

DISEÑOS OBSERVACIONALES: UNA APLICACIÓN EN ESTUDIOS CLÍNICOS

OBSERVATIONAL DESIGNS: AN APPLICATION TO CLINICAL STUDIES

FERNANDO GIMÉNEZ SENA*

*Departamento de Estadística, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Asunción, Paraguay. Email: ferdisena@facen.una.py

Resumen: El diseño es una componente esencial en los estudios estadísticos. Los ensayos clínicos constituyen el estándar de oro pues utilizan el concepto de aleatorización para balancear efectos de variables observadas y no observadas sobre la respuesta de interés, aunque esto no siempre es posible. Una solución posible son los diseños observacionales, que buscan reproducir las condiciones de un estudio aleatorizado, sin la aleatorización. Aspectos principales son el matching y la sensibilidad. El primero busca balancear los posibles efectos de las variables observadas, y el segundo intenta cuantificar la magnitud causal necesaria para invalidar los resultados obtenidos. Estos conceptos son aplicados en un pequeño ensayo clínico exploratorio para evaluar sus cualidades bajo la perspectiva de un estudio observacional. Los resultados muestran que una elección de pares semejantes mejora la comparabilidad entre los grupos de tratamiento y de control. Las conclusiones son altamente sensibles a pequeños cambios que pueden originarse como efecto de variables no observadas.

Palabras claves: *Estudios Observacionales, Matching, Pares semejantes, Análisis de Sensibilidad.*

Summary: Design is a crucial component in statistical studies. In biostatistics, the clinical trials are the gold standard, because the use of randomization to balance either observed and non observed covariates effects on the response. But random assigment of treatment to subjects not always is possible. Observational studies is an alternative, but it lacks randomization. Two key concepts are, matching to balance effects of observed covariates, and sensitivity analysis to measure the amount of the hidden effect that would make conclusions invalid. These concepts are applied to a small exploratory clinical trial in order to evaluate its characteristics under the observational study framework. Results show that it is possible to improve comparability by matching pairs, and conclusions are highly sensitive to small changes due to unobserved covariates.

Keywords: *Observational studies, Matching, Matching pairs, Sensitivity analysis.*

INTRODUCCIÓN

Numerosas instituciones, públicas y privadas, como resultado de sus operaciones, generan una gran cantidad de datos que terminan almacenados en sus archivos, sin extraer la información útil que pueden proporcionar. Ejemplo de tales instituciones son los prestadores de salud, como los hospitales, sanatorios o centros especializados; en estos lugares, las historias clínicas de los pacientes registran numerosas variables relativas a su estado de salud, a los tratamientos que les fueron administrados y a los resultados obtenidos.

El análisis estadístico formal de estos datos suele presentar dificultades, porque no cumplen, en general, con el requisito de asignación aleatoria de tratamientos a los pacientes. Debido a esta falta de aleatoriedad, los grupos potencialmente pierden

comparabilidad; por esta razón, son necesarios ajustes para prevenir que las conclusiones se vean afectadas por sesgos ocultos que no han sido tenidos en cuenta. Esta asignación al azar es característica en los diseños experimentales, cuando esto no es posible, queda la posibilidad de realizar estudios observacionales o quasi-experimentos, en los que se intenta recuperar de algún modo las condiciones de un estudio experimental, incorporando además, procedimientos estadísticos para controlar y cuantificar el grado de distorsión de los resultados obtenidos, que pueden occasionar las variables no observadas.

Por otra parte, si no se tienen en cuenta las limitaciones de los estudios observacionales, la única alternativa válida de análisis se reduce, en general, a los de tipo descriptivo, en los que la información es proporcionada en tablas, cuadros, gráficos y las

medidas descriptivas usuales. Esta situación es satisfactoria cuando no se tiene mayores pretensiones con relación al conocimiento obtenido; pero, si el interés está centrado en conclusiones que orienten la toma de decisiones o en estudiar asociaciones entre variables, factores y respuestas, con la expectativa de encontrar relaciones de causa-efecto, es fundamental contar con herramientas más potentes y adecuadas para obtener la información disponible en los datos.

El objetivo general de este trabajo es explorar las características generales de los estudios observacionales como alternativa válida a los diseños experimentales.

Más específicamente, el objetivo general es desagregado en los dos aspectos siguientes:

(a) Revisar los conceptos de matching y sensibilidad del análisis en un estudio observacional de pares semejantes y (b) Aplicar la teoría revisada para evaluar un estudio clínico piloto real.

El desarrollo es como sigue: Primero se revisa el marco de referencia proporcionado por los estudios experimentales; específicamente, se revisan aspectos relacionados con los diseños de pares semejantes. A continuación, se desarrolla el estudio observacional, junto con sus dos características esenciales, el Matching y la Sensibilidad del Diseño. Finalmente, la teoría de los estudios observacionales, que ha sido revisada, es aplicada para evaluar un estudio experimental piloto para comparar dos tratamientos.

MATERIALES Y MÉTODOS

Estudios de Pares semejantes en general

Los conceptos que siguen han sido desarrollados en Bloom (2006), Greenland (2004), Pirrachio et al. (2012), Rosenbaum (1989), (2002), (2004), (2005 a, b), (2007) y (2010), Rosenbaum y Rubin (1983), en Rubin (2006 a, b) y (2008), Sobel (2006) y Zubizarreta et al. (2011).

El término diseño experimental abarca una gran cantidad de esquemas utilizados para la recopilación de datos. La elección de un esquema particular depende de factores como: el objetivo del estudio, la estructura de la población a estudiar, el tipo de

efecto estudiado y su variabilidad, la validez de los resultados obtenidos, aspectos financieros y consideraciones éticas, entre otros.

Si la respuesta estudiada se espera observar luego de un breve tiempo posterior a la intervención y tanto la enfermedad como el tratamiento tienen poca duración, el diseño pareado y aleatorizado ha demostrado ser un esquema eficiente, el cual es conceptualizado a continuación.

Concepto

En un estudio pareado se toma como unidad de estudio un par de individuos, cada uno identificado como (i, j) ; i denota al par, siendo $i = 1, 2, \dots, I$; similarmente, j identifica al individuo dentro del par, con $j = 1, 2$.

Si Z es la variable que identifica la aplicación, o no, del tratamiento; para el sujeto (i, j) resulta, $Z_{ij} = 1$ si la unidad (i, j) recibe el tratamiento; en caso que reciba el control, entonces $Z_{ij} = 0$; que expresando como un vector toma la forma, $\mathbf{Z} = (Z_{11}, Z_{12}, \dots, Z_{I1}, Z_{I2})^T$. Se observa que $Z_{i2} = 1 - Z_{i1}$.

Sea ahora \mathbf{x} , el vector de variables observadas. Identificando todos sus elementos, se escribe como $\mathbf{x} = (x_1, x_2, \dots, x_k)^T$. El elemento, x_k , denota la k -ésima variable observada en el vector, $k = 1, 2, \dots, K$. El vector de variables toma la forma $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijk})^T$ para indicar las variables asociadas con la unidad (i, j) ; la correspondiente k -ésima variable observada es $X_{i,j,k}$.

Sea R , la respuesta observada. En un par específico, R_{ij} indica la respuesta del (i, j) -ésimo individuo; todas las respuestas se presentan con el vector, $\mathbf{R} = (R_{11}, R_{12}, \dots, R_{I1}, R_{I2})^T$.

Cada respuesta puede representarse como,

$$Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2}) \quad (1)$$

Por otra parte, además de las variables efectivamente observadas, es posible que existan otras no observadas.

Sea u_{ij} , la variable no observada asociada a la unidad (i, j) . Todas las unidades se representan como: $\mathbf{u} = (u_{11}, u_{12}, \dots, u_{I1}, u_{I2})$.

Sin perder generalidad, $j = 1$ identifica siempre

al individuo que recibe el tratamiento y $j = 2$ al individuo que recibe el control. En la práctica, esta asignación es aleatoria.

En un estudio experimental se requiere que los individuos dentro de cada par sean lo más similares posibles. Esta similitud es determinada por medio del matching basado en las variables observadas.

Estadísticos y su distribución aleatorizada

Dentro de cada par, un individuo recibe el tratamiento aleatoriamente. Bajo este esquema, un individuo en el par i tiene probabilidad de recibir tratamiento igual a,

$$\Pr(Z_{i1} = 1) = \Pr(Z_{i2} = 0) = 1/2.$$

Con I posibles pares, y como cada posición puede ocuparse de dos formas posibles, la probabilidad de ocurrir un par específico es entonces, $1/2^I$.

Dependiendo de la asignación del tratamiento a cada individuo, la ecuación 1 tendrá signo positivo o negativo; entonces, esta ecuación proporciona el sentido de la diferencia entre las respuestas de los individuos del par.

En un diseño pareado, cada individuo del par recibe aleatoriamente el tratamiento o el control. Para I pares, se tienen 2^I posibles asignaciones.

Para contrastar la plausibilidad de la hipótesis nula, es necesario un estadístico que proporcione una estimación del efecto del tratamiento, que junto con su distribución muestral, servirán de base para realizar la prueba.

Se consideran: (1) El promedio de las diferencias de respuestas en cada par y (2) el estadístico Rangos con Signo de Wilcoxon.

Distribución aleatorizada del promedio de las diferencias

Sean, \mathcal{F} el conjunto de todos los valores de $(r_p, r_c, \mathbf{x}, \mathbf{u})$ para todos los sujetos y \mathcal{X} el conjunto de todos los s^l posibles valores que puede tomar $\mathbf{Z} = (Z_{11}, Z_{12}, \dots, Z_{I1}, Z_{I2})^T$.

Se define entonces, $\mathbf{A} = (\mathcal{F}, \mathcal{X})$.

Suponiendo que la hipótesis nula es cierta, el

promedio de las diferencias,

$$\bar{Y} = \frac{1}{I} \sum_{i=1}^I Y_i$$

tiene distribución para una cola, $\Pr(\bar{Y} \geq y | \mathbf{A})$.

Para dos colas, la distribución del promedio de las diferencias tiene forma,

$$2 \Pr(\bar{Y} \geq y | \mathbf{A}) = \Pr(\bar{Y} \leq -y | \mathbf{A}) + \Pr(\bar{Y} \geq y | \mathbf{A}),$$

que es simétrica alrededor de cero.

Según R. A. Fisher, la aleatorización proporciona una base razonable para realizar inferencias causales.

Si los datos no tienen buen comportamiento, el promedio de las diferencias es uno de los estadísticos menos eficientes; un estadístico robusto alternativo para este caso, es el estadístico Rangos con Signo de Wilcoxon.

Rangos con Signo de Wilcoxon.

Este estadístico se define de la siguiente manera:

Se tienen $i = 1, 2, \dots, I$ pares de individuos.

Dentro de un par, cada individuo es identificado por $j = 1$ y $j = 2$. Entonces, (i, j) identifica al j -ésimo sujeto en el i -ésimo par.

El estadístico de prueba es,

$$T = \sum_{i=1}^I \text{sgn}(Y_i) q_i$$

donde $\text{sgn}(Y_i) = 1$ si $Y_i > 0$, mientras que $\text{gn}(Y_i) = 0$ si $Y_i \leq 0$; $q_i = \text{rango}|Y_i|$; tipificando el estadístico produce,

$$\frac{T - E(T | \mathbf{A})}{\sqrt{\text{var}(T | \mathbf{A})}} \rightarrow N(0, 1)$$

cuando la cantidad de pares $I \rightarrow \infty$.

Emparejamiento y Sensibilidad

Según Cochran (1965), un estudio observacional es una investigación empírica en el cual,

“... el objetivo es dilucidar relaciones de causa y efecto ... (y donde) no es factible utilizar el experimento controlado, en el sentido de

que, sea posible imponer los procedimientos o tratamientos cuyos efectos se desea descubrir; o asignar sujetos de forma aleatoria a los diferentes procedimientos”.

El parecido con un diseño experimental radica en que se trabaja con tratamientos, intervenciones o políticas y se evalúan los efectos que los mismos causan. Un estudio en el que no existe tratamiento, no es experimental ni observacional. En un estudio observacional, no hay control de la asignación de tratamiento a los individuos, por lo que no es posible asegurar que sujetos similares reciban distintos tratamientos, Rosenbaum (2002).

