



SERIES EN MEDICINA INTENSIVA: ACTUALIZACIÓN EN METODOLOGÍA EN MEDICINA INTENSIVA

## Análisis de la causalidad desde los estudios observacionales y su aplicación en la investigación clínica en Cuidados Intensivos



C. Coscia Requena<sup>a</sup>, A. Muriel<sup>a,b,c</sup> y O. Peñuelas<sup>d,e,\*</sup>

<sup>a</sup> Unidad de Bioestadística, Hospital Universitario Ramón y Cajal, Madrid, España

<sup>b</sup> IRYCIS. CIBER de Epidemiología (CIBERESP), Madrid, España

<sup>c</sup> Departamento de Enfermería y Fisioterapia, Universidad de Alcalá, Alcalá de Henares, Madrid, España

<sup>d</sup> Unidad de Cuidados Intensivos y Grandes Quemados, Hospital Universitario de Getafe, Getafe, Madrid, España

<sup>e</sup> CIBER de Enfermedades Respiratorias (CIBERES), España

Recibido el 7 de septiembre de 2017; aceptado el 13 de enero de 2018

Disponible en Internet el 28 de febrero de 2018

### PALABRAS CLAVE

Causalidad;  
Ensayo clínico;  
Estudio  
observacional;  
Confusión;  
Propensión;  
Cuidados Intensivos;  
Epidemiología

**Resumen** Una de las características fundamentales de los ensayos clínicos es la asignación aleatoria de un tratamiento o intervención sobre los pacientes. Esta asignación divide los pacientes en dos grupos que, aunque difieran por el tratamiento recibido, presentan unas características basales homogéneas haciendo que ambos grupos sean comparables y se pueda evaluar el efecto causal del tratamiento. Por otro lado, los estudios observacionales se caracterizan por la asignación no aleatoria del tratamiento y por lo tanto que los grupos de pacientes no solo difieran por el tratamiento recibido, sino también por otras características basales, a menudo relacionadas con la variable de intervención. En numerosas ocasiones, los ensayos clínicos aleatorizados no son factibles por razones éticas, logísticas, económicas o de otro tipo. Uno de los retos de la investigación clínica en Cuidados Intensivos debería ser aprovechar los datos que provienen de la práctica clínica habitual y analizarlos como si fueran ensayos clínicos. Los estudios observacionales utilizando métodos de análisis con índices de propensión (*propensity score*) han ido en aumento en los artículos científicos de Cuidados Intensivos. Los análisis de índices de propensión intentan controlar la confusión en estudios observacionales ajustando la probabilidad de que un determinado paciente esté expuesto. Sin embargo, los estudios con índices de propensión pueden ser confusos, y los intensivistas no están familiarizados con esta metodología y pueden no comprender plenamente la importancia de esta técnica. Los objetivos de esta revisión son: describir los fundamentos de los métodos del índice de propensión; presentar las técnicas para evaluar adecuadamente los modelos de índices de propensión, y discutir las ventajas y los inconvenientes de estas técnicas.

© 2018 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

\* Autor para correspondencia.

Correo electrónico: [oscar.penuelasro@salud.madrid.org](mailto:oscar.penuelasro@salud.madrid.org) (O. Peñuelas).

<https://doi.org/10.1016/j.medin.2018.01.002>

0210-5691/© 2018 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

**KEYWORDS**

Causality;  
 Clinical trial;  
 Observational study;  
 Confounders;  
 Propensity score;  
 Intensive Care;  
 Epidemiology

**Analysis of causality from observational studies and its application in clinical research in Intensive Care Medicine**

**Abstract** Random allocation of treatment or intervention is the key feature of clinical trials and divides patients into treatment groups that are approximately balanced for baseline, and therefore comparable covariates except for the variable treatment of the study. However, in observational studies, where treatment allocation is not random, patients in the treatment and control groups often differ in covariates that are related to intervention variables. These imbalances in covariates can lead to biased estimates of the treatment effect. However, randomized clinical trials are sometimes not feasible for ethical, logistical, economic or other reasons. To resolve these situations, interest in the field of clinical research has grown in designing studies that are most similar to randomized experiments using observational (i.e. non-random) data. Observational studies using propensity score analysis methods have been increasing in the scientific papers of Intensive Care. Propensity score analyses attempt to control for confounding in non-experimental studies by adjusting for the likelihood that a given patient is exposed. However, studies with propensity indexes may be confusing, and intensivists are not familiar with this methodology and may not fully understand the importance of this technique. The objectives of this review are: to describe the fundamentals of propensity index methods; to present the techniques to adequately evaluate propensity index models; to discuss the advantages and disadvantages of these techniques.

© 2018 Elsevier España, S.L.U. y SEMICYUC. All rights reserved.

**Introducción**

En la investigación clínica desarrollada en las Unidades de Cuidados Intensivos (UCI), uno de los objetivos habituales es evaluar la asociación causal que existe entre un tratamiento o intervención (exposición) y el desenlace de salud de un paciente (muerte curación, alta en UCI). Los ensayos clínicos representan el diseño de investigación de referencia cuando se quiere evaluar la eficacia que produce el tratamiento sobre el evento de interés, ya que se reduce la probabilidad de aparición de sesgos de selección o confusión. Los ensayos clínicos consisten en asignar un tratamiento aleatoriamente a un conjunto de pacientes con un estadio de la enfermedad similar para estimar el efecto que este presente sobre el desenlace; habitualmente en UCI este desenlace es mortalidad durante la estancia en UCI, o durante los 28 días tras su ingreso, o la necesidad de realizar un tratamiento más agresivo, como por ejemplo la traqueotomía<sup>1-3</sup>.

