

Del estudio observacional y el ajuste a la decisión clínica. Parte 2

From Observational Studies and Adjustment to Clinical Decision. Part 2

ARTURO CAGIDE¹, MTSAC

RESUMEN

En esta segunda parte de la revisión sobre el análisis de estudios observacionales nos referimos a cada uno de las herramientas con que contamos, sus ventajas y desventajas comparativas. Y concluimos con la interpretación de los resultados de los estudios que nos sirvieron de ejemplo en la parte inicial.

Palabras clave: Estudios observacionales - Análisis multivariado - Ajuste

ABSTRACT

In this second part of the review on the analysis of observational studies, we refer to each of the tools we have available, their comparative advantages and disadvantages. And we conclude with the interpretation of the results of the studies that served as examples in the initial part.

Key words: Observational studies - Multivariate analysis - Adjustment

Continuamos en esta segunda parte con la explicación de los métodos estadísticos, desde los más usuales hasta los más novedosos, para el análisis de estudios observacionales.

Score de propensión (Propensity Score, PS)

El objetivo de esta metodología estadística es igualar los confundidores de modo que resulten balanceados entre los grupos con intervención y control.

La Figura 1 ilustra el procedimiento. Como primer paso se estima estadísticamente la asociación entre los confundidores (ahora variables independientes) y la exposición (ahora variable dependiente), empleándose también en este caso el análisis multivariado (regresión logística).

De su resultado se deriva el **score de propensión** o **propensity score** (PS) que es la probabilidad de quedar expuesto a la intervención, ajustando por la presencia de los confundidores.

Una ventaja que ofrece el PS sobre el análisis multivariado es que el número de las variables independientes no será limitado por la prevalencia del tratamiento, ya que este, intervención o factor pronóstico, a diferencia del *outcome*, siempre tendrá suficiente número de observaciones. (1)

Hay cierto debate acerca de qué variables deben incluirse en el cálculo del PS. En general deberían ser todas aquellas que el investigador considera que condicionan un determinado tratamiento o intervención. En principio las variables determinantes del *outcome* deberían también incluirse. (2,3)

El problema, como en cualquier análisis multivariado, es que solo se consideran las variables independientes conocidas y disponibles, de modo que el PS puede presentar defectos en la predicción de la exposición al tratamiento o intervención como resultado de esa situación.

Para la evaluación del modelo en el cálculo del PS puede aplicarse la curva ROC y, si bien hay discrepancias en el valor a considerar como adecuado, la mayoría de los autores lo ubican en un área bajo la curva de 0,80.

Desde aquí, existen varios procedimientos para evaluar la asociación de la intervención con el punto final. (4) (Figura 1)

a) Apareamiento

Siguiendo la metodología anterior, cada individuo quedará caracterizado, según sus características basales o confundidores, por un determinado PS. Algunos individuos del grupo intervención tendrán similar PS que los

REV ARGENT CARDIOL 2024;92:389-395. <http://dx.doi.org/10.7775/rac.es.v92.i5.20823>

Recibido: 06/06/2024 - Aceptado: 02/08/2024

Dirección para correspondencia: Arturo Cagide. E-mail: arturo.cagide@hospitalitaliano.org.ar



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

©Revista Argentina de Cardiología

1 Hospital Italiano de Buenos Aires

del control, de modo que es posible aparear pacientes de ambos grupos según su PS. (Figura 2)

Se genera de esta forma una subpoblación (apareada) cuyos confundidores estarán balanceados entre ambos grupos de tal forma que ya no influyan en la estimación de la asociación de la intervención con el *outcome*.

Ahora bien, un número significativo de individuos de ambos grupos, tratados y no tratados, será excluido al no disponerse del correspondiente par. Ese número de excluidos se relaciona en forma directa con el grado de desbalance de confundidores entre ambo grupos de la muestra original del estudio.

