es identificar algunos de los problemas más comunes de los análisis de supervivencia en el campo de los EE.CC, y dar algunas pautas útiles para evitarlos y solucionarlos con alternativas disponibles. Una descripción detallada de los métodos de supervivencia más comunes está fuera de este objetivo, para ello se recomienda revisar algunos trabajos enfocados a lectores no estadísticos⁴⁻⁶. De manera análoga, el uso de otras técnicas menos habituales pueden consultarse en otras fuentes más especializadas⁷⁻⁹.

Descripción de la función de supervivencia: método de Kaplan-Meier

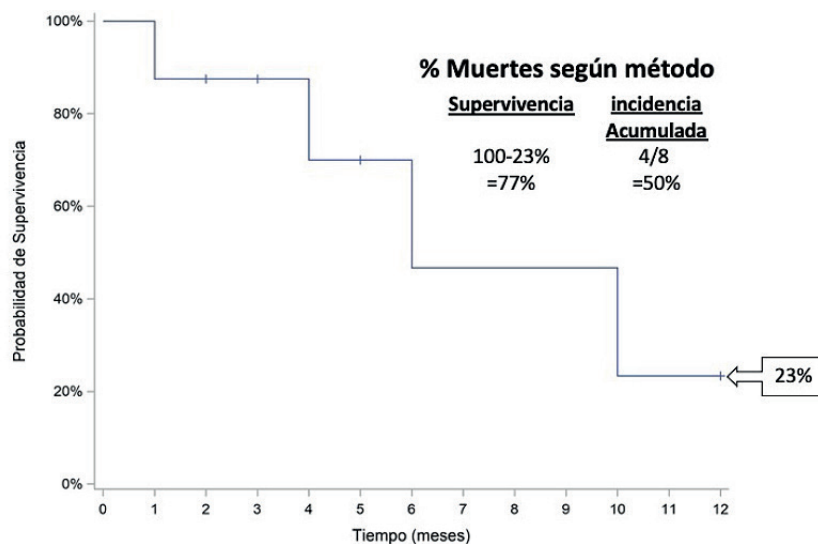
La función de supervivencia se acostumbra a describir utilizando el método de Kaplan-Meier (KM)¹⁰, la cual permite un manejo adecuado de los datos censurados. En los EE.CC, las censuras se identifican como los pacientes sin evento al final de su seguimiento. La curva de KM muestra el tiempo específico de aparición de eventos y censuras, a diferencia del método actuarial que los muestra agrupados por intervalos predefinidos. Una asimetría en la distribución temporal de las censuras entre grupos y/o la violación de que los pacientes en seguimiento deberían tener la misma probabilidad de tener eventos que los censurados, comprometería totalmente la validez de los resultados. A partir de KM se puede estimar la función de supervivencia en un punto temporal determinado (MA, del inglés *milestone analysis*), así como el tiempo en el que el 50% de los pacientes disponibles a riesgo han tenido el evento de interés (o mediana)¹¹.

En general, uno de los problemas principales de la interpretación de las curvas de supervivencia estriba en dar excesivo valor a un único punto de la curva en lugar de evaluar todo el recorrido temporal de la misma. En este sentido es destacable que tanto el MA como la mediana se interpretan habitualmente de manera errónea, tal y como se discute a continuación¹².

Los valores del MA se interpretan ingenuamente como porcentajes de pacientes sin el evento de interés (sobre el total de pacientes), es decir, como si en realidad fuera directamente el valor inverso de una incidencia acumulada (IA). Esto no es así, y la explicación reside en que KM describe la probabilidad de sobrevivir en un momento dado condicionado a estar vivo en el punto temporal anterior. Así, el denominador va descendiendo a lo largo del tiempo debido a eventos y censuras previos. De este modo, los valores descritos como porcentaje de eventos a partir de una curva de supervivencia serán superiores que los que se obtendrían utilizando una IA sobre el total de pacientes que iniciaron

el estudio. En la Figura 1 se representa un ejemplo de mortalidad a los 12 meses en el que, de un total de 8 sujetos, 4 presentan el evento y 4 censuras. La diferencia entre los dos métodos es más que obvia: 77% vs. 50%.

