



# Lectura crítica en pequeñas dosis

## Índices de propensión. El deseo de parecerse al ensayo clínico

M. Molina Arias

Publicado en Internet:  
17-marzo-2015

Manuel Molina Arias:  
mma1961@gmail.com

Servicio de Gastroenterología. Hospital Infantil Universitario La Paz. Grupo de Trabajo de Pediatría  
Basada en la Evidencia AEP/AEPap. Madrid, España. Editor de [www.cienciasinseso.com](http://www.cienciasinseso.com)

### Resumen

#### Palabras clave:

- Aleatorización
- Índice de propensión
- Regresión logística

Los estudios observacionales carecen de aleatorización. Por este motivo, no existe la seguridad de que los diferentes grupos del estudio tengan la misma distribución de covariables basales, lo que tiene como consecuencia que no puedan compararse los resultados de forma directa sin llevar a cabo alguna forma de ajuste para variables de confusión o modificadoras de efecto.

En el presente trabajo se describe el uso de los índices de propensión (*propensity score*) como método de ajuste alternativo al uso clásico del ajuste por regresión logística.

## Propensity score analysis. The desire to resemble clinical trials

### Abstract

#### Key words:

- Randomization
- Propensity score
- Logistic regression

Observational studies lack randomization. Therefore, there is no assurance that the different study groups have the same distribution of baseline covariates, which has the consequence that results cannot be compared directly without performing some form of adjustment for confounding or modifying effect.

In this paper the use of propensity scores is described as an alternative method to the classic use of adjustment by logistic regression.

## ÍNDICES DE PROPENSIÓN

El ensayo clínico aleatorizado se considera el diseño más idóneo para estimar el efecto de las intervenciones sanitarias. Además, es el único tipo de estudio que permite establecer una relación de causalidad entre exposición o intervención y resultado<sup>1</sup>.

Esto es en gran parte gracias a uno de los aspectos clave del ensayo clínico: la aleatorización. Mediante la asignación aleatoria de los participantes al grupo de intervención o al de control se consigue que los dos grupos sean homogéneos en la distribución de sus características basales. Esto es así no solo para las covariables conocidas por el investigador, que se sabe pueden influir en el resultado,

Cómo citar este artículo: Molina Arias M. Índices de propensión. El deseo de parecerse al ensayo clínico. Rev Pediatr Aten Primaria.2015;17:87-90.

sino también para las desconocidas. De esta forma, cualquier diferencia que pueda observarse en los resultados entre los dos grupos se deberá probablemente a la intervención en estudio.

Por su parte, los estudios observacionales carecen de aleatorización, por lo que los participantes de los diferentes grupos pueden diferir sistemáticamente en algunas de sus características basales. Esta es la razón por la que los resultados en los dos grupos de un ensayo clínico pueden compararse directamente, mientras que en los estudios observacionales hay que tener en cuenta estas posibles diferencias y realizar un ajuste por posibles variables de confusión o modificadoras de efecto. Además, esta falta de aleatorización y el riesgo inevitable de sesgo impide a los estudios observacionales lo que solo está permitido a los ensayos clínicos: establecer relaciones causales entre exposición y efecto.

Por estos motivos no es de extrañar que se desarrollen métodos para controlar en lo posible la influencia de variables de confusión. Quizás el más conocido de todos sea el ajuste mediante un modelo de regresión logística, en el que las variables que pueden influir en los resultados se utilizan como predictores del modelo y la variable principal de resultado como regresor.

Sin embargo, gracias a la potencia de cálculo de los paquetes estadísticos y ordenadores actuales, cada vez vemos con más frecuencia otro método de ajuste, del que nos vamos a ocupar en el presente trabajo: el uso de índices de propensión, más conocidos por su nombre en inglés, *propensity score*. Veamos qué son y cómo se utilizan.