Por múltiples razones puede no ser posible controlar la asignación de tratamiento; por ejemplo, porque el tratamiento es perjudicial o costoso, o una intervención posiblemente está condicionada por motivos políticos o socioeconómicos, o existan restricciones éticas para su aplicación en forma aleatoria, etc.

Un estudio observacional es sesgado si los grupos de tratamiento y de control son diferentes antes del tratamiento, de forma que esta diferencia tiene importancia en la respuesta observada. El sesgo se llama abierto cuando es evidente en los datos reco-
pilados; en contrario, el sesgo es oculto cuando los datos no lo evidencian, debido a que la información necesaria no ha sido observada o registrada.

El sesgo abierto es controlado recurriendo al matching o a la estratificación y se inicia antes de diseñar el estudio. Por otra parte, por medio del estudio de la sensibilidad del diseño se investigan los posibles efectos del sesgo oculto. Estos dos conceptos son revisados a continuación.

Matching

Concepto

Es una técnica utilizada para conseguir que dos o más grupos de estudio sean semejantes.

En el contexto de estudios clínicos, permite que los grupos experimental y control sean comparables. Esta comparabilidad es factible por medio de emparejamientos individuales, 1-a-1, alternativamente son posibles utilizar, emparejamientos 1:n (1

tratamiento con n controles) o n:1 (n tratados con 1 control), o emparejamientos de relación variable y emparejamientos grupo a grupo de individuos.

Si es la única técnica utilizada para lograr la comparabilidad, puede resultar inadecuada para sustituir la aleatorización como medio de balance entre los individuos.

Emparejar los individuos del grupo tratado con individuos del grupo de comparación, significa igualarlos en base a la similitud de ciertas variables. La tarea es relativamente sencilla cuando la cantidad de variables es reducida.

Como los emparejamientos son factibles utilizando unas pocas variables, puede que no sea posible tener en cuenta todas las diferencias que potencialmente podrían distorsionar de manera significativa los resultados.

La ausencia de variables importantes en el análisis da lugar a dos tipos de distorsión. La forma más simple se denomina "confusión directa", que aparece cuando existe una variable subyacente que explica la variable y la respuesta al mismo tiempo. La segunda ocurre cuando la relación entre la variable y la respuesta es a través de una variable intermedia, Cox y Wermuth (2007).

Población antes del matching

Antes del matching existen en la población L individuos. Cada individuo l , tiene asociado una variable observada \mathbf{x}_l y una variable no observada \mathbf{u}_l , con $l = 1, \dots, L$.

La variable indicadora Z_l , con $Z_l = 1$ señala que la unidad l ha recibido el tratamiento de interés, con probabilidad π_l , independiente de otros individuos. Cuando $Z_l = 0$, se indica que el sujeto recibe el tratamiento control con probabilidad de asignación igual a $1 - \pi_l$.

Se tiene también que r_{nl} designa la respuesta que potencialmente se observaría en un individuo si recibiera tratamiento, mientras que r_{cl} denota la respuesta potencial si el individuo hubiera recibido el control. La respuesta observada se designa por $R_l = Z_l \times r_{nl} + (1 - Z_l) \times r_{cl}$

La probabilidad que un sujeto reciba el tratamiento de interés está dada por,

$$\pi_i = \Pr(Z_i | r_{T_P}, r_{C_P}, \mathbf{x}_i, \mathbf{u}_i).$$

Que resulta en,

$$\begin{aligned} \Pr(Z_1 = z_1, \dots, Z_L = z_L | r_{T_1}, r_{C_1}, \mathbf{x}_1, \mathbf{u}_1, \dots, r_{T_L}, r_{C_L}, \\ \mathbf{x}_L, \mathbf{u}_L) = \pi_1^{z_1} (1 - \pi_1)^{1-z_1} \times \dots \times \pi_L^{z_L} (1 - \pi_L)^{1-z_L} \\ = \prod_{l=1}^L \pi_l^{z_l} (1 - \pi_l)^{1-z_l}. \end{aligned}$$

Matching ideal

Si existen dos individuos k y l , uno de los cuales recibe el tratamiento, con $\pi_k = \pi_l$ se tiene que $Z_k + Z_l = 1$. La formación de un par semejante es complejo, pues \mathbf{u}_k y \mathbf{u}_l no son observables. Sólo se registra r_{T_k} o r_{T_l} y r_{C_k} o r_{C_l} por lo que únicamente es posible emparejar en base a las variables observadas \mathbf{x} , de forma que $\mathbf{X}_k = \mathbf{X}_l$. Finalmente se determina (Z_k, Z_l) por asignación aleatoria de los tratamientos.

Si la asignación aleatoria no es posible, como ocurre en los estudios observacionales, se evidencia el problema de encontrar pares de individuos que tengan la misma probabilidad de asignación de tratamientos. Se buscan entonces, condiciones en las que estas probabilidades sean aproximadamente iguales.

Si existen dos sujetos para los que $\pi_k = \pi_l$ Rosenbaum (2010) muestra que,

$$\begin{aligned} \Pr(Z_k = 1, Z_l = 0 | r_{T_k}, r_{C_k}, \mathbf{x}_k, \mathbf{u}_k, r_{T_l}, r_{C_l}, \mathbf{x}_l, \mathbf{u}_l) = \\ \pi_k^{z_k} (1 - \pi_k)^{1-z_k} \times \pi_l^{z_l} (1 - \pi_l)^{1-z_l} = \\ \pi_l^{z_k+z_l} \times (1 - \pi_l)^{(1-z_k)+(1-z_l)} \end{aligned}$$

y si además, exactamente uno de los individuos recibió el tratamiento de interés, (esto es $Z_k + Z_l = 1$), resulta,

$$\Pr(Z_k = 1, Z_l = 0 | r_{T_k}, r_{C_k}, \mathbf{x}_k, \mathbf{u}_k, r_{T_l}, r_{C_l}, \mathbf{x}_l, \mathbf{u}_l; Z_k + Z_l = 1) = \frac{1}{2}$$

es decir, si es posible encontrar en el conjunto de L individuos, un número de $2I$ sujetos diferentes y emparejados en I pares, entonces se podrá reconstruir la distribución de las asignaciones de tratamientos

Z. En consecuencia, será posible aplicar el análisis de un experimento aleatorizado.

El matching con una sola variable en general produce grupos de tratamientos y control que en conjunto están balanceados con relación a variables observadas. Puede ocurrir que pares individuales cercanos, difieran ampliamente en algunas variables específicas. Para la construcción de grupos más próximos se recurre a medidas que cuantifiquen las diferencias entre los individuos.

Estas medidas penalizan las diferencias grandes, y permiten la búsqueda de individuos que sean en cierto sentido cercanos. Los conjuntos similares, en particular pares similares, son construidos utilizando un algoritmo de optimización. El emparejamiento de los individuos de un par puede tener impacto en la distribución de las variables. Por ejemplo, conducir a la pérdida de información cuando la cantidad de unidades tratadas es diferente a la cantidad de controles utilizados.

Según Rubin (1973), hay un límite en el beneficio que puede conseguirse con el matching, aunque esto puede ser resuelto utilizando otros esquemas para hacer que los individuos sean semejantes.

El uso de la aleatorización para obtener el balance en las variables, es mucho más potente que el matching obtenido con las probabilidades de asignación de tratamientos. Esto se debe a que, el primero balancea tanto variables observadas y variables no observadas, mientras que el segundo sólo actúa sobre las observadas.

Puntaje de propensión

El puntaje de propensión es una herramienta para realizar el matching, es utilizado como una medida de la similitud entre dos individuos.

El puntaje asigna un número a cada individuo que servirá de criterio para construir los emparejamientos de sujetos. Las variables del estudio observacional son colapsadas en un único número que resume su asociación conjunta con las condiciones del tratamiento, Hansen (2008).

En general, el puntaje es aplicado a tratamientos de tipo binario. Hirano e Imbens (2004) hacen extensiones para el caso en que el tratamiento es

continuo.

Definición

Con una muestra de tamaño L de una población infinita, el Puntaje de Propensión es la probabilidad condicional de recibir el tratamiento dada las variables observadas, esto es:

$$e(\mathbf{x}) = Pr(Z_i = 1 | \mathbf{x}) \quad (2)$$

En un ensayo aleatorio controlado, la probabilidad de asignación del tratamiento a un individuo es conocida. En contrario, en un estudio observacional, esta probabilidad es desconocida; además, su estimación es compleja pues depende del conjunto de variables observadas, así como de las variables no observadas.

Propiedades

1) Balance

Individuos tratados e individuos de control que tienen el mismo puntaje de propensión, tienen la misma distribución en las variables observables \mathbf{x} .

$$Pr(Z = 1 | \mathbf{x}; e(\mathbf{x})) = Pr(Z = 0 | \mathbf{x}; e(\mathbf{x})).$$

2) Matching Ideal

Bajo ciertas condiciones, permite obtener un "matching ideal". Si el modelo en el cual se supone que los individuos son comparables, es efectivamente cierto, entonces el matching ideal puede conseguirse utilizando solamente un indicador, denotado por $e(\mathbf{x})$.

Por Matching Ideal se entiende que, los individuos se asemejan con respecto a la probabilidad de asignación del tratamiento a un individuo.

La estimación del puntaje de propensión, denotado por $\hat{e}(\mathbf{x})$, se basa en algún modelo que relaciona la probabilidad de asignación de tratamientos con las variables observadas. Un modelo común es el Logit, aunque es posible utilizar otros métodos. También es posible utilizar la estratificación como una alternativa al matching como herramienta para

asemejar.

Cochran (1965), afirma que utilizando cinco estratos, formados a partir de una variable continua, se puede remover el 90% del sesgo presente en una variable.

Matrices distancia

La Matriz Distancia es otra herramienta que sirve para determinar la similitud entre individuos. Cada fila de la matriz corresponde a un individuo que recibe el tratamiento y cada columna está asociada con un individuo que potencialmente es un control. Esta distancia se define como,

$$d_{ik} = [\hat{e}(\mathbf{x}_l) - \hat{e}(\mathbf{x}_k)]^2.$$

Para tomar en consideración las posibles diferencias en variables se utiliza un "calibrador", denotado por w , de tal modo que para dos individuos, k y l con \mathbf{x}_k y \mathbf{x}_l , respectivamente, con el calibrador w se tiene que,

- si $|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| > w$, entonces se fija la distancia entre ambos en ∞ .
- si $|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| \leq w$, entonces se fija la distancia entre ambos en d_{ik} , que es una medida de la proximidad entre ambos individuos.

El calibrador es frecuentemente definido como un múltiplo de la desviación estándar del puntaje de propensión. Modificando los valores del multiplicador se obtienen diferentes calibraciones de la importancia relativa del puntaje de propensión y de las variables.

Un criterio sugerido por Rosenbaum (2010), consiste en iniciar con una amplitud de 20% del error estándar del puntaje de propensión e ir ajustando según la necesidad, hasta obtener un balance satisfactorio.

Gu y Rosenbaum (1993) expresan que, igualando con el puntaje de propensión en general es suficiente. Sin embargo, cada par individual puede ser muy distinto en términos de las variables.

Por otra parte, el matching dentro de intervalos de puntajes estrechos de calibración también tienden a balancear la distribución de variables, proporcionando el beneficio extra de pares indivi-

duales más cercanos.

La Distancia de Mahalanobis es una generalización de la distancia al caso de varias variables. La diferencia en desvío estándar pesa igual para cada variable en el vector \mathbf{x} . Además, toma en consideración la correlación entre variables, esto es, dos variables altamente correlacionadas son consideradas como una sola por la distancia de Mahalanobis.

Si $\hat{\Sigma}$ es la matriz de covarianzas muestrales de \mathbf{x} , entonces la distancia de Mahalanobis para un par de individuos k y l se define como,

$$D_{lk} = (\mathbf{x}_k - \mathbf{x}_l)^T \hat{\Sigma}^{-1} (\mathbf{x}_k - \mathbf{x}_l)$$

La Distancia de Mahalanobis fue diseñada para datos normales multivariados, por lo que su utilización con datos no normales puede producir resultados inconsistentes.