De este modo, la conclusión del estudio y su traslado a la práctica, se limita exclusivamente a la muestra apareada, sin poder generalizarse a la totalidad de la población.

b) Análisis multivariado y PS

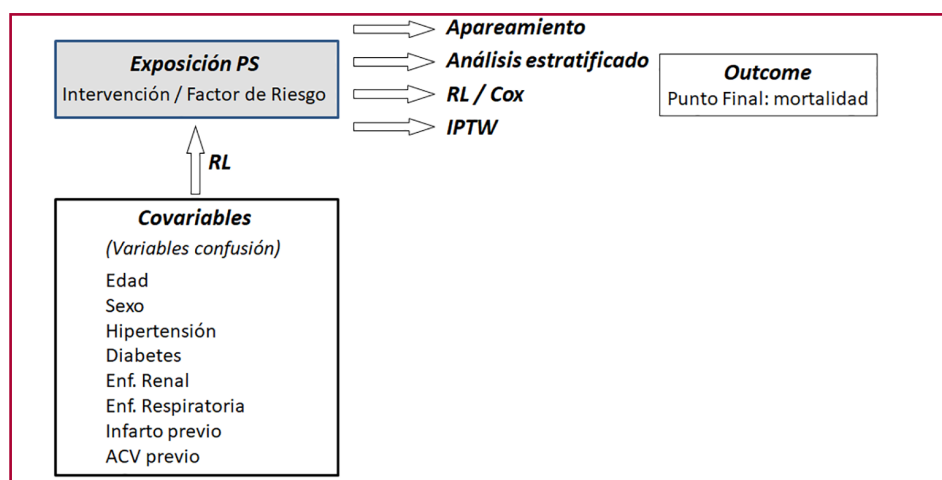
Es posible efectuar un análisis de este tipo incluyendo como variables independientes la intervención (o el

criterio pronóstico según el caso) y, además, el PS, que en este caso representa la totalidad de los confundidores analizados.

c) Análisis estratificado

Con este método, al igual que con el análisis multivariado, se incluyen la totalidad de los individuos del ensayo. El procedimiento consiste en conformar estratos según el PS y en cada uno de ellos estimar la asociación entre la variable de exposición y el *outcome*, para luego obtener un resultado global como expresión de esa asociación. (Figura 3)

El análisis estratificado habitual para el balance de confundidores presenta la dificultad de ajustar solo algunas variables, ya que el número de estratos se eleva exponencialmente al incorporar numerosos criterios con pocas observaciones en cada uno de ellos. Aplicando la estratificación según PS se incorporan todas las características basales de interés, es decir de confundidores.



ACV: accidente cerebrovascular; IPTW: inversa de la probabilidad ponderada de tratamiento; RL: regresión logística

Fig. 1. Score de propensión (PS) y análisis derivados

El objetivo final del estudio es, como el de la Fig. 4 (Parte 1), estimar la posible asociación de la exposición con el *outcome*. Previamente se analiza la asociación de los confundidores con la exposición y se calcula el PS. Desde allí, el PS permite, aplicando diferentes metodologías estadísticas, estimar la asociación independiente de la exposición con el *outcome*.