El índice de propensión (IP) es la probabilidad que tiene cada participante del estudio de ser asignado a cada una de las ramas del estudio en base a sus características basales<sup>2,3</sup>. Gracias a la asignación aleatoria, esta probabilidad es fija y conocida en el caso de los ensayos clínicos. Sin embargo, en el caso de estudios observacionales, es desconocida y debe calcularse a partir de los datos disponibles. Vamos a entenderlo mejor analizando la metodología de un estudio de cohortes retrospectivo sobre el uso de antibióticos empíricos de espectro

amplio o reducido en niños con neumonía extrahospitalaria<sup>4</sup>, cuya valoración crítica ha sido también publicada recientemente<sup>5</sup>.

Los autores diferencian dos cohortes según el tipo de antibiótico administrado y comparan, entre otras variables de resultado, la duración de la estancia hospitalaria. Si comparásemos las estancias de los dos grupos, como haríamos en un ensayo clínico, correríamos el riesgo de obtener un resultado sesgado. ¿Por qué? Pensemos un poco cómo hemos asignado los tratamientos.

El hecho de que un niño reciba antibiótico de espectro amplio o reducido puede depender de la edad, del tiempo de evolución, de los tratamientos previos, etc. Así, puede ocurrir que sistemáticamente haya más tendencia a tratar los más graves con amplio espectro y los más leves con espectro reducido. Como esto estuvo a criterio de cada médico que atendió a cada niño, tenemos que tener la precaución de comparar los niños teniendo en cuenta todos estos factores.

En otras palabras, tenemos que comparar entre sí participantes que tengan una probabilidad similar de recibir uno u otro tratamiento o, lo que es lo mismo, con un IP similar.

Los IP se calculan, de forma más habitual, construyendo un modelo de regresión logística con la intervención como resultado (regresor) y con las covariables basales como predictoras (en el ajuste clásico de los factores de confusión por regresión logística, el regresor no es la intervención, sino la variable principal de resultado). Podemos introducir todas las covariables en el modelo, aunque este es más eficiente si introducimos solo aquellas que creamos que pueden tener efecto confusor en el resultado o en la probabilidad de recibir una u otra intervención. En este modelo de regresión no hay problema de exceso de ajuste por introducir muchos predictores, ya que el propósito es ajustar el modelo lo máximo posible y no predecir datos nuevos a partir de los existentes.

Así, el modelo de regresión nos proporciona la probabilidad de recibir el tratamiento según el valor de las covariables de cada participante (en realidad,

proporciona el logaritmo de la *odds ratio*, pero podemos usar este valor directamente o calcular la probabilidad a partir de él).

En el ejemplo que nos ocupa, los autores construyen su modelo empleando como predictores edad, sexo, raza, financiación, asma, fiebre y taquipnea al inicio, antibióticos previos, diagnóstico de viriasis de vías bajas, ingreso en Unidad de Cuidados Intensivos Pediátricos (UCIP), realización y resultado de hemocultivo, hemograma anormal y nivel basal hospitalario de uso de cefalosporinas.

Es importante, para que los grupos puedan ser comparables, que exista solapamiento de sus IP (habla de la homogeneidad de los grupos), hecho que los autores de nuestro ejemplo comprueban y reflejan en su trabajo.

Una vez que cada participante tiene su IP, debemos agruparlos para poder comparar los resultados. Los métodos más habituales son el de emparejamiento y el de estratificación. Con menos frecuencia en Medicina se usan otros dos métodos más complejos y que no trataremos aquí: el ajuste de covariables por IP y la ponderación por el inverso de la probabilidad de recibir la intervención<sup>6</sup>.

Mediante el **emparejamiento** se hacen conjuntos (generalmente parejas) de sujetos de ambos grupos con IP similares. Esto puede hacerse mediante la estrategia del “vecino más próximo” (la empleada en nuestro ejemplo), en la que se busca uno de un grupo y otro cuyo valor sea cercano (hay que definir cuál es el límite de cercanía) o buscando que las diferencias de los IP de todas las parejas sea la mínima posible, lo que se conoce como emparejamiento óptimo<sup>3,7</sup>.