Una modificación para datos no normales recibe el nombre de Distancia de Mahalanobis basada en rangos.

Tipos básicos de matching

Según Rosenbaum (2010), el matching ideal es el más conveniente pues produce resultados equivalentes a un diseño aleatorizado, sin embargo, no es posible estar completamente seguros de que en la práctica se cumplan las condiciones necesarias para tal matching.

A continuación se describen brevemente los conceptos de: (a) matching óptimo de pares, (b) matching óptimo con múltiples controles y (c) matching óptimo completo.

Matching óptimo de pares

En este tipo de matching, cada individuo tratado es emparejado con un control distinto con el objeto de minimizar la distancia total dentro del par. Alternativamente, se utiliza una Función de Penalización que asigna un valor a los pares que violan la restricción determinada por el calibrador.

Rosenbaum (2010) propone,

$$1000 \times \max(0, |\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| - w)$$

por lo que si,

$$|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| \leq w$$

entonces la penalización es cero.

Por otra parte, si

$$|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| > w$$

entonces la penalización está dada por:

$$1000 \times (|\hat{e}(\mathbf{x}_k) - \hat{e}(\mathbf{x}_l)| - w).$$

La penalización es agregada a la distancia de Mahalanobis o a la distancia de Mahalanobis basada en rangos en el valor correspondiente al par.

Matching óptimo con múltiples controles

En este caso, cada sujeto tratado es emparejado con al menos un, posiblemente con varios, control. Emparejar con relación fija significa que cada individuo tratado tiene el mismo número de controles asociados; mientras que un matching con relación variable permite que cada sujeto tratado pueda ser emparejado con un número diferente de controles.

El matching con relación variable es ligeramente más complejo, pues es necesario determinar quienes son emparejados y también cuantos serán en cada emparejamiento. Una estrategia útil es imponer restricciones en la elección de los controles.

Matching óptimo completo

Con este procedimiento, a diferencia de la asemejación con un número variable de controles, el esquema permite que un control sea emparejado con varios sujetos tratados. Este matching puede resultar mucho mejor que el matching óptimo de pares o el matching con un número variable de controles. Como el matching completo tiene como casos especiales los procedimientos previamente vistos, los conjuntos similares serán al menos tan buenos como ellos.

Los conjuntos resultantes pueden tener distintos tamaños, por lo que es posible la existencia de ciertas ineficiencias.

En un sentido específico, este esquema es un diseño óptimo para los estudios observacionales. Define una estratificación que es una partición de los elementos en términos de las variables y con el requerimiento de que cada estrato debe contener, al menos, una persona tratada y un control.

Comprobación del balance

Una vez obtenido el grupo de controles para los emparejamientos, es importante verificar si se ha obtenido el balance en las variables observadas. En caso que algunas variables no hayan sido balanceadas, es posible realizar algunos ajustes para lograr el balance.

Rosenbaum (2010) señala tres procedimientos: el matching casi exacto, el matching exacto y el balance fino.

Una modificación de la diferencia absoluta estandarizada de las medias es utilizada como medida del desbalance en las variables. Los valores se calculan antes y después del matching, Rosenbaum y Rubin, (1985).

Sean:

- \bar{x}_{tk} la media de la variable k en el grupo tratado, antes del matching.
- \bar{x}_{ck} la media de la variable k en el grupo de control, antes del matching.
- \bar{x}_{cmk} la media de la variable k en el grupo asemejado de control.

Sean entonces, la diferencia absoluta estandarizada antes del matching:

$$sd_{bk} = \frac{|\bar{x}_{tk} - \bar{x}_{ck}|}{\sqrt{(s_{tk}^2 + s_{ck}^2)/2}}$$

mientras que, la diferencia absoluta estandarizada después del matching es:

$$sd_{mk} = \frac{|\bar{x}_{tk} - \bar{x}_{cmk}|}{\sqrt{(s_{tk}^2 + s_{ck}^2)/2}}.$$

La ventaja principal de la diferencia absoluta estandarizada sobre la diferencia no estandarizada es que las variables con escalas diferentes pueden ser mostrados en un mismo gráfico para su comparación.

El balance puede observarse gráficamente en un diagrama de cajas de las diferencias estandarizadas agrupadas en "antes del matching" y "después del matching". Otra representación útil es el gráfico

QQ de los valores-p comparados con la distribución uniforme.

Sensibilidad del diseño

La sensibilidad del diseño investiga el posible efecto del sesgo oculto en los resultados.

Los sesgos abiertos pueden ser controlados adecuadamente con una cuidadosa aplicación del emparejamiento, la estratificación y el ajuste basado en modelos. Sin embargo, la posible presencia de efectos asociados a variables no observadas requiere atención adicional. Dos posibilidades de control son, (a) el análisis de sensibilidad y (b) la elaboración de teorías.

El Análisis de Sensibilidad indica la magnitud de sesgo oculto que es necesario para alterar las conclusiones cualitativas del estudio. Los estudios observacionales pueden ser muy distintos en su sensibilidad al sesgo oculto, por tanto, es importante conocer si un estudio particular es sensible a pequeños sesgos o insensible a sesgos muy grandes.

Un Análisis de Sensibilidad para estudios observacionales intenta detectar en qué medida los sesgos ocultos de variada magnitud pueden alterar las conclusiones.

Se busca determinar cómo las conclusiones basadas en un conjunto de suposiciones son modificadas si este conjunto fuera alterado en diversos grados.

El marco de referencia esta dado por dos modelos que expresan en términos generales la semejanza de los individuos a ser estudiados. El modelo simple o "ingenuo", que afirma: "los individuos que parecen semejantes, efectivamente lo son", mientras que, el modelo realista que sostiene: "los individuos que parecen semejantes, en realidad podrían no serlo".

Modelo Simple para el análisis de sensibilidad

Se define un modelo "simple" como aquel que describe individuos que parecen semejantes y efectivamente lo son. Si el modelo simple es cierto, en un experimento de pares aleatorizados, la distribución del vector que representa las asignaciones de los tratamientos, Z , puede reconstruirse por "matching"

para las variables observadas x .

La sensibilidad de un estudio observacional a los sesgos originados por variables no observadas, es la magnitud del alejamiento con relación al modelo simple, que debe estar presente para que se modifiquen las conclusiones del estudio.

Se entiende por Modelo Simple a la siguiente situación:

Dos individuos k y l con las mismas variables observadas x_k y x_l , tienen la misma probabilidad de recibir el tratamiento condicionado a (r_p, r_C, x, u) , esto es:

$$\pi_k = \pi_l$$

$$\pi_k = \Pr(Z_k = 1 | \mathbf{r}_{Tk}, \mathbf{r}_{Ck}, \mathbf{x}_k, \mathbf{u}_k)$$

$$\pi_l = \Pr(Z_l = 1 | \mathbf{r}_{Tl}, \mathbf{r}_{Cl}, \mathbf{x}_l, \mathbf{u}_l)$$

El modelo simple toma en consideración las mismas probabilidades pero bajo la suposición de que este modelo podría no ser el adecuado en un grado expresado por un parámetro $\Gamma \geq 1$.

Los odds para dos individuos que tienen el mismo conjunto de variables observadas es igual a,

$$\frac{\pi_k}{1 - \pi_k} \text{ para el individuo } k,$$

$$\frac{\pi_l}{1 - \pi_l} \text{ para el individuo } l.$$

Ambos difieren a lo sumo en un múltiplo de Γ , en consecuencia:

$$\frac{1}{\Gamma} \leq \frac{\frac{\pi_k}{1 - \pi_k}}{\frac{\pi_l}{1 - \pi_l}} \leq \Gamma$$

siempre que $x_k = x_l$.

Cuando $\Gamma = 1$ entonces $\pi_k = \pi_l$, esto es, el modelo simple es cierto.

Cuando $\pi_k \rightarrow 0$ o $\pi_k \rightarrow 1$, entonces $\Gamma \rightarrow \infty$; en consecuencia resulta que $1 \leq \Gamma < \infty$.

Análisis de sensibilidad en un diseño observacional con pares semejantes

Si además, se cumple que exactamente un individuo

recibe el tratamiento y el otro recibe el control, es decir, $Z_k + Z_l = 1$, entonces la probabilidad de que k reciba tratamiento y l reciba el control esta dado por,

$$\Pr[Z_k = 1; Z_l = 0 | \mathbf{r}_{Tk}, \mathbf{r}_{Ck}, \mathbf{x}_k, \mathbf{u}_k; \mathbf{r}_{Tl}, \mathbf{r}_{Cl}, \mathbf{x}_l, \mathbf{u}_l] \\ = \frac{\pi_k}{\pi_k + \pi_l}.$$

Si el modelo de sensibilidad es cierto, entonces se verifica que:

$$\frac{1}{1 + \Gamma} \leq \frac{\pi_k}{\pi_k + \pi_l} \leq \frac{\Gamma}{1 + \Gamma},$$

para los individuos emparejados, en el cual uno es tratado y el otro sirve como control.

Cuando $\Gamma \rightarrow \infty$, el límite inferior en la ecuación anterior tiende a cero, mientras que el límite superior tiende a uno.

En un estudio se emparejan $2I$ individuos del total L en toda la población.

Enumerando los individuos emparejados en I pares de dos individuos, donde $i = 1, 2, \dots, I$ y $j = 1, 2$ de modo que $x_{i1} = x_{i2}$ y $Z_{i1} = 1 - Z_{i2}$ en cada uno de los I pares.

Si las condiciones del modelo simple son ciertas, entonces la distribución de las asignaciones de los tratamientos en los I pares cumple,

$$Z_{i1} \text{ son mutuamente independientes, } \dots i = 1, \dots, I, \quad (3)$$

$$Z_{i2} = 1 - Z_{i1}, \quad (4)$$

$$\frac{1}{1 + \Gamma} \leq \frac{\pi_{i1}}{\pi_{i1} + \pi_{i2}} \leq \frac{\Gamma}{1 + \Gamma}. \quad (5)$$

Esta distribución es similar a la distribución de asignaciones en un experimento de pares aleatorizados, la diferencia radica en que la probabilidad de asignación de tratamientos en el diseño experimental es $1/2$ para $i = 1, 2, \dots, I$, mientras que en las ecuaciones (1) a (3), las probabilidades de asignación de tratamientos a individuos pueden ser diferentes de un par a otro y son desconocidas, pero están acotadas por $1/(1 + \Gamma)$ y $\Gamma/(1 + \Gamma)$.

Se supone ahora, que ha sido calculado un valor-*p* para el diseño observacional emparejado en base a las variables observadas.

Si el modelo simple es cierto, esto es $\Gamma = 1$, las inferencias tendrían las propiedades usuales del diseño aleatorizado.

Si $\Gamma > 1$, es posible calcular como podrían cambiar estos valores-*p*, (es posible hacer lo mismo con las estimaciones puntuales y los intervalos de confianza), para un valor específico de Γ .

Todo estudio es sensible a un sesgo suficientemente grande. En consecuencia siempre existe algún valor de Γ tal que para ese valor y otros mayores, el intervalo de posibles *p*-valores incluye pequeños valores y al mismo tiempo valores grandes.

El análisis de sensibilidad muestra como la inferencia puede cambiar con Γ .

RESULTADOS Y DISCUSIÓN

Caso de Estudio

El estudio a evaluar se ha realizado en el lapso de un año y medio, entre 2010 y 2011. Fue conducido por un centro de tratamiento especializado de lesiones, usualmente en las extremidades inferiores, que presentan dificultades en su curación.

El diseño es experimental, exploratorio y comparativo, ciego simple, en el cual fueron seleccionados, según criterios de inclusión y exclusión especificados, 30 pacientes con ciertas características bien definidas, en particular, pacientes cuya cicatrización no ha sido superior a 1 mm^2 una vez transcurrido 30 días de tratamiento. La mitad de ellos recibe un tratamiento estándar (Tratamiento de Control, en lo que sigue simplemente Control), mientras que la otra mitad, recibe una terapia alternativa. El propósito del estudio fue comprobar si la terapia alternativa, (Tratamiento Nuevo, en lo que sigue simplemente Tratamiento), mejora los resultados obtenidos.