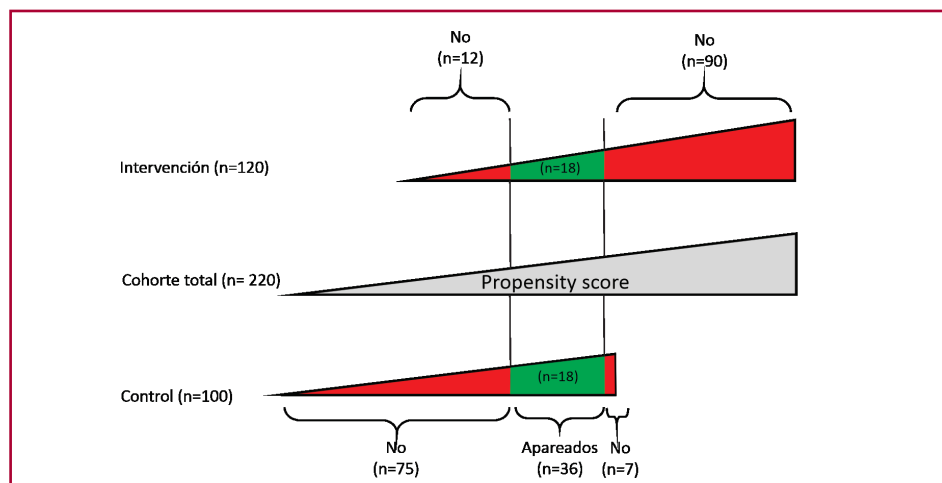


Fig. 2. Apareamiento

Ejemplo teórico de un estudio de 220 pacientes, 120 con intervención y 100 controles. Se grafica, en una escala de valores crecientes, el PS de toda la muestra (centro), de los individuos intervenidos (superior) y de los controles (inferior). Cada individuo con intervención se aparea con un control con igual o muy cercano PS. Se integran 18 pares; la muestra se reduce a 16 % del número original. (36/220). En la muestra apareada los confundidores de ambos grupos, intervención y control, quedan "balanceados".

d) Inversa de la probabilidad ponderada de tratamiento (IPTW)

Este procedimiento, a diferencia del apareamiento, pero al igual que el análisis estratificado, incluye la totalidad de la muestra en estudio.

Ahora bien, mientras que con el apareamiento el ajuste se logra reduciendo la población hasta que los confundidores se igualen en los grupos a comparar, con la IPTW ese objetivo se alcanza incrementando mediante una argucia matemática dicha población con individuos con similar tasa de confundidores. (5)

La Figura 4 es un ejemplo teórico que compara un grupo con intervención con un grupo control. La edad,

dicotomizada en <50 y ≥50 años, es para este ejemplo, el único confundidor. En el grupo intervenido hay cuatro individuos, 3 con <50 años y 1 con ≥50 años; en el control hay 5, 1 de <50 y 4 ≥50 años.

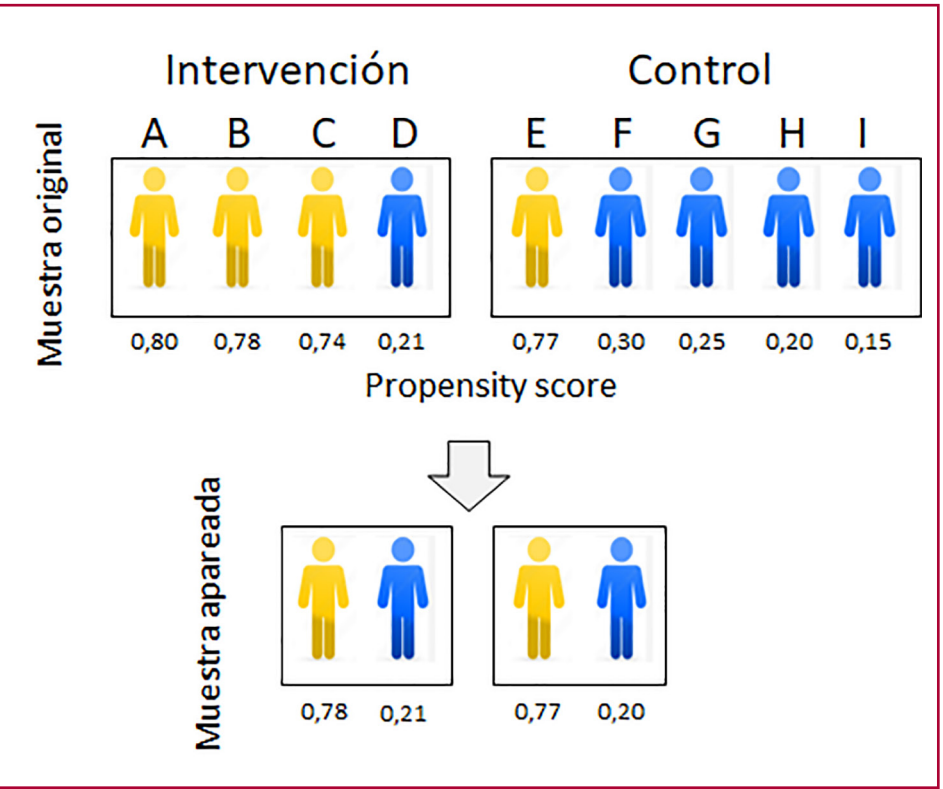
Se debe ajustar la edad para que ambos grupos puedan ser comparados en cuanto a un determinado *outcome*, supongamos mortalidad. Para ello en cada uno se estima el PS según se detalló anteriormente el cual seguramente será diferente en cada individuo.

