

Vivir para contarlo

Dres. Víctor Dayan¹, Abayubá Perna^{2,3}

Durante la Segunda Guerra Mundial, el servicio militar de Estados Unidos solicitó la ayuda de un grupo de estadísticos para analizar los datos recabados durante el conflicto y asesorar sobre la mejor estrategia para reforzar sus aviones⁽¹⁾. Dentro de este grupo asesor, se encontraba el Dr. Abraham Wald, reconocido estadístico austriaco. Al examinar los aviones que retornaban de combate, Wald notó que los principales sitios de impacto de las balas eran en alas, cola, nariz y fuselaje (figura 1). La conclusión que se desprendía de forma natural era que el refuerzo en estos aviones debería concentrarse en estos sitios, que recibían la mayor cantidad de impactos. Sin embargo, esa conclusión iba en sentido opuesto a la creencia de Wald. El método para definir los sectores a reforzar en estos aviones, conllevaba la limitación de usar solo los datos provenientes de aquellos que lograron retornar. Lo que pudieron concluir, era que los aviones que recibían impactos de bala en nariz, alas, fuselaje y cola lograban retornar a casa. Por lo tanto, el refuerzo debía colocarse en aquellas zonas en las que no había impactos de bala (como en el motor). A partir de ese momento, se plasma un concepto clave en la vida diaria y en el análisis crítico de trabajos científicos, el sesgo de supervivencia.

En forma simple, el sesgo de supervivencia jerarquiza el error que cometemos en concentrarnos en los resultados de aquellos que “lo lograron” (o sobrevivieron), sin considerar todos aquellos que “lo intentaron”. La sección de autoayuda de las librerías representa un fiel ejemplo de sesgo de supervivencia. En ella encontramos libros escritos por gente exitosa que nos comparte su receta. Al leerla, vemos que incluye ingredientes básicos (autosuperación, inteligencia emocional, trabajo en equipo, liderazgo, entre otros) propuestos por el autor como claves para su éxito, sin decirnos que la misma es-

trategia fue usada por miles de personas sin el resultado esperado.

¿Cómo aplica esto en la vida diaria? Supongamos que usted es el dueño de una librería que tiende a recibir muchos clientes regulares. Estas son las personas a quienes les gustó tanto su lugar que tienden a repetir la experiencia. Es tan importante conocer qué es lo que hace que estos clientes vuelvan, como saber por qué aquellos que vinieron una vez, no volvieron más. Si nuestra estrategia de expansión y *marketing* se basara solo en lo que dicen los clientes que repiten, estaríamos cometiendo el mismo error que Wald señaló a propósito de los aviones durante la Segunda Guerra Mundial. Para los profesionales que se dedican a *marketing* y administración de empresas este concepto es algo de todos los días. Sin embargo, para nosotros, los médicos, su presencia puede pasar desapercibida y conducirnos a validar conclusiones sesgadas que nos lleven a realizar o indicar procedimientos terapéuticos vacíos de beneficio clínico.

El sesgo de supervivencia adquiere su mayor relevancia en dos situaciones: comparación de resultados funcionales y comparación de supervivencia.

En lo que respecta a los resultados funcionales (subjetivos u objetivos), este sesgo aparece cuando se compara una terapéutica intervencionista con el tratamiento médico o placebo. Por ejemplo, supongamos que comparamos la presencia de ángor en pacientes con coronariopatía sometidos a tratamiento quirúrgico versus tratamiento médico óptimo en un ensayo clínico randomizado (ECR). Al ser un ECR, ambos grupos estarán balanceados en cuanto al porcentaje de pacientes de alto riesgo. La mortalidad a corto plazo será mayor en los pacientes sometidos a tratamiento quirúrgico (debido a la mortalidad operatoria). Sin embargo, a mediano plazo la presencia de ángor es significativamente menor en los pacientes sometidos a cirugía. Descar-

1. Centro Cardiovascular Universitario, Universidad de la República. Montevideo, Uruguay.

2. Hospital de Clínicas, Universidad de la República. Montevideo, Uruguay.

3. Unidad de evaluación, Fondo Nacional de Recursos. Montevideo, Uruguay.

Correspondencia: Dr. Víctor Dayan: Correo electrónico: victor_dayan@hotmail.com

Los autores declaran no tener conflictos de intereses.