Una vez seleccionado el paciente, y luego de una evaluación para determinar su línea base, es asignado aleatoriamente a uno de los dos tratamientos. Debe notarse que por requerimientos del diseño, están fijados la cantidad de pacientes para cada grupo de tratamiento; esto es, 15 individuos

recibirán el control y 15 recibirán el tratamiento experimental.

El paciente es mantenido en el estudio durante tres meses, su estado clínico es controlado durante este período, y recibe, en intervalos regulares, la curación con el tratamiento designado. Un médico se encarga de la condición clínica, mientras que una enfermera especializada se ocupa de realizar las curaciones, además de evaluar y registrar la evolución de la lesión para su posterior estudio.

La respuesta de interés se define como la disminución del área de la úlcera durante el período de estudio, la cual se estima para cada paciente, como la diferencia entre el área de la lesión al inicio del estudio (AT1) y el área al final (ATMAX). En cada caso, el tiempo final de estudio puede ser distinto, un individuo puede culminar su tratamiento antes de los tres meses ya sea por curación de su lesión o por abandono del estudio. En este estudio no se presentaron casos de abandono, por lo que la culminación se ha debido únicamente por curación de la dolencia.

Edad(EDAD), sexo(SEX), residencia(COMUN), peso(PESO), talla(TALLA), índice de masa corporal(IMC), cantidad de antecedentes mórbidos(AMSUM), cantidad de medicamentos que recibe al momento de ingresar al estudio (MESUM), cantidad de antecedentes quirúrgicos relacionados con la herida(AQSUM), presión arterial sistólica(PASIS), presión arterial diastólica(PADIA), presión en el brazo (PBRAZO), presión en el tobillo(PTOBIL), índice tobillo-brazo(ITB), nivel de creatinina(CREA), nivel de hematocritos(HCTO), nivel de hemoglobina glicocilada(HCLI01), hemoglucotest(HGT), tiempo de evolución de la dolencia(EVOL) y área inicial (AT1).

El estudio es realizado en un marco experimental, el "entorno controlado" fijado por medio de los criterios de selección y exclusión que busca recoger una muestra relativamente homogénea de individuos con el objeto de que las diferencias observadas sean deducidas como debidas únicamente a los efectos originados por los distintos tratamientos. Su característica exploratoria justifica el pequeño

número de unidades seleccionadas y aún así, ha sido considerablemente difícil llegar al número de sujetos previsto. Una consecuencia importante del tamaño reducido de la muestra es la dificultad para controlar los efectos de la gran cantidad de variables asociadas con los individuos y que potencialmente podrían tener algún efecto no previsto sobre la evolución de la lesión.

Análisis estándar

En el grupo de control, la menor disminución del área de la lesión es de $0,18 \text{ mm}^2$, mientras que la máxima observada $14,13 \text{ mm}^2$, en promedio la disminución ha sido de $3,24 \text{ mm}^2$ con error típico igual a $0,94 \text{ mm}^2$, indicando que el valor real se encuentra dentro de los límites $1,23 \text{ mm}^2$ y $5,25 \text{ mm}^2$ del intervalo de confianza del 95%. La disminución promedio eliminando el 5% más extremo se reduce a $2,18 \text{ mm}^2$, unos 12% inferior, que sugiere valores extremos relativamente importantes. La mitad de las disminuciones es mayor a $2,27 \text{ mm}^2$.

Por otra parte, en el grupo de tratamiento se registra $-0,39 \text{ mm}^2$ como valor mínimo, que en realidad significa un aumento de la superficie al final del tratamiento, siendo el máximo observado $15,74 \text{ mm}^2$, lo que indica un rango de $16,12 \text{ mm}^2$ que es mayor al observado en el grupo de control. El intervalo de confianza del 95% tiene límites $1,63 \text{ mm}^2$ y $7,06 \text{ mm}^2$ con centro en $4,35 \text{ mm}^2$ superior a la disminución promedio del grupo control; el promedio truncado es similar, aproximadamente 4 mm^2 . Sin embargo, el 50% de los pacientes tiene una reducción menor o igual a $1,71 \text{ mm}^2$.

Estudiando la disminución del área de la lesión ($\Delta\text{AREA} = \text{AT}-\text{ATMAX}$) se han encontrado los siguientes resultados.

En la Tabla 1 se ve la reducción media en cada grupo; la mayor disminución corresponde al grupo

Tabla 1. Disminución del área de la lesión por Grupo de tratamiento.

| | Grupo | N | Media | Err. tip. media |
|-------------------|-------|----|--------|-----------------|
| Cambio en el área | TRAT | 15 | 4.3467 | 1.26 |
| | CONT | 15 | 3.2417 | 0.94 |

Tabla 2. Prueba de igualdad de la disminución promedio del área en cada grupo.

| Dif. media | gl | Sig (bilat.) | Err. tip. Dif. | L. inf. 95 | L. sup. 95 |
|------------|----|--------------|----------------|------------|------------|
| 1.105 | 28 | 0.488 | 1.57 | -2.12 | 4.33 |

de interés, pero también presenta mayor variabilidad. El coeficiente de variación observado en el grupo de tratamiento es 12% mayor que en el grupo de control.

De la Tabla 2 en el que se muestra la diferencia de la disminución promedio del área de la lesión, por medio del estadístico de Student se concluye que la diferencia $1,1 \text{ mm}^2$ no puede asociarse con efectos sistemáticos y el valor es considerado como resultado de la variación aleatoria.

El procedimiento se desarrolla en cuatro etapas, (1) análisis de las variables, (2) descripción de la respuesta estudiada, (3) construcción de pares semejantes, y (4) evaluación de la sensibilidad. Se comparan los resultados de tres modelos.

1) Mod-1: Modelo 1

Incorpora todas las variables de interés, con las siguientes características: La variable IMC (índice de masa corporal) que es derivada de las variables PESO y TALLA; la variable ITB (índice tobillo-brazo) derivada de PTOBIL y PBRAZO; no incluye la variable AT1.

2) Mod-2: Modelo 2

Es el Modelo 1 del cual se eliminan las variables menos significativas.

3) Mod-3: Modelo 3

Es el Modelo 2 con el agregado de la variable AT1 (Área inicial de la lesión).

Los resultados se obtienen utilizando los programas estadísticos R y SPSS.

Matching

El matching consiste en seleccionar un grupo de

Tabla 3. Distribución de las covariables por Grupo.

| Variable | Min | 1Q | Mediana | 3Q | Max |
|-------------|-------|-------|---------|-------|-------|
| EDAD-cont | 44.0 | 64.0 | 66.0 | 78.0 | 89.0 |
| EDAD-trat | 51.0 | 59.0 | 69.0 | 79.0 | 89.0 |
| IMC-cont | 19.7 | 22.6 | 27.77 | 27.8 | 34.3 |
| IMC-trat | 19.7 | 21.8 | 23.45 | 26.4 | 34.1 |
| PASIS-cont | 100.0 | 130.0 | 130.0 | 140.0 | 170.0 |
| PASIS-trat | 100.0 | 128.0 | 140.0 | 150.0 | 162.0 |
| PADIA-cont | 74.0 | 78.0 | 82.0 | 84.0 | 96.0 |
| PADIA-trat | 70.0 | 80.0 | 82.0 | 86.0 | 100.0 |
| ITB-cont | 0.57 | 0.63 | 0.69 | 0.75 | 0.79 |
| ITB-trat | 0.40 | 0.58 | 0.66 | 0.73 | 0.79 |
| CREAT-cont | 0.7 | 1.1 | 1.10 | 1.3 | 2.9 |
| CREAT-trat | 0.8 | 1.0 | 1.10 | 1.2 | 2.9 |
| HCTO-cont | 32.0 | 36.0 | 41.0 | 42.0 | 44.0 |
| HCTO-trat | 31.0 | 36.0 | 40.0 | 42.0 | 44.5 |
| HGLI01-cont | 5.5 | 6.0 | 7.10 | 8.3 | 11.6 |
| HGLI01-trat | 5.5 | 6.3 | 7.10 | 8.5 | 10.2 |
| HGT-cont | 72.0 | 135.0 | 142.0 | 209.0 | 433.0 |
| HGT-trat | 70.0 | 81.0 | 121.0 | 176.0 | 294.0 |
| EVOL-cont | 0.0 | 2.0 | 3.0 | 4.0 | 6.0 |
| EVOL-trat | 3.0 | 3.0 | 4.0 | 5.0 | 6.0 |
| AT1-cont | 0.3 | 0.8 | 2.55 | 5.3 | 14.4 |
| AT1-trat | 0.2 | 0.6 | 3.60 | 10.4 | 15.9 |

individuos, con características específicas, para aplicarles el tratamiento de interés. Similarmente, se elige otro grupo de forma tal que, sus individuos sean lo mas parecidos, en término de las características específicas, al grupo de tratamiento. A este grupo se administra el tratamiento de control.

El matching es válido bajo el siguiente principio: cuando las características observadas son balanceadas en los dos grupos, entonces están balanceadas con respecto a todas las características relevantes a la respuesta de interés. En consecuencia, cuanto mayor sea el número de variables utilizadas para el balance, mayor es el beneficio obtenido.

Una característica adicional necesaria es que exista una superposición de los valores de cada variable en el grupo de tratamiento y de control. A continuación se revisan gráficamente las distribuciones de variables en cada grupo.

En la Tabla 3 se muestran los estadísticos que identifican la distribución de las variables continuas observadas. Las variables IMC, PADIA, ITB, EVOL y HGT sugieren diferencias sistemáticas en sus distribuciones. Las variables IMC e ITB en el grupo tratamiento tienen valores consistentemente menores que en el grupo de control. Por otra parte, los cuartiles de las variables PADIA, HGT y EVOL tienen valores mayores en el grupo de tratamiento. En particular, el hemoglucotest (HGT) es mucho mayor para los individuos tratados; así mismo, el tiempo de evolución de la lesión en el grupo de tratamiento, es mayor que en el grupo control.

Las restantes variables no muestran patrones notables. Finalmente, en el caso del área inicial de la lesión, AT1, el punto de corte para el tercer cuartil es el doble en el grupo tratado comparado con el grupo control.

De la Tabla 4 se desprende que la distribución de la variable sexo es idéntica en ambos grupos, aunque hay una mayor cantidad de varones que

Tabla 4. Distribución de Sexo por Grupo.

| Sexo | Control | Tratamiento |
|-----------|----------|-------------|
| Femenino | 3 (20%) | 3 (20%) |
| Masculino | 12 (80%) | 12 (80%) |

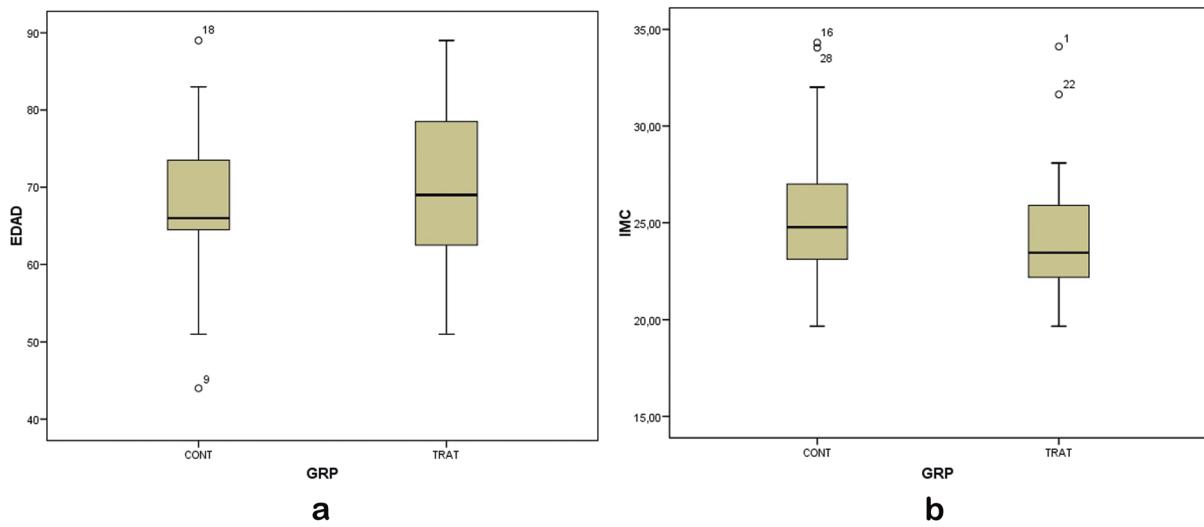


Gráfico 1. Diagrama de Cajas para Edad e IMC. **a)** Cajas para Edad por grupos. **b)** Cajas para IMC por grupos.

mujeres, esta distribución es consistente con la experiencia.