Si se aplicara la estrategia de apareamiento, se podrían integrar dos pares de 2 pacientes cada uno, tratado y control, (Figura 4) que compartieran similar PS: la muestra quedaría limitada a solo 4 individuos.

Fig. 3. Análisis estratificado
Se agruparon los individuos en quintiles según su PS. En cada uno de ellos se estimó el riesgo relativo (RR) para el outcome (por ejemplo, mortalidad). Luego se calcula el RR de la muestra ajustado por PS, y con ello de las variables de confusión incluidas en él.

PS (quintilos)	Outcome		RR
	Intervención	Control	
1°	%	%	1°
2°	%	%	2°
3°	%	%	3°
4°	%	%	4°
5°	%	%	5°

Fig. 4. Apareamiento
Ejemplo teórico de un estudio que compara la intervención (4 individuos) con el control (5 individuos). En cada uno se calculó el PS. La variable de confusión es la edad: amarillos <50 años, azul 50 años. Para ajustar la edad se aparean individuos de cada grupo con valores similares de PS. La muestra original de 9 se reduce a 2 pares (4 individuos) La muestra, ajustada para edad, ha quedado significativamente reducida en su número.



Veamos ahora el ejemplo de la Figura 5.

A diferencia de la Fig. 4, supongamos ahora que hay nuevamente un solo confundidor, la edad.

La probabilidad de recibir intervención (es decir el PS) con <50 años es 0,75, y con ≥50 años 0,20.

Obsérvese que con igual probabilidad de intervención (PS = 0,75), 3 con <50 años recibieron intervención y solo 1 control; a su vez con ≥50 años y la misma probabilidad de intervención (PS = 0,20) y de control (1-PS = 1-0,20 = 0,80) 4 recibieron control y solo 1 intervención.

Aplicando IPTW:

- **<50 años:** la inversa de la probabilidad de recibir tratamiento es 1,33 (1/0,75): cada individuo intervenido de <50 años ahora quedará representado por 1,33, de modo que la suma de todos ellos es 4. La inversa de la probabilidad de control es 4 (1/1-0,75) (la probabilidad de control es 1 menos la probabilidad de intervención, (1-PS): el único individuo de < 50 años del grupo control quedara representado por 4 (Figura 5).
- **≥50 años:** la inversa de la probabilidad de intervención es 5 (1/0,20): en el grupo intervención el único individuo de ≥50 años será representado por 5. La inversa de la probabilidad de control es 1,25 (1/1-0,20): cada individuo del control con ≥50 años quedará representado por 1,25, de modo que la suma de todos ellos es 5.

Obsérvese que en la muestra original la relación de los individuos de <50 años con y sin tratamiento era 3

a 1, quedando, luego de aplicar IPTW, totalmente igualada en 4. A su vez, en ≥50 años con y sin tratamiento la relación original de 1 a 4 quedó igualada en 5.

De modo que mediante la IPTW se alcanza una pseudo randomización al generar una **pseudo población** en la cual el ajuste de la variable edad es matemáticamente perfecto. (Figura 5).

La situación descripta, es particular al considerar solamente un solo confundidor: antes del ajuste el PS era igual para todos los individuos de <50 años (0,75); los de ≥50 años también compartían igual PS (0,20).

En un contexto real la probabilidad de tratamiento queda condicionada a múltiples confundidores cuyo efecto global queda representado en el PS, el cual será diferente para cada caso (Figura 6).

El ajuste mediante la IPTW es similar a lo explicado a propósito de la Figura 5, aplicando el PS individual de cada paciente. Obsérvese que el ajuste reduce el desbalance, aunque persisten ciertas diferencias en la distribución de la edad entre los grupos con y sin tratamiento, debido a que otros confundidores, además de la edad, condicionan el PS. (6)

La estrategia detallada a propósito de la edad debe aplicarse con todas las variables o confundidores disponibles contemplados en el cálculo del PS.