Recibido Set 8, 2020; aceptado Oct 10, 2020

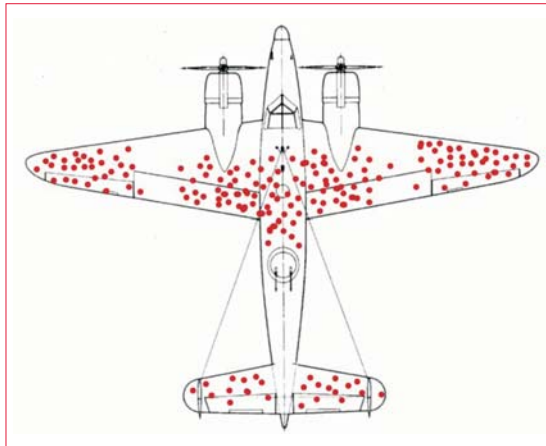


Figura 1. Ilustración de un avión de la Segunda Guerra Mundial. Los puntos rojos representan los sitios de impacto de las aeronaves que retornaron a la base. Tomado de https://en.wikipedia.org/wiki/Survivorship_bias

tando el posible efecto placebo, uno podría decir que algunos de los pacientes más graves (que están en igual proporción en ambos grupos) del grupo quirúrgico fallecieron (lo que explica la mayor mortalidad a corto plazo), permaneciendo aquellos con mejor estado funcional en el seguimiento a mediano plazo. Esta mortalidad temprana generalmente no ocurre en el grupo control bajo tratamiento médico óptimo, sobreviviendo entonces los pacientes de mayor riesgo para su evaluación funcional a mediano plazo. En otras palabras, se introduce un efecto de “selección natural” por el cual los pacientes más débiles de una de las ramas de estudio fallecen tempranamente dejando un grupo más apto para ser evaluado a mediano plazo.

La otra situación en la que el sesgo de supervivencia toma importancia, es cuando se evalúa un objetivo según tiempo al evento (sobrevida) en presencia de riesgos competitivos (RC). Los RC son aquellos eventos que cuando ocurren excluyen la posibilidad de que el sujeto de estudio tenga el evento de interés. Por ejemplo, si nuestro evento de interés es el tiempo a la ocurrencia de muerte cardiovascular, infarto de miocardio o ataque cerebrovascular en un ECR, aquellos pacientes que mueran de causa no cardiovascular, no tendrán la posibilidad de presentar el evento de interés. En este caso, la muerte no cardiovascular es un RC. El no considerar los RC al evaluar los resultados de supervivencia de un ECR, puede llevar a conclusiones erróneas. De hecho, casi 50% de los estudios que incluyen análisis de supervivencia por Kaplan-Meier se encuentran sesgados por RC y en un tercio de ellos, el riesgo de eventos está sobrestimado significativamente⁽²⁾. A continuación veremos como no considerar estos RC puede subestimar el riesgo de supervivencia.