En los Gráficos 1a y 1b se muestran las distribuciones de la edad y del IMC, donde en líneas generales se observa que los valores respectivos son similares en los grupos de tratamiento y control.

Similares distribuciones presentan las variables que registran la presión arterial sistólica y diastólica, tal como se ve en los Gráficos 2a y 2b, aunque

puede notarse un desbalance mas importante en el comportamiento de la presión sistólica

Los diagramas de cajas y bigotes para la distribución de características asociadas con la presencia de azúcar en la sangre muestran valores que se superponen para ambos grupos, en el caso de la hemoglobina glicosilada la superposición es casi exacta, mientras que para la prueba de glucosa existe alta concentración en el rango comprendido

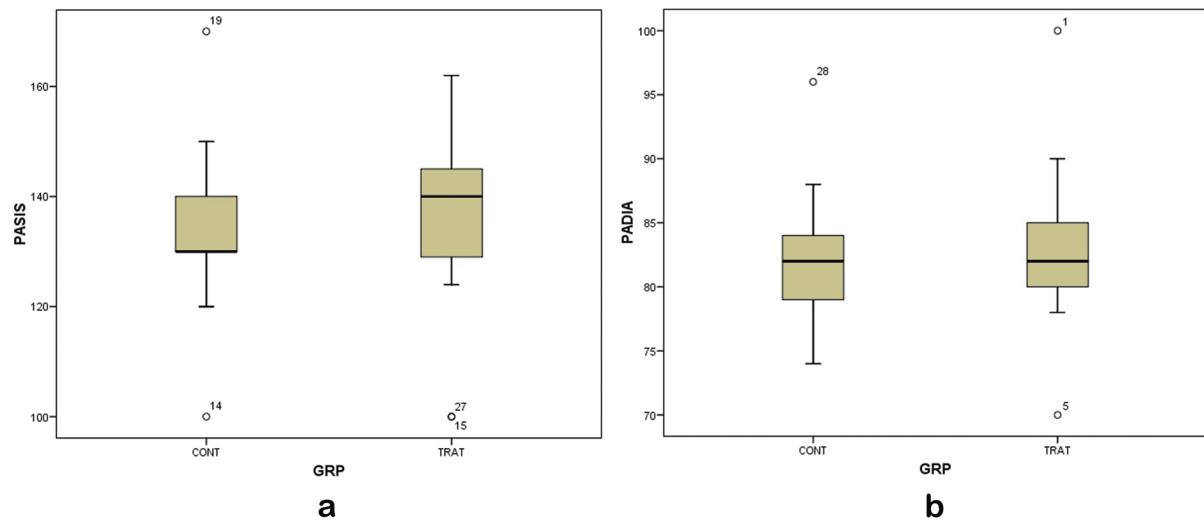


Gráfico 2. Distribución de la Presión Arterial Sistólica y Diastólica. **a)** Cajas para Presión Sistólica. **b)** Cajas para Presión Diastólica.

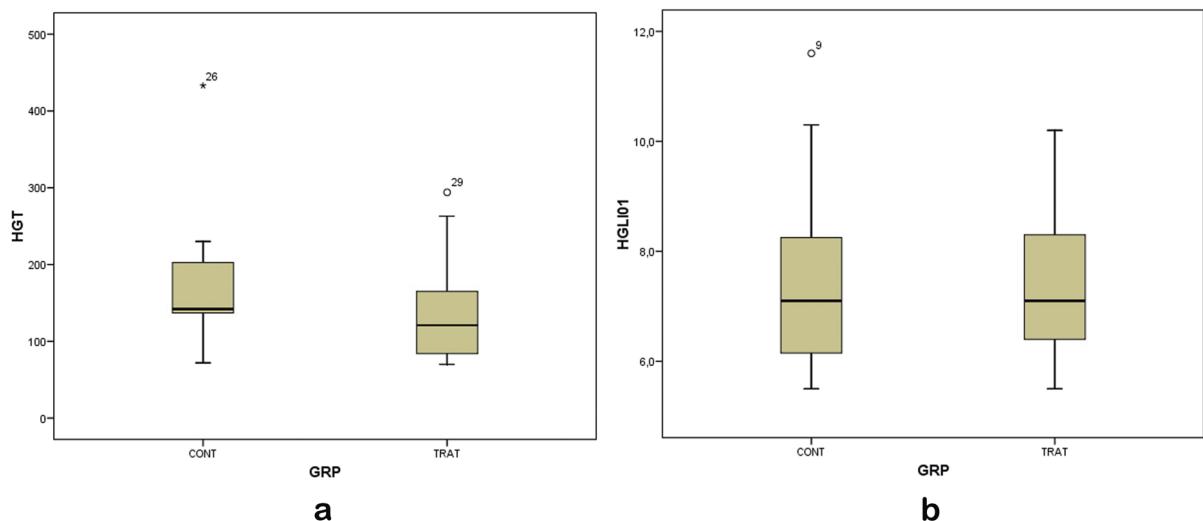


Gráfico 3. Distribución para Glucosa en sangre. **a)** Cajas para HemoGlu Test. **b)** Cajas para Hemoglobina Glicocilada.

entre el segundo y tercer cuartil, ver el Gráfico 3.

En las Tablas 5, 6 y 7, se observa que los valores de las variables se superponen en ambos grupos, aunque también pueden notarse desbalances para ciertos valores.

Tabla 5. Cantidad de Antecedentes Médicos por Grupo.

| Cantidad | Control | Tratamiento |
|----------|------------|-------------|
| 1 | 1 (06.7%) | 2 (13.3%) |
| 2 | 13 (86.6%) | 12 (80.0%) |
| 3 | 1 (06.7%) | 1 (06.7%) |

Al evaluar la distribución del tiempo de evolución de la lesión, Tabla 8, se nota que los pacientes que pertenecen al grupo de tratamiento tienen heridas que son más antiguas que las del grupo de

control. En este sentido, el 36% de los individuos de control con tiempo de evolución menor o igual a tres no tienen contraparte en el grupo que recibe el tratamiento experimental.

Tabla 7. Cantidad de Antecedentes Quirúrgicos por Grupo.

| Cantidad | Control | Tratamiento |
|----------|-----------|-------------|
| 0 | 5 (33.3%) | 7 (46.7%) |
| 1 | 7 (46.7%) | 4 (26.7%) |
| 2 | 3 (20.0%) | 3 (20.0%) |
| 3 | 0 (00.0%) | 1 (06.7%) |

Finalmente, el Gráfico 4 que describe la distribución del área de la lesión al inicio del tratamiento muestra que existe una proporción de pacientes en el grupo de tratamiento que tienen áreas iniciales

Tabla 6. Cantidad de Medicamentos por Grupo.

| Cantidad | Control | Tratamiento |
|----------|-----------|-------------|
| 2 | 0 (00.0%) | 3 (20.0%) |
| 3 | 7 (46.7%) | 2 (13.3%) |
| 4 | 4 (26.7%) | 4 (26.7%) |
| 5 | 2 (13.3%) | 1 (06.7%) |
| 6 | 1 (06.7%) | 3 (20.0%) |
| 7 | 1 (06.7%) | 1 (06.7%) |
| 9 | 0 (06.7%) | 1 (06.7%) |

Tabla 8. Evolución de la lesión por Grupo.

| Evolución de la lesión | Control | Tratamiento |
|------------------------|-----------|-------------|
| 0 (perdido) | 1 (06.7%) | 0 (00.0%) |
| 1 (<14 días) | 1 (06.7%) | 0 (00.0%) |
| 2 (<1 mes) | 3 (20.0%) | 0 (00.0%) |
| 3 (<3 meses) | 5 (33.3%) | 4 (26.7%) |
| 4 (<5 meses) | 2 (13.3%) | 7 (46.7%) |
| 5 (<1 año) | 2 (13.3%) | 3 (20.0%) |
| 6 (\geq 1 año) | 1 (06.7%) | 1 (06.7%) |

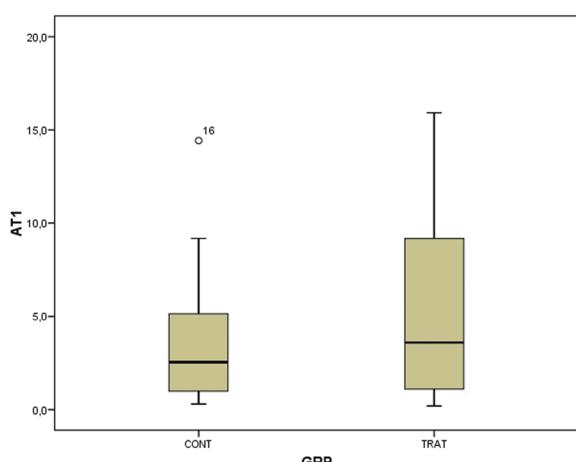


Gráfico 4. Distribución del Área inicial de la lesión.

que son mayores a los del grupo control. Esto es, el valor medio en el grupo de control es $3,07 \text{ mm}^2$ comparado con los $4,07 \text{ mm}^2$ del grupo tratamiento. Se observa también que el límite superior del intervalo de confianza para el grupo de control casi coincide con el límite inferior del intervalo para el grupo que recibe el tratamiento en estudio. Además, la evaluación del grado de superposición en los valores observados, indica que en el grupo control los valores mínimos y máximos son 0 y 6 respectivamente, mientras que en el grupo tratamiento el mínimo es 3 mm^2 y el máximo es igual al grupo control, esto significa que solo el cincuenta por ciento del soporte está superpuesto.

En términos generales, de las consideraciones hechas a las tablas y gráficos mostrados previamente, las variables muestran distintos grados de superposición en sus valores luego de observar sus distribuciones en cada grupo de tratamiento. En ciertos casos, las variables muestran desbalances entre pacientes que podrían ejercer un efecto modificador sobre los resultados observados.

Estimación del puntaje de propensión

El hecho que exista superposición en los valores de las covariables, permite en teoría concretar el matching. Sin embargo, la cantidad de variables disponibles hace que en la práctica sea muy difícil lograrlo separadamente para cada variable, esto es,

lograr un matching multivariado es extremadamente complicado, si no imposible, cuando la dimensión del problema es grande, y más aún si las variables son continuas.

No obstante la dificultad, y tal como se ha visto en la revisión de la teoría, es posible resolver el problema utilizando el puntaje de propensión, reduciendo el problema multivariado a otro unidimensional. El puntaje de propensión es la probabilidad estimada de asignar un individuo al tratamiento condicional al valor de la covariables observada.

Como se ha dicho, el problema de calcular el puntaje de propensión implica decidir (1) que modelo utilizar y (2) cuales variables serán incluidas en el modelo para encontrar la estimación. La primera cuestión es relativamente simple, usualmente son apropiados el modelo logístico o el modelo probit, y como ambos producen resultados similares su elección no representa una dificultad importante.