La asignación del tratamiento es la característica principal de los ensayos clínicos. La elección del tratamiento de cada paciente se realiza de forma aleatoria para que pacientes tratados y no tratados presenten características homogéneas y en consecuencia el efecto del tratamiento no se vea confundido por las características de los pacientes.

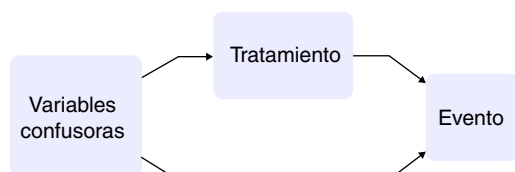
Aunque se consideren los estudios de más calidad, para la estimación de la causalidad, los ensayos clínicos también presentan algunas limitaciones: el tamaño muestral es reducido y difícilmente alcanzable, cuentan con una baja validez externa, los criterios de inclusión reducen el marco de la población a analizar, ya que tienden a excluir pacientes con edades avanzadas y con más de una enfermedad, mujeres; además, los ensayos clínicos presentan aspectos éticos a tener en cuenta y el tiempo de seguimiento es reducido<sup>4</sup>.

En el caso de la UCI, existen además otras dificultades que hacen que el diseño de ensayos clínicos en nuestra disciplina sea particularmente difícil, como son: falta de nosografía segura —los pacientes ingresados en UCI presentan síndromes (por ejemplo, síndrome de distrés respiratorio agudo [SDRA]) en vez de enfermedades—, dificultades para identificar grupos control adecuados, el uso concomitante de diferentes terapias (a menudo la intervención no es un medicamento sino una actitud terapéutica), la aleatorización antes del tratamiento y la solicitud de consentimiento informado es difícil debido al momento en que se produce<sup>5</sup>.

Una posible solución a algunas de las dificultades de los ensayos clínicos son los estudios observacionales. Son un tipo de investigación cuya selección del tratamiento está condicionada a las características basales del paciente. El propio médico determina el tipo de tratamiento en función de las características que presenta el paciente. Esta es una de las principales diferencias entre ambos estudios<sup>6-9</sup>.

Los estudios observacionales presentan varias ventajas en comparación con los ensayos clínicos: el marco de la población es más amplio, así como el tiempo de seguimiento, y presentan un tamaño muestral mayor. Por otra parte, el hecho de que la asignación del tratamiento no sea aleatoria conlleva que la estimación de la causalidad sea sesgada y, en consecuencia, que pacientes tratados y no tratados difieran no solo en el tratamiento que reciben sino también en las características basales. Si estas variables basales tuvieran asociación con el evento de interés, se estaría indicando que son variables confusoras entre la exposición y el desenlace<sup>10,11</sup> (fig. 1).

Las técnicas estadísticas están en continuo desarrollo y se han ido incorporando nuevos métodos cada vez más utilizados en la investigación clínica que permiten estimar la causalidad en estudios observacionales teniendo en cuenta



**Figura 1** Diagrama representativo de la relación entre las variables confusoras y el evento en los estudios observacionales.

los posibles factores de confusión, concretamente el índice de propensión (IP) (*propensity score*) y los modelos marginales estructurales (*marginal structural models*).

Los artículos publicados utilizando los métodos de IP han ido incrementándose cada vez más en la literatura científica de Cuidados Intensivos en los últimos 10 años (fig. 2). Dado que la metodología de los IP no es aún muy familiar para los intensivistas, en este artículo se presenta el marco conceptual en que se basan y se indican las mejores prácticas para su aplicación. Por tanto, los objetivos de esta revisión son describir: 1) los fundamentos de los métodos del IP; 2) los métodos principales para el uso de la puntuación de la propensión (emparejamiento, estratificación, ajuste por covariable y ponderación por el inverso de la probabilidad de tratamiento (*inverse probability of treatment weighting* [IPTW]) en un ejemplo de Cuidados Intensivos; 3) el análisis estadístico ante la presencia de confusores tiempo-dependientes, y 4) los puntos fuertes y débiles de estas técnicas.

## El índice de propensión en los estudios observacionales

El IP, según Rosenbaum y Rubin<sup>12</sup>, es un método que permite estimar la probabilidad de asignación a un tratamiento condicionada a las características basales que el paciente presente. Este índice es balanceado, puesto que pacientes tratados y no tratados con IP similares tendrán características basales equivalentes. De este modo se estaría controlando el efecto de la confusión producido por las características basales.

Para que se pueda aplicar el IP, se deben cumplir dos condiciones: que no haya factores de confusión no medidos y, además, que cada sujeto presente una probabilidad distinta de cero de recibir un tratamiento. Si se cumplen estas dos condiciones, se puede decir que la asignación del tratamiento es independiente con respecto al resultado, cuando se está condicionando por las covariables.

En los ensayos clínicos el IP es conocido, puesto que se conoce con qué probabilidad se asigna un tratamiento a un paciente (por ejemplo, si es un estudio con dos ramas de tratamiento, la probabilidad de asignación de un sujeto será 0,5 respectivamente). Sin embargo, en los estudios observacionales este índice es desconocido. Se podría estimar mediante regresión logística: en el que la variable dependiente es el tratamiento y las variables independientes son las covariables basales posiblemente confusoras.