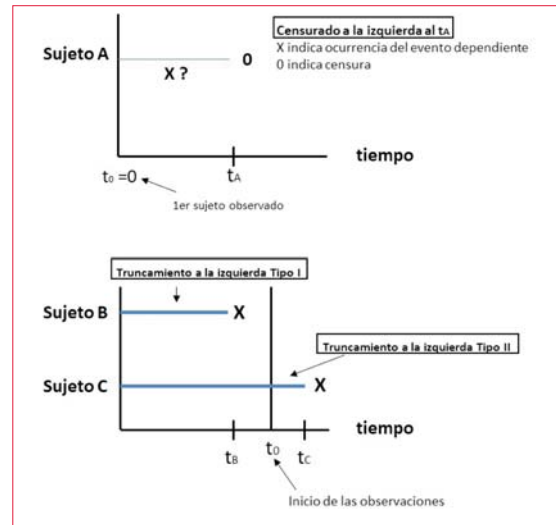


Figura 2. Truncamiento y censura a la izquierda. La existencia de censura a la izquierda y de truncamiento a la izquierda tipo II nos lleva a una estimación incorrecta del tiempo de exposición, mientras que una situación frecuente de truncamiento a la izquierda tipo I afecta la representatividad de la muestra (muestra sesgada). Tomado de Kleinbaum D, Klein Mitchel (2012, Third Edition) *The Cox Proportional Hazards Model and Its Characteristics*. In: *Survival Analysis: A Self-Learning Text (Statistics for Biology and Health)*. USA, Springer⁽⁵⁾.

Miyasaka y colaboradores⁽³⁾ realizaron un estudio poblacional acerca de la incidencia de demencia desde el diagnóstico de fibrilación auricular. Dicho estudio fue realizado según tiempo al evento y la presencia de muerte fue censurada. Como resultado, la incidencia de demencia a los cinco años estimada fue de 10,5%. Al censurar los pacientes que murieron (y no tuvieron demencia), se asume que estos pacientes nunca podrían haber tenido el evento de interés. Lo cual es correcto por un lado (el que muere no puede desarrollar demencia), pero, por otro, es asumir que si estuviera vivo no hubiera tenido demencia. Como resultado, el análisis de supervivencia por Kaplan-Meier sobrestima el riesgo final del evento, o, a la inversa, subestima la supervivencia libre del evento⁽⁴⁾.

Antes de profundizar en el concepto de censura es importante señalar que el ejemplo anterior nos permite plantear dos conceptos cruciales e íntimamente relacionados que hay que tener presentes en el análisis de supervivencia y que están vinculados al “factor de exposición”, en este caso ser portador de fibrilación auricular; a) la representatividad de la muestra, b) el tiempo de exposición, que operativamente es la diferencia entre la fecha en que el individuo comenzó con la exposición (fecha de diagnóstico de la fibrilación auricular) y tuvo el evento dependiente (fecha de diagnóstico de demencia), o

Tabla 1. Sobrevida libre de muerte cardiovascular en una población de 100 pacientes.

<i>Año</i>	<i>Número</i>	<i>Evento CV</i>	<i>Censura</i>	<i>Probabilidad</i>	<i>Probabilidad acumulada</i>
1	100	3	0	97/100	0,97
2	94	5	3	89/94	0,92
3	84	4	5	80/84	0,88
4	70	3	10	67/70	0,84
5	62	6	5	56/62	0,76

CV: cardiovascular.

en el caso que no haya desarrollado demencia, la fecha de últimas novedades (censura a la derecha).

La metodología cuantitativa (inferencial) tiene un objetivo básico: extrapolar los resultados de una muestra a toda la población. Para que ello sea posible, la muestra (n) de pacientes con fibrilación auricular debería representar a toda la población (N) de portadores de fibrilación auricular. Por ejemplo, si un número importante de pacientes de la población (N) con fibrilación auricular (pero sin un correcto diagnóstico de esta condición) hubiera fallecido antes o al momento de seleccionar la muestra, se estaría atentando contra la "buena" representatividad de la misma. Esto se denomina "truncamiento a la izquierda tipo 1". Por otro lado, en aquellos individuos en que se hizo diagnóstico de fibrilación auricular, un buen investigador se preguntaría ¿estamos estimando correctamente desde cuándo estos individuos tienen esta condición? Por ejemplo, a un individuo teórico (individuo C en la figura 2) le diagnosticamos la fibrilación auricular en diciembre, pero en realidad la desarrolló en marzo: ese individuo tendrá en diciembre un "tiempo de exposición real de nueve meses (que en un estudio real nosotros desconocemos), pero lo vamos a considerar como tiempo "cero" (cuando lo diagnosticamos con fibrilación auricular): esto es "truncamiento a la izquierda tipo 2" (figura 2).