El segundo aspecto es más serio, pues la elección de las variables debe asegurar la suposición de que el sesgo de selección es eliminado condicionando en el puntaje de propensión. Con este propósito se consideran tres conjuntos, ver Tabla 9,

Tabla 9. Variables para estimar el Puntaje de Propensión.

| Variable | Mod-1 | Mod-2 | Mod-3 |
|----------|-------|-------|-------|
| EDAD | X | X | X |
| SEX | X | X | X |
| COMUN | X | -- | -- |
| IMC | X | X | X |
| AMSUM | X | X | X |
| MESUM | X | X | X |
| AQSUM | X | X | X |
| PASIS | X | -- | -- |
| PADIA | X | -- | -- |
| ITB | X | -- | -- |
| CREA | X | -- | -- |
| HCTO | X | -- | -- |
| HCGLI01 | X | X | X |
| HGT | X | X | X |
| EVOL | X | X | X |
| AT1 | -- | -- | X |

de covariables utilizando el modelo logístico, se retienen aquellas variables que muestren significación mayor, aunque se han utilizado niveles relativamente bajos comparados con los niveles estándar utilizados en la literatura, como son los niveles comunes 0,05, 0,01 o 0,001. Esto se justifica por el hecho de que no se busca un modelo parsimonioso, sino que en lo posible se intenta retener aquellas variables que modifican los valores de la probabilidad de asignación.

El procedimiento requiere que inicialmente se realice la estimación de los puntajes de propensión para cada individuo, que en otras palabras es la probabilidad de que un individuo reciba el tratamiento de interés dado los valores de sus covariables.

Para obtener la estimación se recurre al modelo logístico expresado como:

$$\log\left(\frac{\pi}{1-\pi}\right) = \mathbf{X}\boldsymbol{\beta} \quad (6)$$

Se designa $\hat{\pi} = \hat{e}(\mathbf{x})$ al valor estimado del puntaje de propensión para cada individuo.

La probabilidad de asignación de tratamientos se calcula con los tres modelos previamente enumerados.

• **Modelo 1**

$$\begin{aligned} \log\left(\frac{\pi}{1-\pi}\right) = & \beta_0 + \beta_1 SEX + \beta_2 COMUN \\ & + \beta_3 EDAD + \beta_4 IMC + \beta_5 AMSUM + \\ & + \beta_6 MESUM + \beta_7 AQSUM + \beta_8 PASIS + \\ & + \beta_9 PADIA + \beta_{10} ITB + \beta_{11} CREAT + \\ & + \beta_{12} HCTO + \beta_{13} HGLI01 + \beta_{14} HGT + \\ & + \beta_{16} EVOL \end{aligned}$$

Este modelo está fuertemente parametrizado. Como modelo alternativo se define el Modelo 2 en el cual se eliminan aquellas variables con menor nivel de significación.

• **Modelo 2**

$$\begin{aligned} \log\left(\frac{\pi}{1-\pi}\right) = & \beta_0 + \beta_1 SEX + \beta_2 EDAD + \beta_3 IMC \\ & + \beta_4 AMSUM + \beta_5 MESUM + \\ & + \beta_6 AQSUM + \beta_7 HGLI01 + \beta_8 HGT + \\ & + \beta_9 EVOL \end{aligned}$$

A este modelo se le incorpora la variable AT1, que corresponde al área inicial de la lesión, resul-

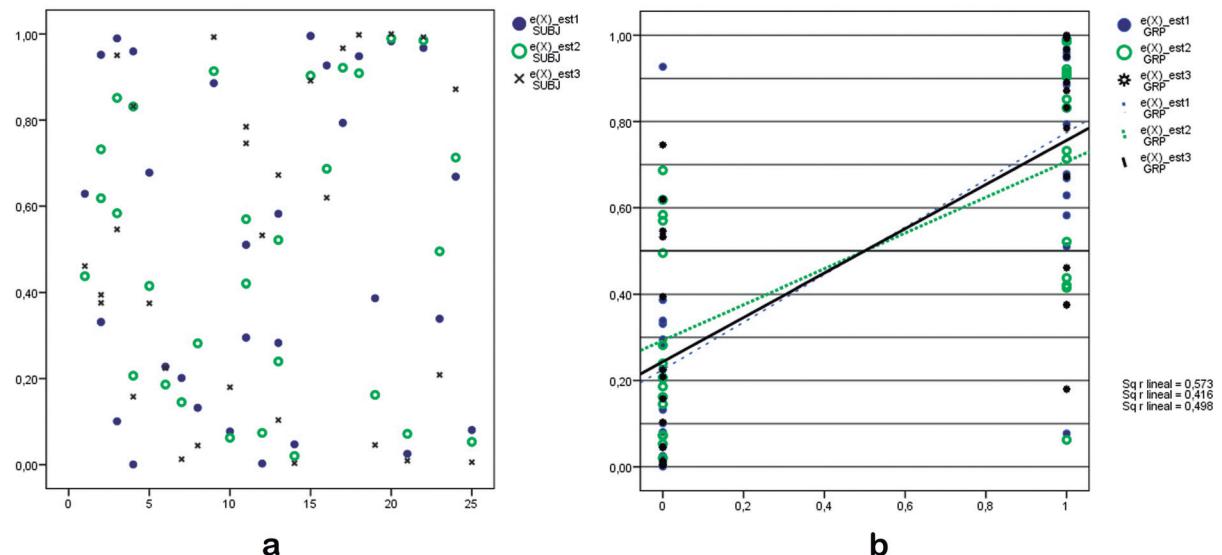


Gráfico 5. Puntajes de Propensión estimados. **a)** Propensión por individuos. **b)** Propensión por grupos.

tando el Modelo 3.

• Modelo 3

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 SEX + \beta_2 EDAD + \beta_3 IMC + \beta_4 AMSUM + \beta_5 MESUM + \beta_6 AQSUM + \beta_7 HGLI01 + \beta_8 HGT + \beta_9 EVOL + \beta_{10} AT1$$

De la inspección de los resultados son notorios los siguientes aspectos:

- 1) El modelo logístico, Modelo 1, proporciona 93% de clasificaciones correctas para cada grupo, TRAT y CON, esto sugiere que están incorporadas las variables que efectivamente influencian la probabilidad de pertenecer a un grupo específico. En otras palabras, una baja cantidad de individuos correctamente clasificados hubiera sido un indicio de la existencia de factores adicionales que afectan el puntaje de propensión.
- 2) Es notable el efecto que tiene cada variable individual sobre el puntaje de propensión estimado para cada paciente; los Gráficos 5 y 6 no sugieren patrón alguno en el comportamiento; para ciertos sujetos el efecto resultante es un aumento en la probabilidad de recibir el tratamiento, mientras que en otros el efecto es totalmente opuesto. Quizás la inclusión de efectos aleatorios asociados con los pacientes brinde mayor visión, pero en este caso la cantidad de unidades disponibles no lo permite.
- 3) Desde el punto de vista de la construcción de modelos, la única variable que muestra significación estadística es EVOL (tiempo de evolución de la lesión). En este caso, la cantidad de unidades correctamente clasificadas es del 67% para los sujetos del grupo control, y del 73,3% en el grupo tratado, alcanzando un global de 70%. Este valor es apenas 3% inferior a la tasa global de clasificación correcta obtenida con el Modelo 2 y el Modelo 3 de la Tabla 9.
- 4) Si bien las tasas de clasificación correctas

son similares en los modelos intermedios y el modelo del punto tres anterior, los cambios en el valor de la probabilidad de recibir el tratamiento son muy importantes, y no parecen seguir un patrón determinado.

- 5) En el Gráfico 5b, se compara la distribución de $\hat{e}(x)$ por grupo. En la misma se evidencia que los puntajes asociados con el grupo de tratamiento son en promedio superiores que los correspondientes al grupo de control. Resulta importante señalar que existe apreciable superposición de las probabilidades observadas en ambos grupos, sin embargo la concentración es mayor hacia los valores inferiores en el grupo de control, y hacia el extremo opuesto en el grupo de tratamiento.
- 6) Los Gráficos en 6, particularmente 6b, muestran nuevamente que el puntaje de propensión es consistentemente superior en el grupo de tratamiento que en el grupo de control.

Matching de las unidades

Una vez calculadas las probabilidades de que un individuo reciba el tratamiento se procede a realizar los emparejamientos o matching. El procedimiento consiste en seleccionar un individuo del grupo de control y emparejarlo a uno que ha recibido el tratamiento. Existen numerosos algoritmos para concretar los emparejamientos, y difieren entre si en la forma que definen un entorno del individuo que recibe el tratamiento y la ponderación que se le asigna. Entre los distintos métodos se tienen, (a) el matching por "vecino más próximo", con sus variantes matching con remplazo y sin remplazo, (b) matching por calibrador y una modificación llamada matching "circular", (c) el matching por estratificación, y (d) el matching del núcleo.

En este caso se utilizará el matching por calibrador, que consiste esencialmente en estimar las distancias entre individuos por medio del puntaje de propensión, las distancias tienen un margen de tolerancia que marca la máxima distancia tolerable

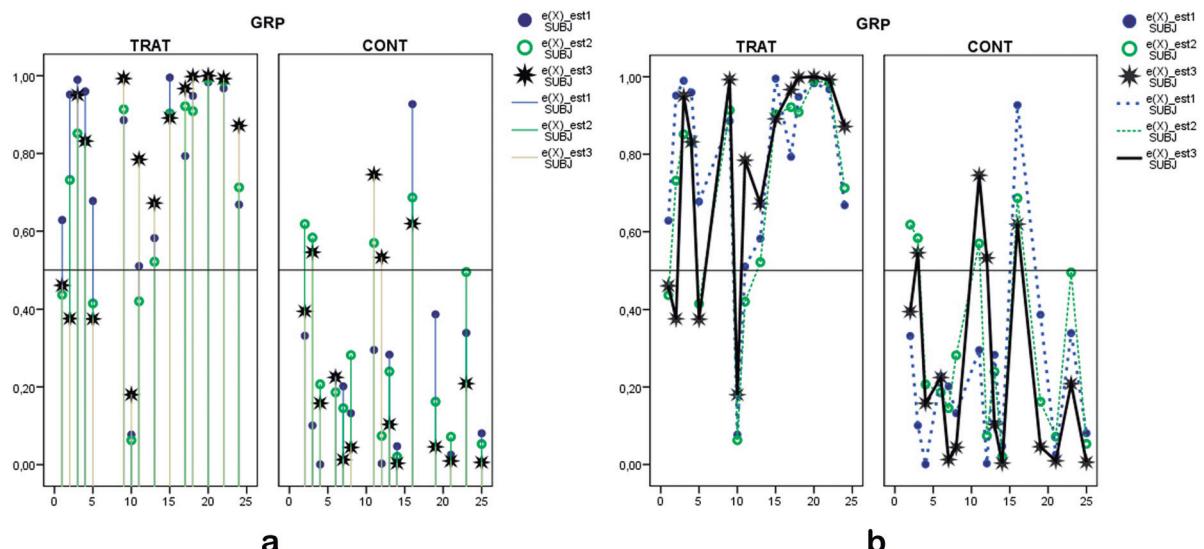


Gráfico 6. Comparación de Puntajes de Propensión estimados. a) Compara individuos. b) Compara grupos.

(el calibrador). El calibrador es similar a permitir que un control utilizado pueda nuevamente ser emparejado con otro, de esta forma se evitan emparejamientos de mala calidad, aunque podría ser al costo de aumentar la varianza de la estimación. El método elige un individuo del grupo control siempre que esté dentro del rango de distancia tolerada y es el más cercano en términos del puntaje de propensión. La modificación del matching circular significa que se empareja no solo el más cercano sino todos los individuos dentro del rango de tolerancia.

Inicialmente se construye la matriz distancia de Mahalanobis basada en rangos para cada modelo logístico estimado; posteriormente se comparan estas distancias con un calibrador, que en general deberá estar dentro del 20% de la desviación estándar del puntaje de propensión, en caso que las distancias entre individuos sea mayor, se impone una penalización. Los cálculos se realizan utilizando rutinas en R; una rutina útil es la llamada "fullmatch" que forma parte del paquete "optmatch".

El algoritmo aplicado a los resultados obtenidos con los modelos para estimar los puntajes de propensión producen los siguientes emparejamientos.