## VARIABLES A CONSIDERAR EN EL CÁLCULO DEL ÍNDICE DE PROPENSIÓN

Las variables incluidas en el IP pueden crear sesgos, modificando la varianza del estimador o el error si no están correctamente elegidas. Para seleccionar dichas variables no existe un consenso, ya que depende del planteamiento clínico, estableciendo las variables con criterio clínico, y del planteamiento estadístico<sup>13</sup>.

Dado un conjunto de variables basales, se pueden clasificar en 1) variables que se relacionan con el evento, 2) variables que se relacionan con la exposición y 3) variables que se asocian tanto con el evento como con la exposición.

Al ser un modelo de regresión logística, el número de variables viene determinado por el número total de expuestos/no expuestos siguiendo la regla de Peduzzi, es decir, 10 pacientes por evento de interés (expuesto/no expuesto).

En estudios de simulación se ha podido comprobar que las variables que se relacionan con el resultado (evento) se deben incluir, así como las variables que presenten asociación con el evento y la exposición<sup>14</sup>. No se deberán incluir aquellas variables que presenten asociación solo con el tratamiento.

La elección de las variables constituye un reto en el que se deben consensuar hallazgos estadísticos y criterio clínico.

## MÉTODOS DE APLICACIÓN DE LOS ÍNDICES DE PROPENSIÓN

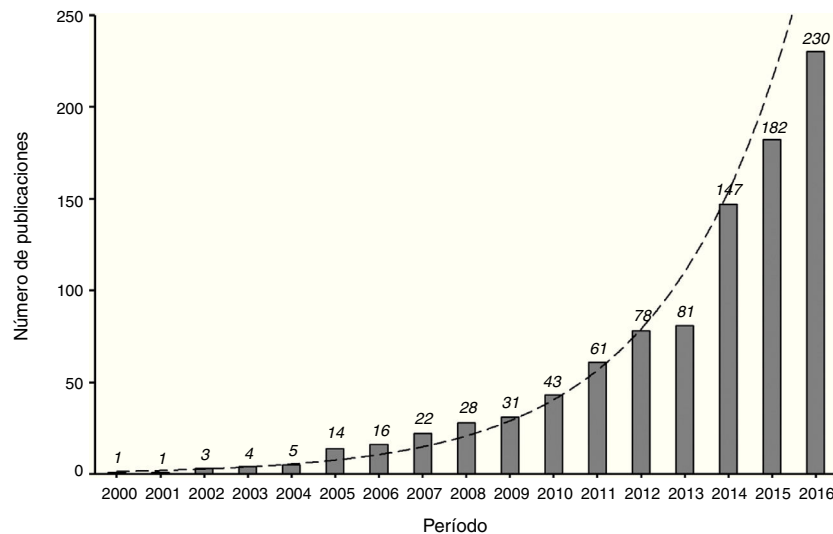
Una vez que el IP ha sido estimado para todos los pacientes, hay distintos métodos de aplicación: mediante emparejamiento (*matching*), estratificación, IPTW, y ajustando el modelo final incluyéndolo como covariable confusora.

### Emparejamiento

El método del emparejamiento se basa en crear una nueva muestra de pacientes en función del IP estimado previamente. Se seleccionan los pacientes con IP similares para crear parejas de pacientes tratados y pacientes no tratados. El objetivo de este método es crear una nueva muestra balanceada y reducir las diferencias que pueda haber entre pacientes de ambos grupos. Este método permite realizar parejas considerando diversos aspectos: se podría tener en cuenta el reemplazamiento, por lo tanto, se podría decidir si un individuo puede encontrar pareja dos veces; se puede emparejar por el «vecino», paciente, que esté más cercano; y también se puede fijar un calibre (*caliper*) que indique la distancia máxima con la que se puede emparejar a dos pacientes, que suele ser el 0,25 por la desviación estándar del IP.

Con esta nueva muestra, con estructura muy similar a la de un ensayo clínico, se estaría solucionando el problema de la confusión y se estarían reduciendo las diferencias de las características de pacientes tratados y controles.

En la muestra emparejada, habría que comprobar, mediante diferencias estandarizadas, que las características basales de los pacientes tratados y los no tratados, después del emparejamiento, sean menores en un 10%<sup>15</sup>.



**Figura 2** Diagrama de flujo de las publicaciones en pacientes críticos adultos con uso de índices de propensión. Fuente: gráfico de elaboración propia de los autores (búsqueda en PubMed hasta marzo de 2017).

Las diferencias estandarizadas se expresan de la siguiente forma para las variables continuas:

$$d = \frac{\bar{x} - \bar{x}_{\bar{T}}}{\sqrt{\frac{S_T^2 + S_{\bar{T}}^2}{2}}}$$

Siendo  $\bar{x}_T$ ,  $\bar{x}_{\bar{T}}$  las medias de las variables en los tratados y en los no tratados, y  $S_T^2$ ,  $S_{\bar{T}}^2$  las varianzas de dichas variables.

Del mismo modo se definen las diferencias estandarizadas para variables dicotómicas:

$$d = \frac{(\hat{p}_T - \hat{p}_{\bar{T}})}{\sqrt{\frac{\hat{p}_T(1-\hat{p}_T) + \hat{p}_{\bar{T}}(1-\hat{p}_{\bar{T}})}{2}}}$$

Donde  $\hat{p}_T$  es la proporción de eventos de una variable para los individuos tratados y  $\hat{p}_{\bar{T}}$  es la proporción para individuos no tratados.