Este ejemplo nos presenta un problema adicional vinculado a la variable dependiente (en este caso tener una demencia): la demencia es una entidad de diagnóstico fundamentalmente clínico (de confirmación anatomopatológica) cuya causa más frecuente es un proceso neurodegenerativo subyacente (existen también las causas vasculares); el proceso neurodegenerativo tiene un período de evolución subclínico, de duración desconocida hasta que se manifiestan los síntomas, que deben alcanzar determinada entidad para que permitan un diagnóstico correcto. Este período puede durar meses o años. Esto quiere decir que en aquellos individuos que en

nuestro estudio clasifiquemos como portadores de demencia, en realidad desconocemos cuándo comenzaron a desarrollarla (en un sentido tal vez exagerado un individuo podría incluso haber comenzado a desarrollarla antes de tener la fibrilación auricular). Analizado desde este punto de vista (o sea adaptar la herramienta estadística al fenómeno biológico que queremos estudiar) y apoyándonos en el gráfico superior de la figura 2, podemos reconocer que nuestra estimación del tiempo de exposición en el caso de un individuo con demencia sería incorrecto: sabemos que el individuo tuvo el evento pero no sabemos exactamente cuándo lo tuvo (figura 2). En esta situación, cabría preguntarse entonces si el individuo no está "censurado a la izquierda", situación que al menos en parte, se podría corregir retrospectivamente considerando una fecha (aproximada) de cuándo comenzaron sus primeros síntomas de deterioro cognitivo. Censura a la izquierda es por lo tanto una situación poco frecuente y se define como aquella en la que el individuo luego de ingresar en el estudio tuvo el evento dependiente pero no sabemos cuándo.

Mucho más frecuente es la "censura a la derecha" durante los estudios de sobrevida. En ellos, un paciente censurado a la derecha es aquel que no tuvo el evento de interés luego del período de observación o que abandonó el estudio. Para la correcta interpretación de los datos arrojados por el análisis de sobrevida mediante Kaplan-Meier, la censura debe ser independiente del evento de interés (por ejemplo: abandona el país por temas laborales y no puede continuar en el estudio). En un trabajo que evalúa la recurrencia de cáncer de mama, un paciente que muere por infarto agudo de miocardio es censurado. Sin embargo, no podemos afirmar que la misma no tenga relación al evento de interés, por lo cual la censura no es independiente y se transforma en un RC.

Veamos un ejemplo para ilustrar el efecto que tiene la censura de pacientes durante el análisis de

Tabla 2. Sobrevida libre de muerte cardiovascular sin eventos censurados en una población de 100 pacientes.

<i>Año</i>	<i>Número</i>	<i>Evento CV</i>	<i>Censura</i>	<i>Probabilidad</i>	<i>Probabilidad acumulada</i>
1	100	3	0	97/100	0,97
2	97	5	0	92/97	0,92
3	92	4	0	88/92	0,88
4	88	3	0	85/88	0,85
5	85	6	0	79/85	0,79

CV: cardiovascular.

<i>Año</i>	<i>Número</i>	<i>Evento CV</i>	<i>Censura</i>	<i>Probabilidad</i>	<i>Probabilidad acumulada</i>
1	100	3	0	97/100	0,97
2	87	5	10	82/87	0,91
3	72	4	10	68/72	0,86
4	58	3	10	55/58	0,82
5	45	6	10	39/45	0,71

CV: cardiovascular.