En la rutina se fuerza a que todos los individuos sean emparejados

Modelo 1

1m8, 2m1, 3m11, 4m2, 5m15, 6m12, 7m3, 8m10, 9m4, 10m5, 11m6, 12m5, 13m3, 14m7, 15m7, 16m8, 17m9, 18m9, 19m10, 20m14, 21m11, 22m2, 23m1, 24m12, 25m4, 26m13, 27m6, 28m14, 29m13, 30m15

Modelo 2

1m6, 2m1, 3m8, 4m2, 5m2, 6m3, 7m3, 8m5, 9m4, 10m5, 11m6, 12m12, 13m10, 14m7, 15m15, 16m8, 17m9, 18m9, 19m10, 20m11, 21m11, 22m14, 23m1, 24m12, 25m4, 26m13, 27m7, 28m14, 29m13, 30m15

Modelo 3

1m6, 2m1, 3m13, 4m2, 5m2, 6m3, 7m3, 8m5, 9m4, 10m5, 11m6, 12m8, 13m12, 14m7, 15m1, 16m8, 17m9, 18m9, 19m10, 20m11, 21m11, 22m14, 23m10, 24m12, 25m4, 26m13, 27m7, 28m14, 29m15, 30m15

Inspeccionando los patrones de emparejamientos obtenidos en cada modelo, resultan que son 17 los pares asemejados que aparecen en los tres modelos, a saber: (2,1), (4,2), (7,3), (9,4), (10,5), (11,6), (14,7), (16,8), (17,9), (18,9), (19,10), (21,11), (24,12), (25,4), (26,13), (28,14) y (30,15).

Tabla 10. Sujetos emparejados en el grupo TRAT.

| ID Paciente | Mod-1 | Mod-2 | Mod-3 | Patrón |
|-------------|-------|-------|-------|------------|
| 1 | 8 | 6 | 6 | >(T:T,T,T) |
| 3 | 11 | 8 | 13 | (T:C,T,T,) |
| 5 | 15 | 2 | 2 | (T:T,C,T) |
| 6 | 12 | 3 | 3 | >(T:T,T,T) |
| 8 | 10 | 5 | 5 | (T:C,T,T) |
| 12 | 5 | 12 | 8 | >(T:T,T,T) |
| 13 | 3 | 10 | 12 | (T:T,C,T) |
| 15 | 7 | 15 | 1 | (T:C,T,T) |
| 17 | 9 | 9 | 9 | (T:C,C,C)< |
| 20 | 14 | 11 | 11 | (T:C,C,C)< |
| 22 | 2 | 14 | 14 | (T:T,T,C) |
| 23 | 1 | 1 | 10 | (T:T,T,C) |
| 25 | 4 | 4 | 4 | (T:C,C,C)< |
| 27 | 6 | 7 | 7 | (T:T,C,C) |
| 29 | 13 | 13 | 15 | >(T:T,T,T) |

Calidad del matching

Los tres conjuntos de emparejamientos, al ser comparados con relación a la pertenencia de los individuos a cada grupo de tratamiento, revelan aspectos interesantes del estudio realizado. En la Tabla 10, se muestra como los individuos del grupo tratamiento son emparejados en virtud del algoritmo de emparejamiento basado en el puntaje de propensión.

Los individuos señalados con ">" han sido consistentemente emparejados con otros pacientes del mismo grupo. Es decir, los valores de probabilidad estimados con el modelo indicado indican que los individuos son propensos a ser emparejados con otros individuos que han recibido el mismo tratamiento.

Por otra parte, los individuos señalados con "<" han sido emparejados sólo con individuos que pertenecen al grupo opuesto. Las otras agrupaciones muestran combinaciones de individuos tratados que según el modelo utilizado son emparejados con sujetos del grupo tratado o del grupo control.

En consecuencia, los pacientes 1, 6, 12 y 29 no tienen otros sujetos "cercanos" que hayan sido asignados al tratamiento de control, y bajo estas

condiciones no tienen posibles pares que hayan recibido el tratamiento estándar.

Sin embargo, tal como se verá a continuación, los resultados en el grupo de control han sido más homogéneos en el sentido de que los emparejamientos han sido consistentes con los tres modelos considerados; de hecho, no se ha observado que un paciente sea emparejado con sujetos del mismo grupo y también con sujetos de otro grupo, como se muestra en la Tabla 11.

En este caso, los pacientes 2, 7, 10, 11, 16, 24, 26, y 30 tienen únicamente pares del grupo tratamiento, mientras que los restantes 4, 9, 14, 18, 19, 21, y 28 han sido emparejados con individuos que recibieron el control; esto es, el algoritmo produce alrededor del 50% de los sujetos emparejados con pacientes del mismo grupo.

Resumiendo, el proceso de emparejar los pacientes por medio del uso del calibrador depende fuertemente del modelo utilizado para estimar las probabilidades de asignación de los tratamientos.

Como la técnica conduce a una selección de individuos con remplazo, resulta que la sucesión de pares no es única, por lo que una alternativa es forzar un emparejamiento que no permita seleccio-

Tabla 11. Sujetos emparejados en el grupo CON.

| ID Paciente | Mod-1 | Mod-2 | Mod-3 | Patrón |
|-------------|-------|-------|-------|------------|
| 2 | 1 | 1 | 1 | (C:T,T,T)< |
| 4 | 2 | 2 | 2 | >(C:C,C,C) |
| 7 | 3 | 3 | 3 | (C:T,T,T)< |
| 9 | 4 | 4 | 4 | >(C:C,C,C) |
| 10 | 5 | 5 | 5 | (C:T,T,T)< |
| 11 | 6 | 6 | 6 | (C:T,T,T)< |
| 14 | 7 | 7 | 7 | >(C:C,C,C) |
| 16 | 8 | 8 | 8 | (C:T,T,T)< |
| 18 | 9 | 9 | 9 | >(C:C,C,C) |
| 19 | 10 | 10 | 10 | >(C:C,C,C) |
| 21 | 11 | 11 | 11 | >(C:C,C,C) |
| 24 | 12 | 12 | 12 | (C:T,T,T)< |
| 26 | 13 | 13 | 13 | (C:T,T,T)< |
| 28 | 14 | 14 | 14 | >(C:C,C,C) |
| 30 | 15 | 15 | 15 | (C:T,T,T)< |

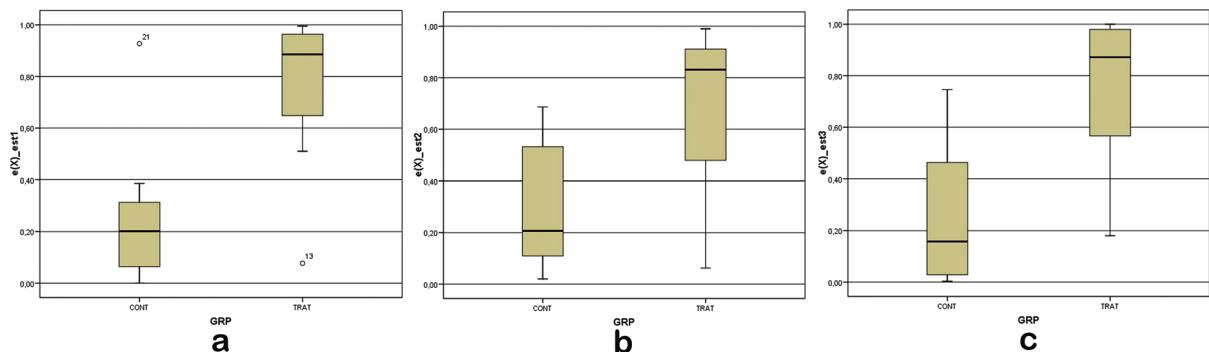


Gráfico 7. Puntaje de propensión estimado por grupo y para cada modelo. a) Modelo 1. b) Modelo 2. c) Modelo 3.

nes repetidas, esta alternativa tiene la desventaja de ser ineficiente cuando la cantidad de individuos disponibles para formar los pares es pequeña, como es este caso en estudio. Una inspección cuidadosa puede mejorar la selección de pares.

Alternativamente, para la comprobación de la calidad del matching se puede recurrir al análisis de la distribución de los puntajes de propensión individuales, tal como se ha visto en el Gráfico 6.

Una comprobación adicional lo da el Gráfico 7 en el cual se ven claramente las diferencias en las distribuciones de los puntajes de propensión según el modelo utilizado.

Un emparejamiento adecuado debe tener un soporte común, más aún, las distribuciones deberían ser similares. De los valores mínimos y máximos mostrados en la Tabla 12, se deduce que para el Modelo 1 un soporte común es el rango (0,08; 0,93), para el Modelo 2 el rango común está dado por (0,06; 0,69), y para el Modelo 3, los valores comprendidos entre 0,18 y 0,75 constituyen el soporte común para los grupos Control y Tratamiento.

Tabla 12. Puntajes de propensión por 100.

| Modelo | Grupo | Min | Max |
|--------|-------|-----|-----|
| 1 | TRA | 8 | 100 |
| | CON | 0 | 93 |
| 2 | TRA | 6 | 99 |
| | CON | 2 | 69 |
| 3 | TRA | 18 | 100 |
| | CON | 0 | 75 |

Habiendo determinado los emparejamientos, definidos los grupos de control y de tratamiento, y evaluado la calidad del matching, el paso siguiente consiste en realizar el cálculo del efecto promedio y determinar el error estándar asociado.

Este paso no se realiza en este trabajo, pues el interés está centrado fundamentalmente en los aspectos del diseño. Se incluye esta sección para completar la exposición.

Una posibilidad sencilla es calcular las medias muestrales en ambos grupos y compararlos utilizando un estadístico apropiado. La dificultad consiste en calcular la varianza del efecto de tratamiento ya que debe incluir la varianza asociada con la estimación del puntaje de propensión, la selección del soporte común y posiblemente el orden en que son elegidos los individuos para el emparejamiento. Lechner (2002) sugiere la utilización de la técnica del bootstrap, aunque Imbens (2004) sostiene que existe poca evidencia que lo justifique.

Sensibilidad

Tal como se ha visto, los pares semejantes tienen por objeto balancear las covariables observadas, esto es válido cuando efectivamente no existe sesgo oculto debido a covariables no observadas. El análisis de sensibilidad es la herramienta para evaluar la magnitud del efecto, debido a covariables no medidas, que es necesario para invalidar el resultado actualmente observado.

A los efectos del análisis de sensibilidad son utilizados modelos que describen el comportamiento

de la probabilidad de asignación de tratamientos a cada sujeto del estudio.

De acuerdo con Rosenbaum (2005a) y Rosenbaum (2010), uno de los modelos utilizados es expresado en términos de las covariables observadas y en este trabajo se identifican como Modelo-1, Modelo 2 y Modelo-3.

Con cada modelo se calcula el puntaje de propensión asociado a cada sujeto, el cual es, por definición, la probabilidad de recibir el tratamiento dependiendo de los valores de las covariables incorporadas en la ecuación. El valor $\Gamma = 1$ corresponde a un estudio libre de sesgo oculto, mientras que valores $\Gamma \geq 1$, indican que, para dos individuos cualquiera, las ventajas (odds) de recibir el tratamiento en estudio difieren cuando más, en un factor de Γ .

Un modelo alternativo para el análisis de sensibilidad se expresa en términos de las covariables no observadas u . Los resultados obtenidos son similares a los que producen los modelos del párrafo anterior.

Una vez obtenidos los pares semejantes, se han definido 15 estratos con 2 individuos cada uno; en cada estrato es calculado el puntaje de propensión, esto es, $\Pr(\mathbf{Z} = \mathbf{z}|\mathbf{x}_i)$.

Estos valores de probabilidad constituyen la distribución del puntaje de propensión para cada configuración de Γ y \mathbf{u} . El análisis muestra la sensibilidad de las inferencias para un conjunto de valores de Γ .

En éste trabajo, el estadístico Rangos con Signo de Wilcoxon, para el caso de pares semejantes, es

suma de 15 variables aleatorias independientes. La probabilidad asociada al valor del estadístico es desconocida debido a la potencial presencia de covariables no observadas, aunque están acotadas superiormente por $p^+ = \Gamma/(1 + \Gamma)$ e inferiormente por $p^- = \Gamma/(1 + \Gamma)$. Se define entonces T^+ , como suma de 15 variables independientes, donde cada una de las variables toma valores distintos de cero con probabilidad p^+ y valores iguales a cero con probabilidad $1 - p^+$. Similarmente se define T^- , reemplazando p^+ por p^- .