El último paso es realizar la estimación del efecto del tratamiento sobre el evento de interés en la muestra emparejada. Si la variable respuesta es continua se podrá estimar la diferencia entre los dos grupos mediante el test de la t de Student o, en caso de no cumplir con la hipótesis de normalidad, mediante el test de Wilcoxon. Si la variable respuesta es dicotómica, se estimará el efecto mediante el test de McNemar.

#### **Inverse probability treatment weighting (IPTW)**

Consiste en estimar pesos basados en el IP, calculados como la inversa de dicha probabilidad para tratados y no tratados, y posteriormente realizar la estimación del efecto del tratamiento ponderando por el peso anteriormente calculado.

#### **Estratificación**

El método de estratificación consiste en dividir la muestra en varias submuestras basadas en el IP. El número de grupos a utilizar suele ser 5, en función de los quintiles o de

otros percentiles del IP estimado. Rosenbaum et al. indican que se consigue reducir el sesgo en un 90%. El efecto del tratamiento se estimará mediante la *odds ratio* (OR) de Mantel-Haenszel (MH).

#### **Índice de propensión**

La última alternativa consiste en calcular el efecto del tratamiento incluyendo el IP como covariable del modelo. Este método tiene la premisa de que la relación entre el desenlace (evento) y la variable confusora debe estar correctamente especificada. Es decir, que se haya realizado de forma correcta; el empleo de métodos no lineales, *splines* o polinomios fraccionales podría ayudar a establecer modelos adecuados.

#### **Comparación entre los distintos métodos**

El método del emparejamiento presenta la ventaja de que, una vez que se han obtenido las parejas, la nueva estructura se asemeja a la de un ensayo clínico. Por otro lado, esta técnica necesita una muestra grande donde haya un mayor número de pacientes no tratados que tratados para que todos los pacientes tratados puedan encontrar pareja. Al emparejar, además, se estaría perdiendo toda la información de aquellos pacientes que no encontraron pareja.

La ventaja principal que presenta el método de estratificación es que permite utilizar toda la muestra de los pacientes. Además, es una técnica más robusta si el IP no está correctamente especificado en comparación con el método del IPTW o el de la covariable.

El método IPTW, según Deb et al.<sup>15</sup>, tiene la desventaja de que los pesos calculados pueden ser inestables si hay pacientes con probabilidades bajas de recibir el tratamiento: si el peso se considera como la inversa del IP, aquellos pacientes con un IP bajo obtendrán un peso mayor que los pacientes con un IP alto.

Considerando el IP como covariable del modelo, se consigue de forma sencilla poder incluir un gran número de

variables en el modelo. Para poder utilizar este método hay que comprobar que la relación entre el IP y el desenlace esté correctamente especificada; la asunción de normalidad debería ser comprobada si el modelo de regresión es lineal. Presenta el inconveniente de que la comprobación del balance de las variables basales (balance diagnóstico) es más complicada que en otros métodos, además de haberse demostrado que produce estimaciones más sesgadas<sup>15</sup>. Por último, la implementación que se tiene de este método no permite la estimación de la reducción absoluta del riesgo o del número necesario para tratar (NNT).

### Índice de propensión y regresión

Bajo ciertas condiciones, las estimaciones obtenidas por los modelos de regresión multivariantes y el IP son coincidentes. La diferencia entre el IP y la regresión logística está muy condicionada al número de eventos y al número de covariables por las que ajustar<sup>16</sup>.

Según Harrell et al.<sup>17</sup>, el número óptimo de variables independientes que se debería incluir en un modelo de regresión logística es 10 veces el número de eventos que se tienen en la muestra, lo que se conoce como la regla del 10. Es decir, si el número de eventos es 30 (fallecimientos, por ejemplo), el número óptimo de variables explicativas para estimar la probabilidad de fallecer debería ser 3. La regresión logística dependerá por lo tanto del número de eventos y en consecuencia del tamaño muestral<sup>18</sup>.

Sin embargo, en la regresión logística del IP, la variable dependiente no es el evento, sino la exposición. El número de variables para estimar la exposición no solo dependerá de la asociación que presenten con el evento, como se decía anteriormente, sino también por el número de exposiciones que haya. Si, por ejemplo, el número de exposiciones fuera 60 (60 fumadores), se debería ajustar el IP con 6 variables.

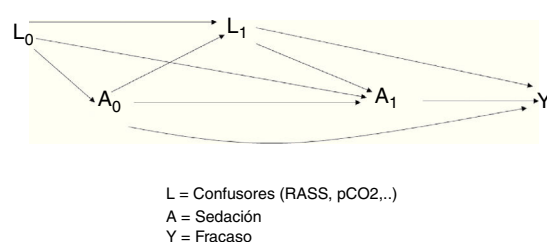
Dependiendo de la prevalencia del evento y del número de expuestos al tratamiento, el IP permite ajustar un modelo con más variables que el modelo de regresión logística.

### Confusores tiempodependientes. Modelos marginales estructurales

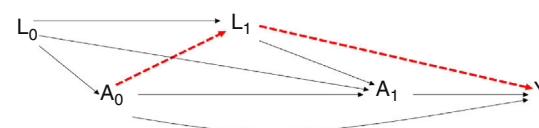
El IP es una técnica estadística que permite estimar el efecto del tratamiento sobre el desenlace de un paciente. Cuando las variables clínicas confusoras son tiempodependientes, la estimación de dicho índice resulta más compleja puesto que los valores de las variables confusoras varían a lo largo del tiempo.