Figura 1: Comparativa de mortalidad estimada según función de supervivencia e incidencia acumulada



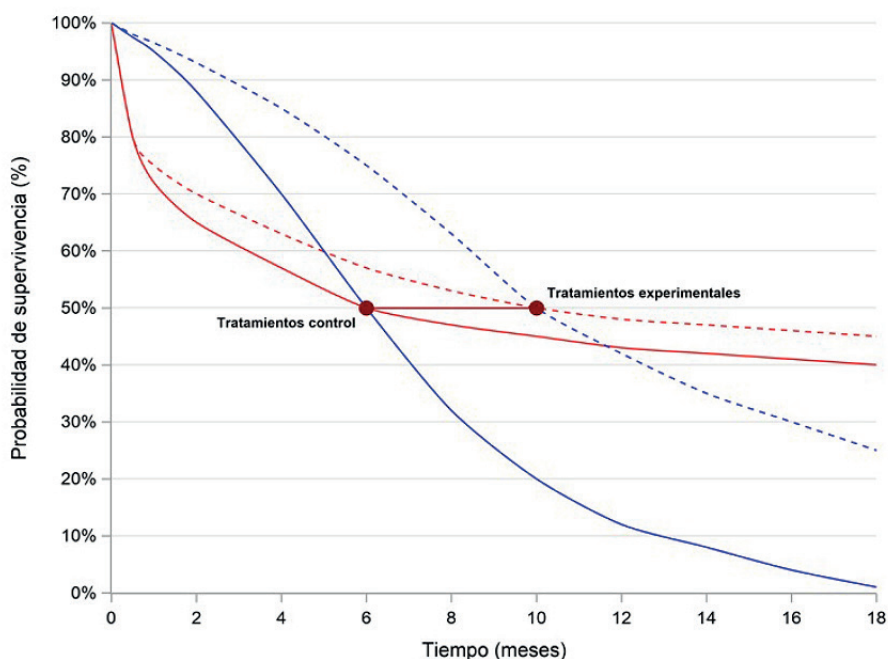
Las medianas se usan de manera habitual para describir el efecto de los tratamientos. Éstas representan el tiempo en el que los pacientes de un grupo llegan al punto medio del eje vertical de la curva de supervivencia. La diferencia de medianas se interpreta como la ganancia en tiempo hasta evento al administrar el tratamiento experimental en lugar del tratamiento control. De hecho, es también relativamente habitual comparar efectos de distintos tratamientos incluso con datos de otros ensayos. En ocasiones, se ha llegado a esgrimir el hecho de que un grupo no llegase a la mediana de la supervivencia como falta de madurez y/o invalidez del análisis. Todas estas interpretaciones son incorrectas y a continuación se justifica con una serie de argumentos.

En primer lugar, al igual que en el caso de MA, la mediana representa un único punto de la curva, por lo que no es correcto interpretarla como un efecto global o una ganancia resumen de todo el trayecto temporal. Además, puede darse el caso que la mediana no sea calculable porque la curva no descienda lo suficiente, o también podría darse el caso que no representara un punto relevante desde el punto de vista clínico. Otro argumento es que se trata simplemente de un estadístico descriptivo, que no tiene en cuenta factores de ajuste o estratificaciones comúnmente usados en los EE.CC. Esto significa que las medianas y sus diferencias están completamente dissociadas del análisis inferencial, que sí que los tiene en cuenta.