Si no existe con quién emparejar, el sujeto no se considera a la hora de comparar los resultados. Esto puede ser un problema si ocurre con frecuencia, ya que el tamaño muestral efectivo para hacer la comparación disminuye y con ello puede disminuir la potencia del estudio. En estos casos puede ser conveniente imputar los datos faltantes mediante cualquiera de los procedimientos habituales<sup>8,9</sup>.

Una vez emparejados, podemos comparar los resultados, generalmente mediante una *t* de Student

para datos pareados (variables continuas) o mediante prueba de McNemar (variables dicotómicas). De aquí se deriva otra ventaja de este método, que es que permite utilizar la misma métrica que emplearíamos en un ensayo: diferencias de medias o reducciones absolutas de riesgos con su inverso, el número necesario a tratar. Esto es más complejo si realizamos el ajuste mediante regresión logística en lugar de con IP.

Mediante la **estratificación** los sujetos se agrupan en subgrupos mutuamente excluyentes en base a sus IP. Lo habitual es ordenar la serie y agruparlos en cinco grupos iguales en base a los quintiles. Dentro de cada estrato, los sujetos de ambos grupos tendrán IP similares, reflejo de una distribución similar de las covariables basales.

El análisis estadístico que debe realizarse con los estratos es similar al empleado en los metaanálisis (test de Mantel-Haenszel), calculándose las diferencias de efecto entre los dos grupos a nivel de estratos y a nivel global.

En nuestro ejemplo de la neumonía extrahospitalaria, se obtuvo como resultado una mayor estancia hospitalaria en los que recibieron tratamiento de amplio espectro. Gracias al método de ajuste mediante IP podemos estar razonablemente seguros que esto no es debido a diferencias basales de gravedad o de otras características de los sujetos del estudio y concluir que el usar antibióticos de amplio espectro no parece acortar la duración del ingreso.

Como reflexión final, el uso de los IP permite a los estudios observacionales convertirse en “casi” estudios experimentales. Este método es más conservador que la regresión logística y es especialmente eficiente en enfermedades poco frecuentes donde existen pocos casos y muchas covariables. No obstante, por mucho que nos empeñemos, nunca podremos tener la seguridad que nos proporciona la aleatorización de los ensayos de que no existan variables desconocidas sin ajustar que puedan sesgarnos los resultados. Este es el motivo de que sea aconsejable que las conclusiones del estudio observacional se confirmen posteriormente con el ensayo clínico aleatorizado correspondiente.

## CONFLICTO DE INTERESES

---

El autor declara no presentar conflictos de intereses en relación con la preparación y publicación de este artículo.

## ABREVIATURAS

---

**IP:** índice de propensión • **UCIP:** Unidad de Cuidados Intensivos Pediátricos.

## BIBLIOGRAFÍA

---

1. Molina Arias M, Ochoa Sangrador C. Ensayo clínico (I). Definición. Tipos. Estudios cuasiexperimentales. *Evid Pediatr.* 2014;10:52.
2. Sainani KL. Propensity scores: uses and limitations. *PMR.* 2012;4:693-7.
3. Austin PC. An introduction to propensity score methods for reducing the effects of confounding. In observational studies. *Multivariate Behav Res.* 2011; 46:399-424.
4. Queen MA, Myers AL, Hall M, Shah SS, Williams DJ, Auger KA, *et al.* Comparative effectiveness of empiric antibiotic for community-acquired pneumonia. *Pediatrics.* 2014;133:e23.
5. Francisco González L, Ochoa Sangrador C. Los antibióticos de amplio espectro no mejoran el tratamiento de la neumonía comunitaria. *Evid Pediatr.* 2014;10:65.
6. Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. *Statist Med.* 2010;29:2137-48.
7. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statist Med.* 2009;28:3083-107.
8. Patrician PA. Multiple imputation for missing data. *Res Nurs Health.* 2002;25:76-84.
9. McCleary L. Using multiple imputation for analysis of incomplete data in clinical research. *Nurs Res.* 2002;51:339-43.