Estimando el grado de ajuste

La precisión del ajuste es un aspecto crítico cuando se aplica la IPTW ya que se incluyen la totalidad de los

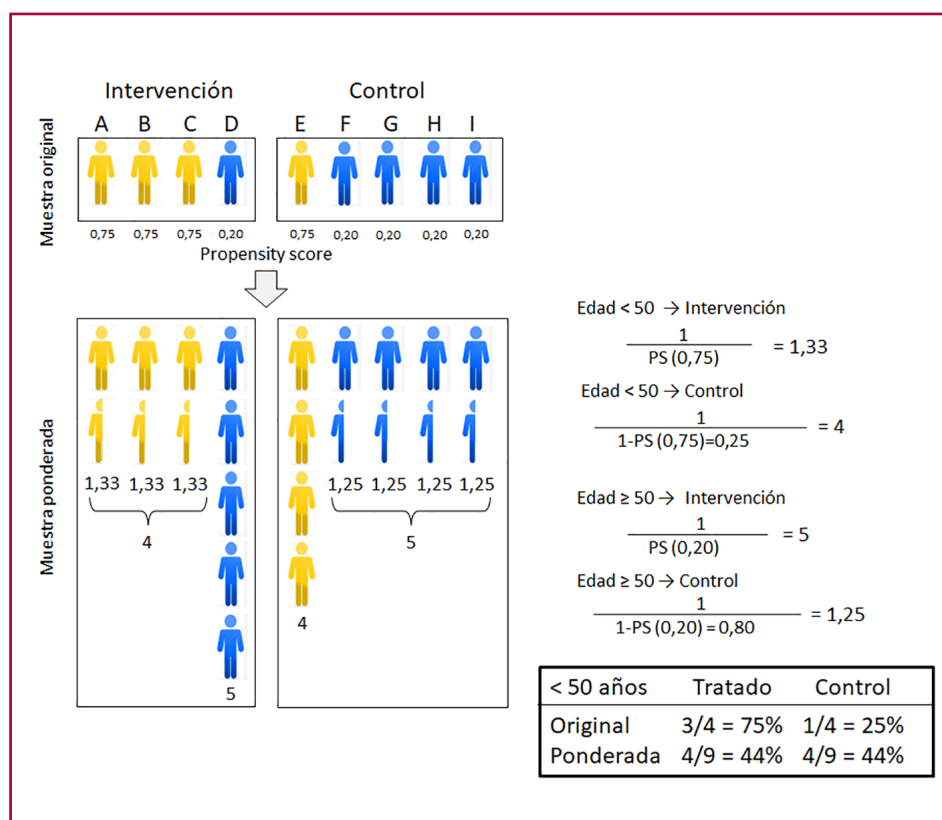


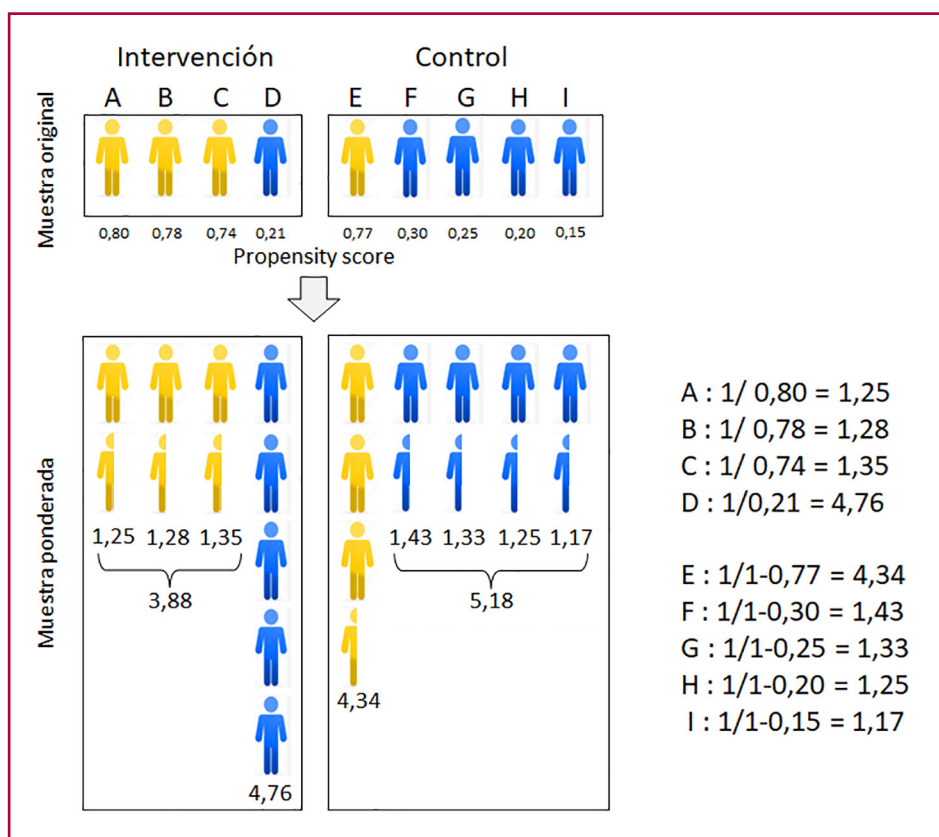
Fig. 5. Inversa de probabilidad ponderada de tratamiento (IPTW)

En la parte superior se presenta la condición de la Fig. 8. Se pretende ajustar la edad (que en este ejemplo teórico es el único confundidor existente), dicotomizada en < 50 (amarillo) y 50 años (azul). Con <50 años la probabilidad de intervención, igual en todos los individuos, es 3/4 = 0,75; con 50 años la probabilidad de intervención, también igual en todos, es 1/5 = 0,20. Las probabilidades calculadas en este caso teórico surgidas de un simple cálculo matemático, constituyen el PS. (La probabilidad de recibir control será 1 — PS)

Las fórmulas indican cómo se calcula la IPTW, a partir del cual se estima cómo quedará representado cada individuo en los grupos intervención y control. EL cuadro muestra la distribución de la edad antes y después del ajuste, que en este caso particular, es perfecta. Explicación en el texto.

Fig. 6. Inversa de la probabilidad ponderada de tratamiento (IPTW)

A diferencia del ejemplo anterior las probabilidades calculadas por el PS son diferentes en cada individuo ya que están condicionadas no solo por la edad sino por la totalidad de las variables de confusión. Su cálculo resulta de un análisis de regresión logística de las covariables con el tratamiento recibido, intervención o control. Aplicando igual metodología que en la Figura 9, el ajuste reduce las diferencias, aunque no es matemáticamente exacto. Explicación en el texto.



individuos en estudio, que seguramente presentarán diferencias sustanciales en la prevalencia de múltiples confundidores. En el caso de apareamiento también puede haber desbalance aunque de menor grado.

Para evaluar el grado de ajuste alcanzado usualmente se emplea la diferencia absoluta estandarizada (DAE), es decir la diferencia medida en unidades de desvío estándar, existente para cada una de las variables de confusión luego del ajuste por IPTW. En general se acepta que el margen que asegura que el ajuste fue adecuado es una diferencia $< 0,10$, aunque en ocasiones se extiende a $< 0,20$, lo que de alguna manera quita consistencia a las conclusiones del estudio.

En ocasiones se grafica la DAE antes y después del ajuste lo que indica el desbalance previo y el éxito del ajuste.

La exactitud del PS para estimar la probabilidad de ser intervenido es un condicionante de la metodología de IPTW. Nuevamente las variables confundidoras no incluidas, desconocidas o no contempladas, constituyen un punto crítico del procedimiento estadístico.

Regresando a los ensayos clínicos desde la perspectiva metodológica

Invitamos a los lectores a volver a la parte 1 de la presente revisión, para entender cabalmente las consideraciones que siguen.