Por otra parte, el valor de las medianas está claramente determinado por la forma de las curvas, y en particular por su ritmo de descenso. Las inconsistencias entre la interpretación del efecto del tratamiento usando solo las medianas, o toda la curva, son frecuentes. En el siguiente ejemplo hipotético (ver Figura 2) se comparan tratamientos experimentales (líneas punteadas) contra controles (líneas sólidas) en dos EE.CC. distintos, uno en el que se identifican los grupos en color rojo y otro en color azul. De manera evidente, la forma de las curvas permitiría concluir que el efecto protector del tratamiento experimental contra su control es mayor en el ensayo 'azul' que en el ensayo 'rojo'. Por el contrario, los valores de la mediana son idénticos en los dos tratamientos controles y también en los dos tratamientos

experimentales, es decir el beneficio en diferencia de medianas sería el mismo. De manera análoga, estas inconsistencias también pueden suceder de manera parecida al evaluar el MA, fruto de utilizar un solo punto temporal de las curvas.

Figura 2: Discrepancias al evaluar el trayecto completo de la curva de KM vs solamente un único punto en la mediana



El uso de las medianas para comparar tratamientos de estudios distintos es todavía más incorrecto y sujeto a más sesgos, sobre todo si no se puede garantizar la misma intensidad en los procedimientos diagnósticos, si los mismos tiempos de evaluación de los eventos son distintos.

Métodos más habituales para la inferencia estadística en los análisis de supervivencia: prueba de logrank y regresión de Cox

Los EE.CC. utilizan como técnicas inferenciales más habituales la prueba (estratificada o no, según proceda) de logrank (o log-rank)¹³ para el cálculo del valor de p, y la regresión de riesgos proporcionales (RP) de Cox¹⁴ para estimar la magnitud del efecto del tratamiento. La prueba de logrank realiza un análisis acumulado a lo largo del tiempo comparando las curvas respecto a los valores de eventos observados contra los esperados si no hubiera relación. Es una prueba relativamente robusta, y por eso muchas veces se tiende a preferirla a la regresión de Cox como criterio de decisión inferencial. Sin embargo, puede tener un rendimiento inferior al habitual en ciertas situaciones, como por ejemplo cuando las curvas tardan en empezar a separarse, cuando las diferencias se van atenuando con el tiempo o bien cuando se cruzan las curvas⁷.

La regresión de Cox es un modelo semi-paramétrico, no asume ninguna distribución de las curvas, pero requiere estrictamente del cumplimiento de los riesgos proporcionales (RP). Se habla de RP cuando la razón de riesgo permanece constante durante todo el seguimiento. Este modelo estima un resumen del efecto del tratamiento a lo largo de toda la escala temporal en un único número: la razón de riesgos comúnmente conocida como *Hazard Ratio* (HR). La asunción de los RP es el punto crítico que puede condicionar la validez de la inferencia y de las estimaciones del riesgo¹⁵. Adicionalmente a la dificultad del cumplimiento de los RP, en muchas ocasiones el efecto del beneficio de los HR se interpreta de manera errónea, como una ganancia (protección) en el eje horizontal que representa el tiempo. Sin embargo, lo

que realmente se evalúa es la comparación de las curvas según el eje vertical. A modo de ejemplo, un HR de 0,60 no indica que los pacientes del grupo experimental vivan un 40% más, sino que tienen un 40% menos de probabilidades de sufrir el evento estudiado a lo largo del seguimiento respecto al grupo control.

Métodos que evalúan el beneficio en escala de tiempo

Métodos basados en la aceleración del tiempo de fallo

Se dispone desde hace tiempo de algunas técnicas alternativas que, aunque menos difundidas, pueden ser perfectamente válidas y aplicables en muchos casos en los que las técnicas más conocidas son inadecuadas o subóptimas¹¹. Si lo que se pretende es valorar la ganancia en tiempo, se podrían utilizar modelos paramétricos denominados aceleración del tiempo de fallo (AFT, del inglés *accelerated failure time*) que estiman ratios de tiempo (TR, del inglés *time ratio*), en lugar de los HR¹¹. El TR se interpreta como una relación de tiempo o velocidad en la que un nuevo tratamiento ayuda a retrasar la aparición de eventos.