*Ejemplo.* Si se quisiera evaluar el efecto causal de la sedación (A) en el fracaso de la ventilación mecánica no invasiva (Y), podrían existir variables confusoras (L), por ejemplo, la escala de sedación de RASS o algunos marcadores respiratorios como el pH o la PaCO<sub>2</sub><sup>18,19</sup>.

Tanto la escala de sedación RASS como los marcadores respiratorios (pH o PaCO<sub>2</sub>) se miden en distintos periodos de tiempo. Por tanto, se considerarán como confusoras tiempodependientes si los valores pasados de la covariable pueden predecir el tratamiento actual o si los valores actuales podrán predecir resultados futuros, condicionados al tratamiento recibido en pasado.



**Figura 3** Diagrama representativo de las variables confusoras tiempodependientes en los estudios observacionales. A<sub>0</sub> es la variable a estudio o exposición; Y es el desenlace de interés; L<sub>0</sub> representa las variables que modifican la variable a estudio en función del tiempo.



**Figura 4** Diagrama representativo de las variables confusoras tiempodependientes en los estudios observacionales. Podría suceder que se estableciera una relación entre la variable de exposición y el desenlace simplemente influido por la existencia de las variables confusoras tiempodependientes (camino alternativo en rojo) y no verdaderamente por la existencia de una relación causal entre la exposición y el desenlace.

En este ejemplo, según se observa en la [figura 3](#), se quiere estimar el efecto del tratamiento de sedación (A), tanto en el instante 0 (A<sub>0</sub>) como en el instante 1 (A<sub>1</sub>). Por otro lado, L<sub>1</sub> representa a las variables confusoras (escala RASS y pCO<sub>2</sub>) medidas en el instante 1, e Y representa el desenlace de interés (fracaso de la ventilación no invasiva). Al estudiar la relación entre el tratamiento en el instante 0 y el desenlace, aparecen dos caminos causales: A<sub>0</sub> – Y y A<sub>0</sub> – L<sub>1</sub> – Y. El camino A<sub>0</sub> – Y es un camino directo.

Sin embargo, en el camino A<sub>0</sub> – L<sub>1</sub> – Y, la relación que se establece entre tratamiento y desenlace pasa por las variables confusoras L<sub>1</sub>. Para estimar el efecto, según los modelos clásicos, se podría utilizar una regresión sin considerar los confusores L<sub>1</sub> puesto que están en la misma cadena causal entre el efecto (A<sub>0</sub>) y el desenlace (Y) ([fig. 4](#)).

Si se estima el efecto sin ajustar por L<sub>1</sub>, se estaría cometiendo un sesgo dado que no se estaría considerando el camino causal que hay entre A<sub>1</sub> e Y que pasa por L<sub>1</sub>. Los métodos de regresión pueden presentar problemas de sesgo aunque no haya confusión residual; esto es debido posiblemente a la confusión creada por las variables tiempodependientes en la asignación del tratamiento previo.

Existen distintas alternativas para poder estimar el efecto de un tratamiento en presencia de confusores tiempodependientes: modelos estructurales marginales y modelos estructurales anidados<sup>20</sup>.

Los modelos marginales estructurales<sup>19</sup> se consideran como alternativa a los modelos de regresión cuando existe una variable confusora tiempodependiente que se encuentra asociada con el desenlace de interés y que también se relaciona con el tratamiento que se está evaluando. La palabra estructural indica un efecto causal, no solo asociación estadística.

En lugar de la distribución conjunta se usan las distribuciones marginales, condicionadas a las variables basales, por lo que se denominan modelos marginales.

### Estimación del efecto mediante un modelo marginal estructural

Para la creación de los modelos marginales estructurales, el primer paso es la estimación del IP mediante modelos de regresión logística o probit, en el cual la variable dependiente es el tratamiento y las independientes son aquellas variables asociadas con el inicio del tratamiento y/o con el evento de interés.

Cada paciente contribuye con tantas observaciones como unidades de tiempo se hayan seguido (meses, semanas... ). Se deberá añadir como covariable la variable tiempo de seguimiento.

El peso de cada paciente y tiempo es estimado según las probabilidades predichas por este modelo, mediante el inverso de la probabilidad de tratamiento que realmente recibió. En este contexto, se asume que no hay eventos competitivos puesto que se considera que cada paciente es seguido hasta que ocurre el evento de interés o se pierde su seguimiento.

Del mismo modo que anteriormente, se estima un peso, para cada paciente, de ser censurado.

Mediante su producto, se combinan los pesos obtenidos del tratamiento y de la probabilidad de censura; de este modo, con esta ponderación, se puede simular una muestra en la que el tratamiento y la censura son independientes de los confusores medidos. En caso de que haya probabilidades muy pequeñas, los pesos extremos se sustituyen por percentiles cercanos<sup>21</sup> (1 y 99 o 2,5 y 97,5) evitando así la presencia de valores de pesos muy grandes.

En el caso de que el evento de interés sea una variable binaria, la OR de la regresión logística entre el evento y el tratamiento, ponderada por el peso combinado, aproxima bien al riesgo relativo instantáneo del modelo de Cox, porque el riesgo de eventos es bajo en todos los meses.

### Asunciones

Para poder utilizar los métodos estadísticos anteriormente descritos se deben verificar ciertas asunciones, como son:

- El orden temporal: es decir, que la exposición al tratamiento ocurra antes que el evento.
- Consistencia: el resultado potencial de un individuo, condicionado al historial del tratamiento observado, coincide con el verdadero resultado observado.
- Positividad: los individuos de la población deben presentar una probabilidad mayor de cero de recibir el tratamiento en cada categoría del tratamiento y en cada uno de los niveles de las variables confusoras. De este modo, el efecto causal promedio del tratamiento se puede estimar dentro de cada subgrupo de la población establecido por las variables confusoras.
- Correcta especificación del modelo: es decir, que se haya estimado el modelo de forma adecuada. La asunción de linealidad se debería evaluar.