Si el tratamiento no tiene efecto, para cada $\Gamma \geq 0$ se tiene,

$$\Pr(T^+ \geq a) \geq \Pr(T \geq a | \mathbf{x}) \geq \Pr(T^- \geq a) \quad \forall a \text{ y } \forall \mathbf{u} \in \mathbf{U}$$

Esta expresión impone límites a los niveles de significación que serían apropiados si las covariables que potencialmente inducen sesgo han sido observadas. Estos límites son calculados para cada valor de Γ .

Utilizando el paquete "rbounds" de R se realizan los cálculos correspondientes, uno para cada modelo especificado previamente, Keele (2010).

Los resultados del análisis de sensibilidad del estudio para comprobar la hipótesis nula de que ambos tratamientos tienen igual efecto terapéutico sobre las lesiones.

En cada par, uno recibe el tratamiento experimental y el otro recibe el tratamiento control. Los emparejamientos se han realizado con respecto a las respectivas covariables incluidas en los modelos especificados.

Tabla 13. Límite inferior (P_{\min}) y superior (P_{\max}) de los valores-p.

| Γ | Modelo 1 | | Modelo 2 | | Modelo 3 | |
|----------|------------|------------|------------|------------|------------|------------|
| | P_{\min} | P_{\max} | P_{\min} | P_{\max} | P_{\min} | P_{\max} |
| 1 | 0.2478 | 0.2478 | 0.3046 | 0.3046 | 0.2302 | 0.2302 |
| 2 | 0.0269 | 0.6851 | 0.0403 | 0.7462 | 0.0234 | 0.6634 |
| 3 | 0.0029 | 0.8811 | 0.0053 | 0.9158 | 0.0024 | 0.8676 |
| 4 | 0.0003 | 0.9558 | 0.0007 | 0.9724 | 0.0003 | 0.9488 |
| 5 | < 0.001 | 0.9836 | < 0.001 | 0.9909 | < 0.001 | 0.9802 |
| 6 | < 0.001 | > 0.99 | < 0.001 | > 0.99 | < 0.001 | 0.9923 |
| 7 | < 0.001 | > 0.99 | < 0.001 | > 0.99 | < 0.001 | > 0.99 |
| 8 | < 0.001 | > 0.99 | < 0.001 | > 0.99 | < 0.001 | > 0.99 |

En la Tabla 13 se consignan los límites inferiores y superiores de los valores-p correspondientes a desvíos con respecto a la asignación aleatoria para varias magnitudes de Γ , su interpretación es tal como se vio en el apartado "Sensibilidad del diseño" en el desarrollo del marco teórico.

Cuando $\Gamma = 1$ ambos límites son iguales y corresponde al valor de un diseño aleatorizado. Los límites acotan los posibles valores-p para la prueba de que no hay efectos de tratamientos.

Cuando el intervalo contiene valores pequeños (por ejemplo, 0,05 o menos) y valores grandes (por ejemplo, 0,10 o más) el estudio se vuelve sensible a los sesgos, y cuanto mayor es Γ mayor es el sesgo necesario para distorsionar los resultados.

En el caso del modelo 1, cuando $\Gamma = 2$, los posibles valores-p asociados con las diferencias entre las disminuciones promedio por grupo, están en el rango 0,03 y 0,68. Esto significa que, con este conjunto de datos, ciertas diferencias son poco probables bajo la hipótesis nula de no tratamiento, mientras que otras diferencias no resultan suficientes para rechazarla.

Similar interpretación es aplicable para los modelos 2 y 3.

Como es de esperar con una muestra de pequeño tamaño, en todos los casos el estudio resulta sensible a los efectos del sesgo oculto de pequeña magnitud; de hecho, para el diseño aleatorizado cuando $\Gamma = 1$, los valores-p 0,24, 0,30 y 0,23 no alcanzan para afirmar la existencia de efectos de tratamientos, resultado que es consistente con lo obtenido en el estudio de base.

CONCLUSIONES

Los estudios observacionales constituyen una alternativa válida para realizar inferencias causales cuando el diseño aleatorizado no es factible por cuestiones prácticas o éticas. Las relaciones de causa y efecto son posibles de determinar siempre que se cumplan ciertas condiciones para su validez.

El primer requerimiento consiste en balancear las covariables observadas de forma tal que las unidades en estudio sean lo más similares posibles, de esta forma se evita introducir sesgos debidos a

las diferencias iniciales entre las covariables.

Una herramienta para lograr el balance es el puntaje de propensión que indica la probabilidad de que un individuo reciba el tratamiento de interés condicional a sus valores de covariables. Este puntaje, a su vez es base para construir un calibrador que permite identificar individuos que no son comparables. El desvío estándar de las estimaciones del puntaje de propensión es utilizado para determinar el rango de valores del calibrador. La aplicación de penalizaciones aplicadas a individuos que no satisfacen el calibrador permite mejorar la construcción de los grupos de tratamiento y de control.

El segundo requerimiento plantea la suposición de existencia de posibles sesgos asociados con variables importantes no medidas. El análisis de sensibilidad cuantifica la magnitud del sesgo que debe estar presente para que los resultados obtenidos sean invalidados. Con este análisis se pretende resolver el problema que encierra la falta de aleatorización.

Los conceptos de matching para emparejar individuos, y el análisis de sensibilidad son utilizados para evaluar un pequeño ensayo clínico para determinar el efecto de una nueva terapia para el tratamiento de úlceras crónicas. En este caso particular, la matriz distancia es calculada en base al estadístico Rangos con Signo de Wilcoxon, pues es más robusto que el estadístico basado en los promedios.

Como resultado del matching se ha obtenido un subconjunto de pares que se han comportado consistentemente con tres modelos de estimación del puntaje de propensión. Adicionalmente se ha constatado la extrema influencia de las covariables a nivel individual ya que los resultados no señalan algún patrón obvio de comportamiento de las probabilidades de asignación de tratamientos. Esto podría sugerir la necesidad de incorporar, además de los efectos fijos de las variables, los efectos aleatorios de las unidades en estudio.

De los resultados del análisis de sensibilidad aplicado a los tres modelos de prueba, se concluye que el diseño no tiene capacidad para discriminar la existencia de diferencias de efectos entre los dos grupos comparados. Los valores-p correspondientes

al diseño aleatorizado en todos los casos tienen valores cercanos a 0,20.

Una consecuencia adicional es que el estudio es muy sensible a pequeños sesgos que podrían alterar de manera significativa los resultados obtenidos. La evaluación del estudio se ha basado en asemejar pares de individuos en una relación 1:1. El esquema corresponde a un matching óptimo completo.

La teoría y aplicaciones de los estudios observacionales constituyen una área de la estadística que recibe mucha atención en la actualidad; particularmente, la utilización del matching por medio del puntaje de propensión. En consecuencia, existe una diversidad de caminos posibles.

Posteriores trabajos pueden orientarse a la investigación de la asignación por relaciones variables, e incluso aplicar las técnicas correctivas ante la presencia de desbalances en las covariables, como la técnica del balance fino. También resulta interesante explorar los emparejamientos de grupos de riesgo (risk set matching) o los emparejamientos tomando en cuenta las dosis de tratamiento administrado.

Adicionalmente, el estudio de los algoritmos de optimización y generación de rutinas en R constituyen líneas posibles de estudio.

Se discute también la posibilidad de que el puntaje de propensión se modifique a lo largo del tiempo, como resultado de la experiencia del investigador; ésta es una línea que presenta posibilidades de desarrollo futuro.

Finalmente, se encuentra en investigación las posibles aplicaciones de los estudios observacionales en grandes conjuntos de datos, como los que son originados en la internet.

LITERATURA CITADA

- Bloom, H. (2006). The core analytics of randomized experiments for social research. *MDRC Working paper on research methods*. www.mdr.org.
- Cochran, W.G. (1965). The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society, A* 128: 134-155.
- Concato, J., Shah, N. & Horwitz, R. (2000). Ran-
- domized, controlled trials, observational studies, and the hierarchy of research designs. *The New England journal of medicine*, 242(25): 1887-92.
- Cox, D.R. & Wermuth, N. (2007). Some interpretation issues connected with observational studies. *Bulletin of the International Statistical Institute*, 62: 905-912.
- Gu, X.S. & Rosenbaum, P. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2: 405-420.
- Greeland, S. (2004). An overview of methods for causal inference from observational studies. In Gelman, A. & Meng, X.L. (eds): *Applied Bayesian modeling and causal inference from incomplete-data perspectives*. New Jersey: John Wiley and sons. 440 pp.
- Grimes, D.A. & Schultz, K.F. (2002). A overview of clinical research: the lay of the land. *The Lancet*, 359: 57-61.
- Hansen, B. (2008). The prognostic analogue of the propensity score. *Biometrika*, 95(2):481-488.
- Hirano, K., Imbens, G.W. (2004). The propensity score with continuous treatments in Missing data and Bayesian methods in practice. In Gelman, A. & Meng, X.L. (eds): *Applied Bayesian modeling and causal inference from incomplete-data perspectives*. New Jersey: John Wiley and sons. 440 pp.
- Imbens, G. (2004). Non parametric estimation of average treatment effects under exogeneity: A review. *The review of economics and statistics*, 86(1): 4-29.
- Jüni, P., Altmand, D., Egger, M.,(2001). Assessing the quality of controlled clinical trials. *British Medical journal*, 323: 42-46.
- Keele, L. (2010). *An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data*. http://www.personal.psu.edu/ljk20/rbounds vignette.pdf. 15 pp.
- Lechner, M. (2002). Some practical issues in the evaluation of heterogeneous labour market

- programmes by matching methods. *Statistics in Society*, A 165: 59-82.
- McKinlay, S. (1975). The design and analysis of observational studies-A review. *Journal of the American Statistical Association*, 70(351): 503-520.
- McNeil, D. (1996). *Epidemiological research methods*. New Jersey: John Willey and sons. 316 pp.
- Pirracchio, R., Resche, M. & Chevret, M. (2012). Evaluation of the propensity score methods for estimating marginal odds ratios in case of small sample size. *BMC Medical Research Methodology*, 12: 70.
- Rosenbaum, P. (1989). Optimal matching for observational studies. *Journal of the American Statistical Association*, 84(408): 1024-1032.
- Rosenbaum, P. (2002). *Observeational studies*. Edition 2. New York: Springer. 377 pp.
- Rosenbaum, P. (2004). *Randomized experiments and observational studies: Causal inference in statistics*. <http://www-stat.wharton.upenn.edu/ rosenbap/ExperAndObsHand.pdf>. 10 pp.
- Rosenbaum, P. (2005 a). Observational study. *Encyclopedia of Causal statistics in behavioral sciences*, 3: 1451-1462.
- Rosenbaum, P. (2005 b). Sensitivity analysis in Observational study. *Encyclopedia of Causal statistics in behavioral sciences*, 4: 1809-1814.
- Rosenbaum, P. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477): 191-200.
- Rosenbaum, P. (2010). *Design of observational studies*. New York: Springer. 384 pp.
- Rosenbaum, P. & Rubin, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41-55.
- Rosenbaum, P. & Rubin, D. (1985). Constructing a control group by multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39: 33-38.
- Rubin, D. (1973). Matching to remove bias in observational studies. *Biometrics*, 29: 159--183.
- Rubin, D. (2006a). *Matched sampling for causal effects*. New York: Cambridge University Press. 502 pp.
- Rubin, D. (2006b). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in medicine*, 26(1): 20-36.
- Rubin, D. (2008). For objective causal inference, design trumps analysis. *The annals of applied statistics*, 2(3): 808-840.
- Sobel, M. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476): 1398-1407.
- Zubizarreta, J., Reinke, C., Kelz, R., Silber, J. & Rosenbaum, P. (2011). Matching for several sparse nominal variables in a case--control study of readmission following surgery. *The American Statistician*, 65(4): 229-238.