Estos métodos, aunque no tienen la restricción de los RP, tienen también una limitación clave: se debe realizar la selección apropiada de la distribución estadística de entre varias posibles. Esta decisión, si se determina *post-hoc* basada en los datos observados, es muy cuestionable metodológicamente, al menos en estudios confirmatorios. Del mismo modo, si se basa en asunciones a partir de datos previos, los datos del estudio podrían diferir de las asunciones iniciales y, por lo tanto, cuestionar la selección hecha *a priori* y los resultados consecuentes⁵. Por lo tanto, solo se recomienda su uso si no se espera que ocurran esas limitaciones.

Métodos basados en las medias de tiempo restringido

Las limitaciones anteriores pueden solventarse utilizando los métodos basados en el tiempo de supervivencia medio restringido (RMST, del inglés *restricted mean survival time*). El RMST es el tiempo promedio transcurrido sin un evento hasta un horizonte de tiempo preespecificado denominado *tau*. Los métodos basados en el RMST no asumen ninguna distribución de los datos, y se puede calcular usando el área bajo la curva de supervivencia. Además, permiten la modelización de covariables y, como tiene en cuenta toda la información de supervivencia antes de *tau*, evalúan realmente la ganancia de tiempo hasta el evento de forma global. La comparativa entre grupos se puede realizar tanto en una escala aditiva de diferencias de promedios restringidos o de la razón (*ratio*) de los mismos^{16,17}. La principal limitación es la predefinición del tiempo *tau*, pero en los EE.CC. no debería ser un problema predefinir ese punto de corte bajo un criterio de relevancia clínica razonable.

Reconstrucción de curvas de supervivencia

Actualmente, existe metodología reconocida y ampliamente difundida que permite reconstruir numéricamente las curvas de supervivencia a partir de los gráficos de KM¹⁸. De esta forma, los resultados originales pueden ser reevaluados y, en su caso, comparados con otras técnicas de análisis que no violen los requisitos de la aplicación, o bien que evalúen la ganancia de escala temporal que la mayoría de los análisis originales no permitieron.

Conclusiones

Los resultados de los EE.CC. que utilizan técnicas de supervivencia se interpretan de manera errónea en bastantes ocasiones. La utilización de los métodos descriptivos de la función de supervivencia debería limitarse exclusivamente a ese uso, y sobre todo se debería evitar la extrapolación y comparación entre estudios de resultados a partir de un solo punto de la curva. Para la inferencia y estimación del efecto de tratamientos, los métodos más habituales como la prueba de logrank y la regresión de riesgos pro-

porcionales Cox siguen siendo válidos mientras se puedan garantizar las asunciones de aplicabilidad, y en caso contrario debería recurrirse a otras alternativas. Si el objetivo de un estudio es evaluar la ganancia en tiempo, dadas las limitaciones de los métodos *accelerated failure time* parece recomendable decantarse por los métodos *restricted mean survival time*, que además no requiere el cumplimiento de estrictas asunciones de aplicabilidad.