sobrevida o eventos. En la tabla 1 se muestra una población de 100 pacientes en la que se estudia la sobrevida libre de muerte cardiovascular (CV). Se comienza con la población total (100 pacientes) y en cada año se registran los eventos ocurridos (muerte CV) y los pacientes que deben censurarse (porque no tuvieron el evento al finalizar el período de observación, porque abandonaron el estudio o porque murieron de causa no CV, y, por lo tanto, no siguen en el estudio). Si bien no es la intención del presente trabajo instruir sobre la parte matemática implicada en la confección de las tablas de sobrevida, vale la pena aclarar que la probabilidad acumulada es igual a la multiplicación de la probabilidad de cada uno de los períodos. La probabilidad de cada uno de los períodos se calcula según el número de eventos/la población en riesgo (el número de individuos del período anterior menos los que sufrieron el evento en el período anterior, menos aquellos censurados en este período). Imaginemos por un momento que los pacientes que censuran son solo aquellos que sufren muerte no CV. Como resultado se ve a que los cinco años de seguimiento, la sobrevida libre de muerte CV es de 76%. En otras palabras, existe un riesgo de 24% de sufrir muerte CV. Ahora imaginemos que no hay eventos

censurados porque ningún paciente muere de causa no CV (tabla 2). En este caso, a los cinco años de seguimiento, la sobrevida libre de muerte CV es de 79%, o sea, existe una probabilidad de 21% de morir de causa CV. ¿Qué conclusión debemos sacar de esto? La estimación de sobrevida mediante Kaplan-Meier en presencia de RC tiende a subestimar la sobrevida libre de evento (o a sobrestimar el riesgo del evento). Cuanto mayor es el porcentaje de eventos censurados respecto a los eventos registrados, mayor será esta discrepancia. En la tabla 3 se ilustra lo anterior. Tiene representados los mismos datos que la tabla 1 pero con mayor número de eventos censurados. Como se ve, a los cinco años, la sobrevida libre de muerte CV en este caso es de 71%. Lo más eficiente es entonces intentar reducir los eventos censurados mediante un seguimiento a largo plazo (disminuyendo de esta forma los pacientes que no tuvieron el evento), no perder pacientes durante el mismo, e incluir los RC en el evento compuesto. La otra forma es mediante el análisis de la frecuencia de incidencia acumulada (CIF), que será tema de discusión en otro apartado. En conclusión, al enfrentarnos como médicos a una curva de sobrevida debemos evaluar los siguientes aspectos:

- ¿Nuestra muestra es representativa (criterios de inclusión, truncamiento a la izquierda tipo 1)?
- ¿Puede existir un sesgo en nuestra estimación del “tiempo de exposición” (truncamiento a la izquierda tipo 2, censura a la izquierda)?
- ¿Hay posibilidad de que exista un evento lo suficientemente frecuente que compita y por lo tanto evite la posibilidad de ocurrencia del evento dependiente?
- ¿Los métodos del trabajo informan acerca del ajuste y corrección por RC?
- Número de eventos censurados, teniendo en cuenta que un índice mayor de 0,5 (eventos censurados/eventos de interés) puede sobrestimar el dato de sobrevida.

Víctor Dayan, <https://orcid.org/0000-0002-5470-0585>
Abayubá Perna, <https://orcid.org/0000-0001-7062-1310>

Este artículo fue aceptado para su publicación por:
Editor jefe Dr. Gerardo Soca.

Bibliografía

1. **Wald A.** A reprint of a method of estimating plane vulnerability based on damage of survivors. Alexandria: Center for Naval Analyses; 1980.
2. **Van Walraven C, McAlister FA.** Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. *J Clin Epidemiol.* 2016;69:170-3.e8. doi:10.1016/j.jclinepi.2015.07.006
3. **Miyasaka Y, Barnes ME, Petersen RC, Cha SS, Bailey KR, Gersh BJ, et al.** Risk of dementia in stroke-free patients diagnosed with atrial fibrillation: data from a community-based cohort. *Eur Heart J.* 2007; 28:1962-7. doi:10.1093/eurheartj/ehm012
4. **Pintilie M.** Análisis de riesgos competitivos. An introduction to competing risks analysis. *Rev Esp Cardiol.* 2011; 64(7):599-605. doi:10.1016/j.recesp.2011.03.017
5. **Kleinbaum DG, Klein M.** The cox proportional hazards model and its characteristics. En: Kleinbaum DG, Klein M. *Survival Analysis: a Self-learning text.* 3ª ed. Suiza: Springer; 2012: 83-116.