- Ausencia de confusión residual: se puede comprobar mediante un análisis de sensibilidad puesto que no se puede asumir utilizando solo los datos de los pacientes.

### Aplicación del índice de propensión en pacientes ingresados en Unidades de Cuidados Intensivos

La ventaja fundamental del IP es que resume en un solo número las características basales de los individuos en los que se realiza la investigación. Este índice resume con un número la probabilidad de ser expuesto.

Por ejemplo, si el objetivo del estudio es evaluar el efecto de la ventilación no invasiva en niños que ingresan en UCI en un estudio observacional de más de 30.000 pacientes de los cuales casi un tercio recibe como primera opción ventilación no invasiva, este tipo de pregunta de investigación se podría haber planteado, con dificultades, mediante un ensayo clínico<sup>22</sup>.

También se ha utilizado este tipo de análisis para estimar el efecto ajustado del cáncer en la mortalidad en UCI<sup>23</sup>, ya que esta investigación solo se puede contestar en investigación clínica mediante estudios observacionales.

También se podrían utilizar estas técnicas analíticas para estimar el efecto de una complicación como es la debilidad adquirida en la UCI (DACI) en el fracaso del *weaning* o en la mortalidad del paciente durante su estancia en UCI<sup>24,25</sup>.

Vamos a desarrollar este último ejemplo, paso a paso. Se trata de un estudio prospectivo, internacional y multicéntrico realizado con 4.157 pacientes con ventilación mecánica durante más de 12 h. La aparición de DACI durante la estancia se asoció con más fracaso del *weaning* y con más mortalidad en UCI. El IP se estimó con las siguientes variables: edad, SAPS II al ingreso en UCI, motivo principal de inicio de la ventilación mecánica (EPOC, insuficiencia cardíaca, sepsis, SDRA y neumonía).

Para evaluar el efecto de la sedación y analgesia en pacientes con ventilación no invasiva en la necesidad de intubación endotraqueal y ante la presencia de confusores tiempodependientes como son la escala de sedación de RASS (Richmond Agitation-Sedation Scale), la PCO<sub>2</sub> o el RASS es necesario plantear un modelo marginal estructural. Para ilustrar la utilización de estos métodos comparando las distintas opciones del IP, y mostrando los resultados mediante métodos clásicos como la regresión logística, se usará un conjunto de datos de un estudio multicéntrico de UCI con ventilación mecánica.

### Primera parte

Teniendo en cuenta como tratamiento la interrupción de la sedación y como evento el fallecimiento del paciente, se ha aplicado el IP para evaluar la causalidad. Aplicando diferentes modalidades del IP (*matching*, estratificación, IPTW y como covariable del modelo), para este escenario clínico, se han obtenido estimaciones en términos de OR e IC (95%) similares. Estas OR se han comparado con las que se obtienen estimando el efecto sin el IP (tabla 1).

Estos resultados tan similares se deben a que, en relación con las características basales, pacientes tratados y no

**Tabla 1** Relación entre riesgo de mortalidad a día 28 (status día 28) y la interrupción de la sedación según diferentes estrategias de análisis de propensión

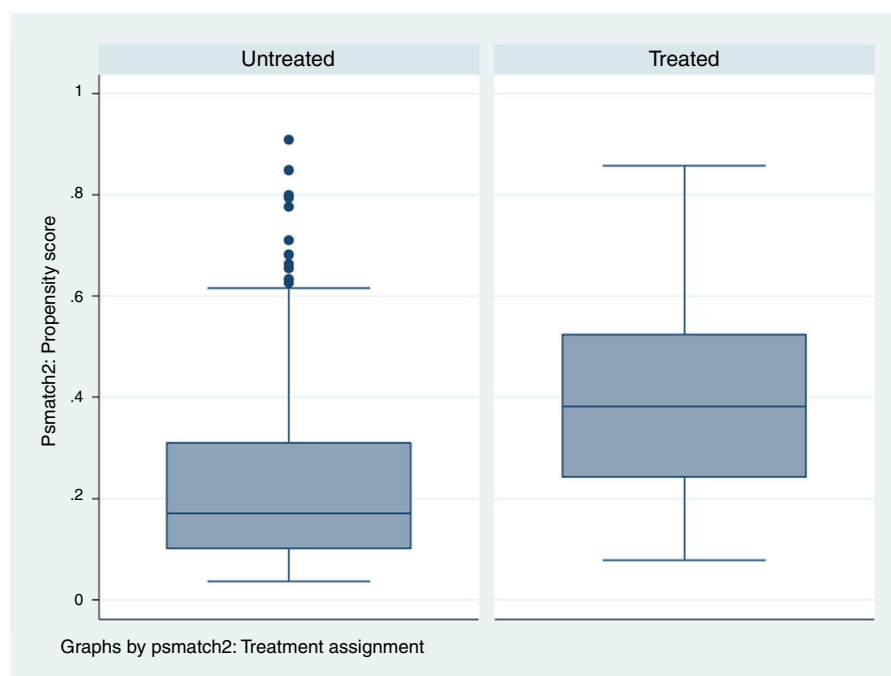
	Aplicaciones del <i>Propensity score</i>								Sin <i>Propensity score</i>			
	Emparejamiento		IPTW		Estratificación MH		Covariable		RL Univariante		RL Multivariante	
	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)
Status día28	0,64	(0,49; 0,83)	0,67	(0,52; 0,85)	0,68	(0,53; 0,86)	0,67	(0,52; 0,85)	0,64	(0,52; 0,80)	0,63	(0,49; 0,81)

IC: intervalo de confianza; IPTW: inverse probability of treatment weighting; MH: Mantel-Haenszel; OR: odds ratio; RL: regresión logística.