Ensayo 1 (7)

- Aplicando el ajuste de confundidores por apareamiento del PS, de los 9586 pacientes que comprenden la muestra, el análisis quedó restringido a 1866: 933 con angioplastia (ATC) y 933 con tratamiento médico (TM), es decir 933 pares. Es evidente que había un marcado desbalance respecto a ciertas características basales, y que la ATC fue reservada a un grupo seleccionado, globalmente de buen pronóstico, por lo que la corrección de esas diferencias es crítica en cuanto a la validez del resultado.
- Dicha corrección fue adecuada ya que la diferencia estandarizada fue $< 0,05$ en la mayoría de las más de 20 variables consideradas
- El HR ajustado del *outcome* mortalidad / infarto no fatal (1,49, TM con respecto a ATC) se calculó en la muestra apareada según el modelo multivariado de Cox, al cual se le incorporaron ciertas variables con la intención de ajustar alguna diferencia remanente.
- El traslado a la clínica de este ensayo queda limitado exclusivamente al grupo de pacientes apareados, 85% del grupo con TM, pero solo el 11% de los del grupo con ATC.

Ensayo 2 (8)

- En este caso el ajuste se efectuó por IPTW.
- La diferencia estandarizada fue $< 0,05$ en las 13 variables consideradas (Tabla 2, Parte 1).

- En la Figura 2 además de la diferencia estandarizada se muestra el efecto sobre el *outcome* con un riesgo relativo de 0,72, indicando que los pacientes con estatinas previas tuvieron un 28% menos de insuficiencia cardíaca como complicación del síndrome coronario agudo.

Ensayo 3 (9)

- El efecto de la intervención se estimó mediante el PS según tres criterios estadísticos. (Tabla 3, Parte 1)
- IPTW (análisis primario): no beneficio de la anticoagulación en CHADS₂ ≤1, pero sí diferencia en las hemorragias mayores, que fueron significativamente más frecuentes con el tratamiento.
- apareamiento y modelo de Cox (análisis secundario y de sensibilidad). El resultado, no beneficio de la anticoagulación con mayor tasa de hemorragias en el CHADS₂ ≤1, fue coincidente con el análisis post IPTW.
- Interesa más allá del resultado en el CHADS₂ ≤1, analizar la Figura 1 (Parte 1), que ilustra la diferencia estandarizada antes y después del ajuste; el límite de 0,1 definió el criterio del grado de ajuste. Obsérvese que en el CHADS₂ ≤1 la diferencia pre ajuste es mucho mayor que en el CHADS₂ = 2, señalando claramente que el primero es un grupo en el cual las variables, probablemente asociadas al tromboembolismo, fueron marcadamente diferentes en los pacientes no anticoagulados (baja prevalencia) en relación con los que no recibieron ese tratamiento (alta prevalencia).

Ensayo 4 (10)

- Es un estudio pronóstico sobre el riesgo del infarto tipo 2 en relación con el tipo 1 empleando IPTW.
- La diferencia estandarizada se ilustra en un gráfico similar al de la Figura 1 (Parte 1), pero es evidente que ambos gráficos son distintos. En primer lugar el límite de aceptación es <0,2, el doble que en el caso anterior. Pero, además, luego del ajuste las diferencias siguieron siendo elevadas, cercanas a 0,10. Tal vez el tamaño de la muestra en relación con el número de covariables fue incorrecto y con ello el cálculo de PS.
- La conclusión del estudio se ve limitada por un ajuste imperfecto de los confundidores (comorbilidades).

Ensayo 5 (11)

- El objetivo en este caso no fue evaluar el efecto de una intervención o el valor pronóstico de un nuevo criterio, sino conocer la evolución de la estenosis aórtica. El problema radica en que en un grupo de pacientes la intervención suspendió la evolución espontánea, de modo que el grupo no intervenido puede haber quedado sesgado (mayor riesgo por haber seleccionado la intervención a individuos de menor riesgo) o su inversa (haber seleccionado para el procedimiento los de alto riesgo).

- La IPTW no tuvo por objeto comparar el grupo tratado con el no intervenido, sino ajustar esas poblaciones para luego cotejar la ajustada con la original sin tratamiento. La Tabla 4 (Parte 1) no detalla la diferencia estandarizada, sino que compara la mortalidad de la muestra ajustada con la no ajustada en subgrupos según gravedad de la valvulopatía.
- Para los autores el grupo no intervenido representa la evolución espontánea real de la estenosis aórtica no afectada por la intervención.