REFERENCIAS

1. Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72(2):511–8.
2. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: Good practice and pitfalls. *Lancet*. 2002;359(9318):1686–9.
3. Abaira V, Muriel A, Emparanza JI, Pijoan JI, Royuela A, Plana MN, et al. Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirements proposal. *J Clin Epidemiol*. 2013;66(12):1340–1346.e5.
4. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival Analysis Part I: Basic concepts and first analyses. *Br J Cancer*. 2003;89(2):232–8.
5. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part II: Multivariate data analysis- An introduction to concepts and methods. *Br J Cancer*. 2003;89(3):431–6.
6. Delgado J, Pereira A, Villamor N, López-Guillermo A, Rozman C. Survival analysis in hematologic malignancies: Recommendations for clinicians. *Haematologica*. 2014;99(9):1410–20.
7. O'Quigley J. Testing for Differences in Survival When Treatment Effects Are Persistent, Decaying, or Delayed. *J Clin Oncol*. 2022;1:1–10.
8. Lin RS, Lin J, Roychoudhury S, Anderson KM, Hu T, Huang B, et al. Alternative Analysis Methods for Time to Event Endpoints Under Nonproportional Hazards: A Comparative Analysis. *Stat Biopharm Res* [Internet]. 2020;12(2):187–98. Available from: <http://dx.doi.org/10.1080/19466315.2019.1697738>.
9. Ristl R, Ballarini NM, Götte H, Schüller A, Posch M, König F. Delayed treatment effects, treatment switching and heterogeneous patient populations: How to design and analyze RCTs in oncology. *Pharm Stat*. 2021;20(1):129–45.
10. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* [Internet]. 1958 Jun;53(282):457. Available from: <https://www.jstor.org/stable/2281868?origin=crossref>.
11. Gregson J, Sharples L, Stone GW, Burman C-F, Öhrn F, Pocock S. Nonproportional Hazards for Time-to-Event Outcomes in Clinical Trials: JACC Review Topic of the Week. *J Am Coll Cardiol* [Internet]. 2019;74(16):2102–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31623769>.
12. European Medicines Agency. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man. Methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials. EMA/CHMP/27994/2008/Rev.1 [Internet]. 2013 [cited 2018 Jan 8]. p. 1–8. Available from: <https://www.ema.europa.eu/en/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using>
13. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard S V, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br J Cancer* [Internet]. 1977 Jan;35(1):1–39. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/831755>
14. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B* [Internet]. 1972;34(2):187–220. Available from: <http://www.jstor.org/stable/2985181>
15. Stensrud MJ, Hernán MA. Why Test for Proportional Hazards? *JAMA* [Internet]. 2020 Apr 14;323(14):1401–2. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32167523>
16. Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* [Internet]. 2013 Dec 7;13:152. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24314264>
17. Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol*. 2014;32(22):2380–5.
18. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol* [Internet]. 2012 Feb 1;12:9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22297116>.

Adendum icb septiembre 2022

Estimado Comité editorial,

Las autoras del artículo publicado en el número anterior, ICB nº 134, de título “*Generación de células madre pluripotentes inducidas (iPSC): a propósito de un proyecto de investigación presentado al CEIm del Hospital Central de la Defensa*”, queremos actualizar la información sobre el **informe favorable preceptivo de la “Comisión de Garantías para la Donación y Utilización de Células y Tejidos Humanos” en determinadas investigaciones**, al haberse modificado el artículo 35 de la Ley de Investigación Biomédica, con la publicación y entrada en vigor de la *Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011 de 1 de junio, de la Ciencia, la Tecnología y la Innovación*, en el mes que se publicó este artículo breve del ICB.

La Ley 17/2022 en la Disposición final primera, modifica la Ley 14/2007, de 3 de julio, de Investigación Biomédica en los siguientes términos:

Uno. Se añade un nuevo apartado 3 al artículo 35, con la siguiente redacción:

«3. No obstante lo dispuesto en los apartados anteriores, no precisarán informe previo de la Comisión los proyectos de investigación con células humanas pluripotentes obtenidas mediante reprogramación celular cuando se trate de:

- a) Su generación mediante la utilización de células somáticas no embrionarias procedentes de donantes sanos o enfermos y su caracterización, incluyendo la formación de teratomas.
- b) Su generación a partir de células registradas, de origen humano, no embrionarias, procedentes de biobanco o de origen comercial.
- c) Su utilización o la de sus derivados, procedentes de biobancos o de investigadores/as en otros países para la investigación *in vitro*, salvo las que tengan por objeto fines reproductivos, producción de gametos y formación de quimeras primarias interespecies.

La autorización de estos proyectos corresponderá al Comité de Ética de la Investigación correspondiente, que podrá elevar consulta a la Comisión de Garantías para la Donación y la Utilización de Células y Tejidos Humanos de estimarlo necesario.

Amelia García-Luque, Ruth María Aparicio-Hernández

Secretaría técnica del CEIm Hospital Central de la Defensa

Servicio de Farmacología Clínica Hospital Central de la Defensa

Entidades patrocinadoras y colaboradoras de ICB digital



Lilly



Menarini



Novartis



Sanofi Aventis