**Tabla 2** Relación entre riesgo de mortalidad a día 28 (status día 28) y la interrupción de la sedación según diferentes estrategias de análisis de propensión

	Aplicaciones del <i>Propensity score</i>		Sin <i>Propensity score</i>	
	OR	IC (95%)	OR	IC (95%)
Emparejamiento	1,72	(0,55; 5,97)	Univariante	2,79 (1,26; 6,17)
IPTW	2,08	(0,83; 5,24)	Multivariante	3,18 (1,42; 7,13)
Estratificación	1,81	(1,14; 7,14)		
Covariable	2,06	(0,84; 5,11)		

IC, intervalo de confianza; IPTW: inverse probability of treatment weighting; OR: odds ratio.

**Figura 5** Gráfico de diagrama de cajas que sirve para comparar —en este caso para observar— que los pacientes tratados (*treated*) con los bloqueantes neuromusculares son distintos a los pacientes que no han sido tratados (*untreated*) con bloqueantes neuromusculares (NMB).

tratados no difieren y por lo tanto el IP estimado es muy similar en ambos conjuntos de sujetos.

## Segunda parte

Si se considerase un conjunto de pacientes cuyas características basales presentaran más diferencias, las estimaciones

obtenidas del efecto del tratamiento no concordarían en los diferentes métodos. Por ejemplo, si se considerase solo aquellos pacientes con enfermedades respiratorias y se quisiera estudiar el efecto de los bloqueantes neuromusculares sobre el delirio del paciente, se obtendrían estimaciones diferentes en los diferentes métodos ([tabla 2](#)).

Los pesos obtenidos mediante el IPTW abarcan los siguientes valores: [1,037; 12,75].

**Tabla 3** Comparación de las diferencias estandarizadas de las variables basales en pacientes tratados y no tratados, antes y después del emparejamiento para comprobar que el índice de propensión (IP) está especificado correctamente. Si estas diferencias son mayores en un 10% significará que el IP no está calculado de modo adecuado

	Antes del emparejamiento			Después del emparejamiento		
	NMB = Sí (n = 148)	NMB = No (n = 390)	Diferencias estandarizadas (%)	NMB = Sí (n = 141)	NMB = No (n = 381)	Diferencias estandarizadas (%)
<i>Variables basales</i>						
Edad (años), media (DE)	59,53 (13,53)	65,33 (13,58)	42,79	60,53 (14,53)	58,96 (13,64)	11,17
SAPS-II puntos), media (DE)	43,74 (15,49)	46,83 (16,76)	19,15	44,99 (15,67)	43,41 (14,78)	10,37
Sexo (V)	56,8	48,5	16,63	55	58,3	6,73
Soporte ventilatorio invasivo (Sí)	21,6	21,3	0,59	23,3	21,7	3,98
Presencia de fracaso cardiovascular (Sí)	68,9	40,8	58,37	67,5	68,3	1,78
Presencia de fracaso renal (Sí)	42,6	21,28	40,87	36,7	40	6,85
Presencia de fracaso hematológico (Sí)	24,3	8,97	25,75	19,2	16,7	6,52
Sepsis durante la ventilación mecánica	43,24	24,1	37	40,88	45,83	10,10
Duración de soporte ventilatorio (días)	15,90 ± 13,14	8,21 ± 14,82	54,91	14,35 ± 11,60	11,97 ± 19,04	15,09

Cuando pacientes tratados y no tratados difieren y el número de eventos y de exposiciones es reducido, se observan diferencias significativas entre los métodos estadísticos.

Para ver cómo se distribuyen los pacientes tratados y los no tratados mediante el IP estimado se estudia la zona de soporte común: es el rango de valores comunes que el IP presenta en ambos grupos.

Por lo tanto, cuanto más parecidos sean los grupos de tratados y no tratados, la zona de soporte común abarcará mayor cantidad de pacientes de ambos grupos.

Si tratados y no tratados son muy similares, la zona de soporte común abarcará todos los individuos porque el máximo y el mínimo de ambos IP serán parecidos. Si ambos grupos son muy distintos, lo más probable es que el número de pacientes que entren en la zona de soporte común sea más reducido (fig. 5) (tabla 3).

Para ver si el orden de los sujetos en el método del emparejamiento influye, se han realizado (simulado) 100 modelos y se han obtenido las correspondientes estimaciones de las OR. Se ha observado que cuando el número de variables por las que se ajusta el IP es elevado, el orden de emparejamiento no influye y se obtienen estimaciones similares. Cuando el número de variables por las que se ajusta el orden de emparejamiento sí que influye, se obtienen estimaciones diferentes.

### Ventajas e inconvenientes de estas técnicas

Se ha comprobado que, dependiendo del número de eventos y de exposiciones que haya, el IP y la regresión logística podrían considerarse equivalentes, obteniendo valores en las estimaciones similares cuando no se observan

diferencias entre pacientes tratados y no tratados y el número de eventos es elevado.

Cuando el número de eventos es reducido, el IP permite ser ajustado por más variables que la regresión logística, y además, permite cuantificar el efecto del tratamiento.