CONCLUSIÓN

Los estudios randomizados constituyen, indudablemente, el mayor nivel de evidencia. Pero son también los de más complejidad y, sobre todo, tiempo de ejecución y costos. Solo unas pocas intervenciones pueden evaluarse por esta metodología.

Los ensayos observacionales, particularmente cuando se dispone de bases de datos prospectivas de gran dimensión, constituyen una opción válida en la evaluación de nuevos criterios pronósticos, o alternativas terapéuticas innovadoras. Es aquí donde la metodología estadística tiene su mayor desafío: corregir variables que pueden afectar la relación en estudio.

Familiarizarse con términos, tablas y gráficos acerca de la estrategia estadística aplicada al efecto es de importancia vital para la toma de decisiones por el clínico. No se trata de comprender la complejidad matemática sino su base conceptual.

No interiorizarse al respecto es quedar condicionado a las opiniones de expertos. Siempre es preferible analizar la información base que recurrir a la interpretada por otros.

Consideraciones éticas

No aplica

Declaración de conflicto de intereses

El autor declara que no tiene conflicto de intereses (Ver formulario de conflicto de intereses del autor en la web)

BIBLIOGRAFÍA

1. Andrew BY, Alan Brookhart M, Pearse R, Raghunathan K, Krishnamoorthy V. Propensity score methods in observational research: brief review and guide for authors. *Br J Anaesth* 2023 131:805-9. <https://doi.org/10.1016/j.bja.2023.06.054>.
2. Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, Fremes SE. A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. *Can J Cardiol* 2016;32:259-65. <https://doi.org/10.1016/j.cjca.2015.05.015>.
3. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. *Eur J Cardiothorac Surg* 2018;53:1112-7. <https://doi.org/10.1093/ejcts/ezy167>.
4. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Propensity Score Methods for Bias Reduction in Observational Studies of Treatment Effect. *Rheum Dis Clin North Am* 2018;44:203-13. <https://doi.org/10.1016/j.rdc.2018.01.002>.

5. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015;34:3661-79. <https://doi.org/10.1002/sim.6607>.
6. Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C, et al. An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J* 2021;15:14-20. <https://doi.org/10.1093/ckj/sfab158>.
7. Hannan EL, Samadashvili Z, Cozzens K, Walford G, Jacobs AK, Holmes DR Jr, et al. Comparative outcomes for patients who do and do not undergo percutaneous coronary intervention for stable coronary artery disease in New York. *Circulation* 2012;125:1870-9. <https://doi.org/10.1161/CIRCULATIONAHA.111.071811>
8. Bugiardini R, Yoon J, Mendieta G, Kedev S, Zdravkovic M, Vasiljevic Z, et al. Reduced Heart Failure and Mortality in Patients Receiving Statin Therapy Before Initial Acute Coronary Syndrome. *J Am Coll Cardiol* 2022;79:2021-33. <https://doi.org/10.1016/j.jacc.2022.03.354>
9. Kanaoka K, Nishida T, Iwanaga Y, Nakai M, Tonegawa-Kuji R, Nishioka Y, et al. Oral anticoagulation after atrial fibrillation catheter ablation: benefits and risks. *Eur Heart J*. 2024;45:522-34. <https://doi.org/10.1093/eurheartj/ehad798>
10. López-Cuenca A, Gómez-Molina M, Flores-Blanco PJ, Sánchez-Martínez M, García-Narbon A, De Las Heras-Gómez I, et al. Comparison between type-2 and type-1 myocardial infarction: clinical features, treatment strategies and outcomes. *J Geriatr Cardiol* 2016;13:15-22.
11. Généreux P, Sharma RP, Cubeddu RJ, Aaron L, Abdelfattah OM, Koulogiannis KP, et al. The Mortality Burden of Untreated Aortic Stenosis. *J Am Coll Cardiol* 2023;82:2101-9. <https://doi.org/10.1016/j.jacc.2023.09.796>