Mediante el método del emparejamiento del IP, se obtiene una estructura semejante a la de un ensayo clínico (las parejas presentan características similares), con la desventaja de que aquellos pacientes que no encuentren pareja serán descartados del estudio.

Los ensayos clínicos aleatorizados presentan la ventaja de que, debido a la asignación aleatoria del paciente, se puede evaluar la relación causal de modo directo. Cuando estos estudios no son viables, los estudios observacionales pueden considerarse como alternativa. Para poder evaluar la causalidad en estudios observacionales, se deberá resolver primero el problema de la confusión que conllevan dichos estudios.

### Conclusiones

Los ensayos clínicos representan el mejor diseño metodológico para analizar la causalidad en investigación clínica. Sin embargo, existen hipótesis que no pueden ser testadas clínicamente por limitaciones éticas, metodológicas, económicas. Para resolver esta limitación los estudios observacionales pueden abordar este problema y simular el escenario hipotético de un ensayo clínico para poder dar respuesta, mediante la aplicación del IP y los modelos marginales estructurales. El interés creciente por esta nueva metodología hace necesaria su difusión por todos los intensivistas para mejorar el conocimiento de esta metodología



y su potencial aplicación en la investigación clínica de los Cuidados Intensivos.

Por ejemplo, Delaney et al.<sup>21</sup> muestran el uso de los métodos estadísticos detallados en este artículo para evaluar el efecto de los corticoides sobre la mortalidad en pacientes que presentan gripe A (H1N1pdm09). Las OR asociadas a dichos modelos varían desde 1,85 (IC 95%: 1,12-3,04) en los modelos clásicos de regresión logística multivariable; 1,71 (IC 95%: 1,05-2,78) en el modelo de regresión logística ajustando por el IP; 1,52 (IC 95%: 0,90-2,58) después del emparejamiento del IP, y 0,96 (IC 95%: 0,28-3,28) en el modelo marginal estructural ajustando por las variables tiempodependientes. Ajustando mediante las variables tiempodependientes en los modelos marginales estructurales, han observado que no existe asociación entre el uso de corticoides y la mortalidad de los pacientes.

### Conflicto de intereses

Los autores del manuscrito declaran no tener ningún conflicto de intereses ni económico, ni comercial ni intelectual en la redacción de este artículo.

### Bibliografía

- Grasselli G, Gattinoni L, Kavanagh B, Latini R, Laupacis A, Lemaire F, et al. Feasibility, limits and problems of clinical studies in Intensive Care Unit. *Minerva Anesthesiol.* 2007;73:595–601.
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol.* 1974;66:688–701.
- Rubin DB. For objective causal inference, design trumps analysis. *Ann Appl Stat.* 2008;2:808–40.
- Holmberg L, Baum M. Can results from clinical trials be generalized? *Nat Med.* 1995;1:734–6.
- Villar J, Pérez-Méndez L, Aguirre-Jaime A, Kackmarek RM. Why are physicians so skeptical about positive randomized controlled clinical trials in critical care medicine? *Intensive Care Med.* 2005;31:196–204.
- Austin P. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res.* 2011;46:399–424.
- Rubin DB. The design versus the analysis of observational studies for causal effects: Parallels with the design or randomized trials. *Stat Med.* 2007;26:20–36.
- Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics.* 1968;24:295–313.
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc.* 1984;79:516–24.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70:41–55.
- D'Agostino RB Jr, D'Agostino RB. Estimating treatment effects using observational data. *JAMA.* 2007;297:314–6.
- Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health.* 2000;21:121–45.
- Patrick AR, Schneeweiss S, Brookhart MA, Glynn RJ, Rothman KJ, Avorn J, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: An empirical illustration. *Pharmacoepidemiol Drug Saf.* 2011;20:551–9.
- Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163:1149–56.
- Pattanayak CW, Rubin DB, Zell ER. Propensity score methods for creating covariate balance in observational studies. *Rev Esp Cardiol.* 2011;64:897–903.
- Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, et al. A review of propensity-score methods and their use in cardiovascular research. *Can J Cardiol.* 2016;32:259–65.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15:361–87.
- Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol.* 2003;158:280–7.
- Muriel A, Penuelas O, Frutos-Vivar F, Arroliga AC, Abraira V, Thille AW, et al. Impact of sedative on outcomes of noninvasive ventilation. *Intensive Care Med.* 2015;41:1586–600.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11:550–60.
- Delaney JW, Pinto R, Long J, Lamontagne F, Adhikari NK, Kumar A, et al. The influence of corticosteroid treatment on the outcome of influenza A (H1N1pdm09)-related critical illness. *Critical Care.* 2016;20:75.
- Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168:656–64.
- Morris JV, Ramnarayan P, Parslow RC, Fleming SJ. Outcomes for children receiving noninvasive ventilation as the first-line mode of mechanical ventilation at intensive care admission: A propensity score-matched cohort study. *Crit Care Med.* 2017;45:1045–53.
- Rosa RG, Tonietto TF, Duso BA, Maccari JG, de Oliveira RP, Rutzen W, et al. Mortality of adult critically ill subjects with cancer. *Respir Care.* 2017;62:615–22.
- Peñuelas O, Muriel A, Frutos-Vivar F, Fan E, Raymondos K, Rios F, et al. Prediction and outcome of intensive care unit-acquired paresis. *J Intensive Care Med.* 2018;33:16–28, <http://dx.doi.org/10.1177/0885066616643529>.